Are We There Yet? Feasibility of Continuous Stress Assessment via Wireless Physiological Sensors

Md. Mahbubur Rahman*, Rummana Bari*, Amin Ahsan Ali*, Moushumi Sharmin*, Andrew Raij†
Karen Hovsepian[∨], Syed Monowar Hossain*, Emre Ertin[◇], Ashley Kennedy[∧]
David H. Epstein[∧], Kenzie L. Preston[∧], Michelle Jobes[∧], J. Gayle Beck*
Satish Kedia*, Kenneth D. Ward*, Mustafa al'Absi‡, Santosh Kumar*

University of Memphis*   University of South Florida†   Troy University[∨]   NIDA IRP[∧]
The Ohio State University[◇]   University of Minnesota Medical School‡

ABSTRACT

Stress can lead to headaches and fatigue, precipitate addictive behaviors (e.g., smoking, alcohol and drug use), and lead to cardiovascular diseases and cancer. Continuous assessment of stress from sensors can be used for timely delivery of a variety of interventions to reduce or avoid stress. We investigate the feasibility of continuous stress measurement via two field studies using wireless physiological sensors — a four-week study with illicit drug users ($n = 40$), and a one-week study with daily smokers and social drinkers ($n = 30$). We find that 11+ hours/day of usable data can be obtained in a 4-week study. Significant learning effect is observed after the first week and data yield is seen to be increasing over time even in the fourth week. We propose a framework to analyze sensor data yield and find that losses in wireless channel is negligible; the main hurdle in further improving data yield is the attachment constraint. We show the feasibility of measuring stress minutes preceding events of interest and observe the sensor-derived stress to be rising prior to self-reported stress and smoking events.

Categories and Subject Descriptors

H.1.2. [**Models and Principles**]: User/Machine Systems

General Terms

Human Factors; Design; Measurement

Keywords

Mobile Health; Stress; Wireless Physiological Sensor

1. INTRODUCTION

Although low to moderate levels of stress can improve task performance and contribute to skill development [24,50], repeated exposure to acute stress can cause significant damage to physical and mental well-being [32, 45]. Acute stress can lead to headaches, trouble sleeping, and fatigue [6, 10, 12, 33]. Repeated stress can cause or worsen cardiovascular diseases and cancer [44,47]. Stress can also precipitate adverse behaviors, such as depression, rage, anxiety, and addiction [5, 14, 15, 20]. As a result, stress contributes significantly to health care costs [38]. Even for healthy people, stress can degrade quality of life by affecting mood and productivity.

Assessment of stress has traditionally relied on surveys and self-reports. Real-time and continuous measurement of stress in daily life can enhance stress awareness, and revolutionize stress research. It can potentially lead to just-in-time intervention not only for stress management but also to manage other conditions that are affected by stress such as smoking, drinking, drug use, depression, migraine, etc. Real-time measurement of stress is a very active area of research. Researchers have developed a webcam-based method to measure stress in confined work environments [48] and a microphone-based method for use in unconstrained acoustic environment [29]. These approaches, however, do not lead to continuous measurement of stress in daily life. For example, stress from acoustics can only be inferred when people are conversing (25.6% of time [42]).

Physiological monitoring [19,22,40] is a promising approach for continuous assessment of stress. Recent work has demonstrated the feasibility of capturing physiological data in the natural environment with real time wireless transmission to smart-phone[1]. In 2010, [18] reported an average of 23.7 hours per participant of continuous measurement in a mobile health (mHealth) study with 19 participants wearing a custom-made physiological sensor suite for five days. They reported several issues that were major obstacles for continuous physiological measurements — loss of wireless connectivity, incorrect positioning of sensors on the body, and individuals forgetting to fully charge the sensors, or accidentally turning them off. In 2011, [40] reported an average of 24.8 hours per participant of good quality data from 21 participants over two days. Recently, [26] reported 45.6 hours per participant of physiological data collected from 4 participants. Although the amount of data collected per participant has been increasing, it is still not known whether physiological data can be collected in the natural field setting for a longer duration from a larger number of participants.

In this paper, we report our experiences from two user studies (with 40 illicit drug users and 30 daily smokers), where 317 hours of good quality electrocardiogram (ECG) data per participant has been collected over 4 weeks of wearing wireless physiological sensors. We analyze the data collected from these two studies to understand the feasibility and challenges for longer-term continuous

---

[1]We note that real-time wireless transmission of physiological data on smart phone is critical to facilitating just-in-time interventions (JITI) for stress management.

stress assessment in the field. This study reveals several lessons for future efforts on physiological monitoring in daily life. It also provides key implications for future stress research and the design of just-in-time interventions for stress management.

First, we find that 11+ hours/day of usable data can now be obtained in a 4-week study. Second, we observe a significant learning effect after the first week, implying that one week may suffice for most users to learn proper wearing of the physiological sensors by themselves. Third, we observe that data yield continues to increase over time even in the fourth week. This may imply that fatigue stage is not reached in 4 weeks; longer studies may be feasible.

Fourth, we propose a framework to analyze sensor data yield and find that losses in the wireless channel are negligible. Fifth, we find that the yield from respiration sensor is higher than that for an electrocardiogram sensor, perhaps due to a more stringent attachment requirement; we lose at least 1.5 hours of ECG data per day due to attachment issues. It implies that the main hurdle in further improving data yield is the attachment constraint. Finally, we show the feasibility of measuring stress in the minutes preceding events of interest. We observe sensor-derived measure of stress to be rising prior to self-reported stress and smoking events.

Organization. We discuss related works in Section 2, data collection procedure in Section 3, and our framework for data yield analysis in Section 4. Section 5 describes how we infer stress from physiological data. Section 6 presents key results, lessons learned, and implications from our investigations. Section 7 describes limitation and future works, and Section 8 concludes the paper.

2. RELATED WORKS

Several studies have demonstrated the feasibility of physical activity monitoring [13, 41, 49] in daily life using accelerometers. Physiological sensors, however, have more stringent skin attachment requirements. Hence, data yield findings from accelerometer studies are not directly applicable to those involving physiological sensors.

Early systems such as Holter monitors enabled collection of physiological data in the field with a bulky sensor suite [21]. To reduce the burden, sensors were subsequently integrated into clothing in products such as Lifeshirt [51], Smart Vest [37], Life Guard [35], and Smart Shirt [17]. Wireless technology eased the burden further by removing wiring. Wireless integration with mobile phones made it possible to process sensor data in real-time on the mobile phone, and prompt users to fix loose sensor attachments. Systems using such technology include Alive Monitor [1], Zephyr BioHarness [3], and several others (see [9] for a survey). The Alive Monitor includes ECG and pulse oximeter sensing and has been used successfully in several field deployments [25, 28]. The BioHarness system includes ECG, respiration, and skin temperature and has also been used successfully in the field [11]. Despite the recent proliferation of wireless physiological sensors that wirelessly connect to mobile phones, there has been surprisingly little work on the analysis of the resultant data for continuous stress assessment.

Recent work has demonstrated the feasibility of capturing physiological data in the natural environment with real time wireless transmission to smart phones. However, long-term collection of continuous physiological data using wearable wireless physiological sensors in the free-living environment are rare as acknowledged in recent literature [26]. Amount of data collected have ranged from 23.7 hours per participant from 19 participants [18] and 24.8 hours per participant from 21 participants [40], to 45.6 hours per participant from 4 participants [26]. In contrast, we report 317 hours per participant from 40 participants. Moreover, to the best of our knowledge, no existing work has systematically analyzed major factors that are responsible for data loss. Such an analysis is critical to identifying major hurdles in improving the data yield further. Finally, this is the first work to demonstrate the feasibility of stress assessments in the minutes preceding self-reported stress and addictive behaviors.

3. DATA COLLECTION

Two user studies were conducted to investigate the relationships among stress, addictive behaviors, and their mediators (e.g., conversations, physical activity, and location), where these factors were measured via wearable sensors, rather than via traditional self-reports. Studies were conducted on 40 illicit drug users (Study 1) and 30 daily smokers and social drinkers (Study 2). Table 1 presents a summary of the demographics of the study participants. Both studies were approved by Institutional Review Boards (IRBs). In the following, we describe the devices and protocols used.

3.1 Devices and Sensor Measurements

Sensor Suite: During the study period, participants wore the AutoSense sensor suite underneath their clothes [16]. AutoSense consists of an unobtrusive, flexible band worn around the chest. It provides respiration data by measuring the expansion and contraction of the chest via inductive plethysmography (RIP) and includes two-lead electrocardiograph (ECG), 3-axis accelerometer, temperature (ambient and skin), and galvanic skin response (GSR) sensors. Although we used a research platform (e.g., AutoSense) that may not be as comfortable as commercial sensors such as Zephyr BioHarness [3], we believe that our feasibility results should still be applicable to studies that use commercial sensors. We chose AutoSense because it provides a significantly longer lifetime (7 days vs. 3 days). The measurements collected by AutoSense are transmitted wirelessly using ANT radio [2] to an Android smart phone. The sampling rates for the sensors are 64 Hz for ECG, 21.33 Hz for respiration, 10.67 Hz for each accelerometer axis and GSR, and 1 Hz for the two temperature sensors and the battery level. These samples are transmitted at the rate of 28 packets/second, where each packet is 8 bytes long and contains 5 samples.

Mobile Phone: Each participant also carried a smart phone. The smart phone had four roles. First, it received and stored data transmitted by the sensor suite. Second, it sampled and stored data from the sensors built into the phone — GPS and accelerometers. Third, participants used the phone to complete system-initiated self-reports. Finally, participants reported the beginning of drinking and smoking episodes by using a button on the smart phone.

3.2 Field Study Procedure

In both studies, participants were trained in the proper use of the devices. They were shown how to remove the sensors before going to bed and how to put them back on correctly the next morning. They were also asked to take it off during shower and any contact sport. Participants received an overview of the smart phone software's user interface, including the self-report interface. Once the study coordinator felt that participants understood the technology, they left the lab and went about their normal life. Participants were asked to wear the sensors during their waking hours, complete self-reported questionnaires when prompted, and record smoking and drinking events.

Participants were asked to return to the lab daily. The study coordinator downloaded the data collected in the previous day and reviewed the physiological measurements to ensure that sensors were working and were being worn properly. On the final day, participants returned study equipment and completed an Equipment and Experience Questionnaire. Lastly, participants were debriefed on

Statistics		Drug Users	Daily Smokers
# of Participants		40	30
# of Males		29	15
# of Females		11	15
# of Drop Outs		4	2
Age		41 ± 10	24.25 ± 6.25
Race	White	19	19
	African-American	20	8
	Asian	0	3
	Refused	1	0
Educational Status	High School Grad	40	0
	University Grad	0	30
Employment Status	Full Time	15	6
	Part Time	10	14

Table 1: Demographics of participants in the two studies.

their experiences and comfort with the study.

3.3 Study Specific Information

3.3.1 Study 1: Illicit Drug Users

We recruited polydrug users from an ongoing study who agreed to wear AutoSense and complete additional self-reports. Since drug use is a rare event, we choose to conduct this study for four weeks to maximize the likelihood of capturing real-life drug use events.

Compensation: Participants receive $10/day for wearing the AutoSense sensor suite (and $5 bonus for 14+ hours of wearing), carrying the study phone, and completing device-prompted study questionnaires consisting of 32 items. In total, participants are paid up to $380 plus bonus (if any) for four weeks of participation.

Self-Report Measures: Participants were requested to voluntarily record on the smart phone whenever they smoked a cigarette, used any substance (e.g., cocaine, heroin or another opioid, marijuana, benzodiazepines, or alcohol) outside of a medical context, or whenever they felt overwhelmed, anxious, or stressed more than usual. Urine drug screens and retrospective drug use interviews occurred 3 times per week to verify drug use self-reports.

3.3.2 Study 2: Daily Smokers and Social Drinkers

We recruited "daily smokers" and "social drinkers" from the student population at a large university (approximately 23,000 students). We choose one week study duration for each participant to cover all days of a week.

Compensation: Participants earned $35 for daily visits ($5/day); and $75 for completing all end-of-study procedures and returning all equipment. Completing a device prompted self-report questionnaire consisting of 42 items was worth $1. An additional $0.25 bonus was awarded if the questionnaire was completed within five minutes. A maximum of 20 requests for self-reports occurred each day. Thus, a participant could earn up to $20/day ($25 with bonus, if any), adding up to $140 over seven days ($20 x 7). In total, participants are paid up to $250 plus bonus (if any). Since wearing physiological sensors and answering 42-items questionnaire more than 13 times/day are highly burdensome, level of compensation was derived from the prevailing wage in similar behavioral science studies that involves wearing physiological sensors [36].

Self-Report Measures: In addition to completing system-initiated self-reports, participants were requested to voluntarily log each instance of smoking or drinking on the phone using touchscreen.

In this paper, we report data on the feasibility of stress assessment preceding voluntarily reported events (e.g., stress, smoking, drug use, and drinking alcohol).

4. DATA PROCESSING

In this section, we propose a framework (see Figure 1) to identify major factors that may be responsible for data loss when using wireless physiological sensors. We also describe the computational procedure we use to quantify their impact.

Study duration: It is the total number of days that participants were active in the study. In Study 1, four weeks of data was collected from 40 participants, for a total of 922 person days. In Study 2, we collected seven days of data from 30 participants, for a total of 210 person days. We report the **usable** data (yield) and loss in units of hours per person per day.
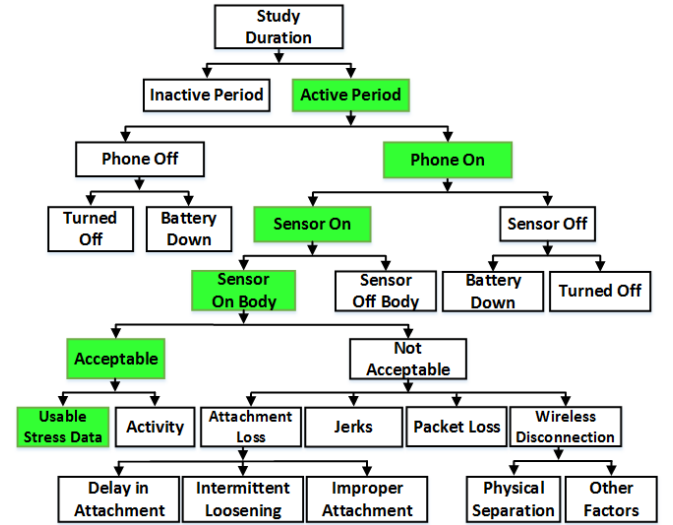

Figure 1: A hierarchical framework for analyzing data loss (and data yield) for physiological data collected in the natural environment. We used this framework to quantify the contribution of various factors in data loss for both studies. We present them in Table 2.

Active/Inactive periods: Because participants were asked to remove the physiological sensors during sleep, the active period per day refers to that part of the day when participants were awake and available for wearing the sensors. We estimate active period to be the period between the first and the last time of the day when acceptable data from any of the physiological sensors (e.g., respiration or ECG) was available. The remaining time of the day (outside the active period) is labeled as inactive period. If participants took the sensors off within the active period (e.g., to take shower), these episodes also contributed to data loss since it occurred within the active period.

Acceptability of ECG and respiration data: For both respiration and ECG, signals are labeled as acceptable if they retain their characteristic morphologies; and unacceptable otherwise. ECG signals are rendered unacceptable mostly due to improper or loose contact of electrodes on body, electrode detachment, loosening of electrical connectors, drying out of gel, or noise from physical movement. Morphology of an acceptable ECG signal corresponds to the standard ECG wave.

Respiration signals are largely affected by misplacement of the chest band and slipping of the band from its expected location on chest. Mere loosening of the chest band sometimes results in a low amplitude signal, but that was considered acceptable if it still retained the characteristic morphology of a respiration signal. Signal saturation to a point where variation is no longer detectable is considered unacceptable. We adopt a method proposed in [39] for

determining acceptability of ECG and respiration signals.

Phone on/off: Within the active period, the period in which the study application was running on the study phone is considered as *phone on*. When the application runs on the phone, it saves phone sensor data even if the body sensors are off or out of radio range. We did not inform participants how to stop the application when the phone was on. But participants could always use the power button on the phone to switch the phone off (which would also stop the application). Time within the active period when the phone was turned off, either intentionally or due to battery drainage, is referred to as *phone off*.

Sensor on/off: *Sensor on* is defined as the period when the study phone receives data from body sensors. *Sensor off* is defined as the period when the study phone is on and the data acquisition application is running, but no data is received from the body sensors for more than one minute. We describe later how we distinguish sensor off from the sensor's being out of wireless range.

Sensor battery down: The wearable sensor suite transmits battery level data. A full charge of our battery is 4.1 volts, nominal operation is 3.7 volts, and the minimum voltage needed for operation is 3 volts. When the battery level is close to 3 volts and the application stops receiving data from the sensors, we define this event as sensor battery down.

Sensor on-body/off-body: Sensor on-body is the time duration when physiological sensors are attached to the participants' body and the phone receives data from the body sensors, even if the data is of poor quality. When the sensors are off body, the data appears saturated (i.e., showing negligible change over time [39]). Because AutoSense has both respiration and ECG sensors, unsaturated data from either of the sensors indicates that the sensor is on-body. Otherwise, it is considered off-body if that time period is within the active duration.

Attachment loss: Attachment loss refers to times when data quality was unacceptable despite the sensors' being attached to the body. It is attributed to three factors:

Delay in attachment occurs when the sensors are being worn but acceptable data from one of the sensors is delayed. Whenever the participants wore the sensors, they were also instructed to visualize the real time signal on the smart phone and fix the attachment, if the signal looked unacceptable.

Intermittent loosening occurs when, after being acceptable for some time, data quality becomes unacceptable intermittently (indicated by restoration of data quality in the same wearing episode). This may be due to movement, ECG electrode gel drying out, or loosening of the electrode attachment or the chest band.

Improper attachment occurs when participants attach sensors improperly and do not fix the attachment for the entire wearing episode.

Loss due to jerks: When data quality becomes unacceptable immediately after the onset of physical activity and again becomes acceptable right after the end of the activity, we define this type of data loss as a loss due to jerks.

Packet loss in the wireless channel: Packet loss (different from disconnection) refers to the time duration when phone is wirelessly connected to the body sensors but some data is lost through the wireless communication channel. Packet loss could occur due to the presence of obstacles between the lines of sight of the devices. We recover lost data using interpolation when the signal retains appropriate morphology even after interpolation. Otherwise, we do not recover them and label these packets as lost packets.

Wireless Connection loss: During the active period, participants were instructed to carry the study phone to ensure that data from the body sensors could be received on the phone in real-time. A green icon (similar to the Wi-Fi icon) was displayed on the application to inform participants about the status of the wireless connection. We logged each disconnection and reconnection time stamp on the phone and use these time stamps to identify data loss due to wireless disconnections, which can result from the following two factors:

Physical separation: Wireless disconnection can result if participants walk away from the phone while wearing the sensors, causing the distance between the phone and sensors to exceed the allowable wireless range. We attribute a connection loss to physical separation, if physical movement (detected from accelerometer that is sampled on both the wearable sensor suite and on the study phone) is observed on the wearable sensors, but not on the study phone, preceding the event of a connection loss.

Other factors: Wireless disconnection can also result from wireless interference or issues with the wireless radio software (on either the sensor or the phone).

Physical Activity Detection: Physical activity episodes need to be detected to exclude data from stress assessment. We adapted a threshold based approach to physical movement detection proposed in [8]. To train our model, we collected training data from seven subjects during walking and running (354.16 minutes), and stationary (1,426.50 minutes) states and then trained a model to distinguish physical movements from stationary states. Participants wore the chest sensors and carried the phone, each of which has a 3-axis accelerometer. The processing of signal includes filtering of raw data and drift removal from the filtered data. Finally, we computed a feature, i.e., standard deviation of magnitude, which is independent of the orientation of the accelerometers [8]. In 10-fold cross-validation test, our model provided 97% accuracy for on-body sensor and 96% for phone in the pocket or purse.

Systematic Filtering of Physical Activity: We did not expect (and instruct) our participants to avoid physical movement during their daily activities as these activities are integral parts of the usual daily life. But, it is well-known that physical activity activates physiology and can easily confound the assessment of stress [40]. It was shown in [22] that the reliability of stress assessment reduces significantly in the presence of physical activity, even in a lab environment. As a result, data affected by physical activity is usually filtered out before applying a stress model on physiological data.
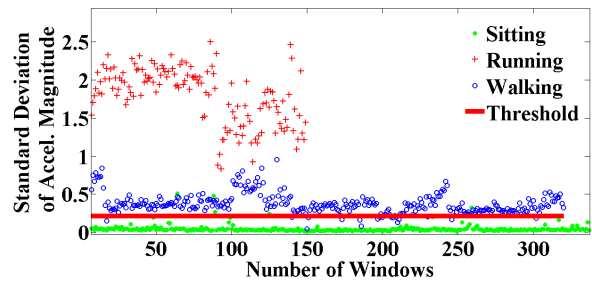

Figure 2: Threshold to detect physical activity. Standard deviation of accelerometers magnitude greater than 0.21348 is labeled as nonstationary (i.e., walking or running) and the others are labeled stationary. We find that daily smoker population was physically active for 16.83% (20% for drug users) when they were wearing our sensors.

We developed an admission control criterion for identifying episodes of physical activity that may confound the assessment of stress to automatically filter out unusable data. Figure 2 shows the threshold for assessment of physical activity. Using this method, we find that 17% of ECG data is affected by activity in daily smokers (Study 2).

5. STRESS INFERENCE FROM PHYSIOL-OGICAL DATA

Stress estimation from ECG has traditionally relied on a single feature (e.g., heart rate or heart rate variability (HRV), respiratory sinus arrhythmia (RSA) [30, 31]. Machine learning models that identify a more specific fingerprint of physiological activation began emerging in the past decade [19]. We use the stress model in [40] since it has been validated in both the lab and the natural field environment. This model provides a continuous measure of stress that is normalized to be between 0 and 1.

This machine learning model predicts whether a one minute measurement corresponds to a physiological response to a stressor. It was trained and tested using physiological data from a 21 person lab study where participants were carefully exposed to three diverse and validated stressors (public speaking, mental arithmetic, and cold pressor challenges) while physiological data and self-reports were collected. This model can classify stress and non-stress minutes with 90% accuracy for 10-fold cross validation (92% accuracy for 66%-34% split), and showed 0.72 correlation with self-reports. Features used in the model included heart rate variability (HRV), respiratory sinus arrhythmia (RSA), minute ventilation, and IE ratio, among several others. To show the generalizability of the model in field, the same 21 participants wore the sensors 12-14 hours daily for 2 days in field and provided stress ratings 25 times/day when prompted. It was found that the average rating produced by the stress model had a correlation of 0.7 with the average rating of self-reported stress [40]. We note, however, that although we use a specific stress model for the analysis reported in this paper, our goal here is to analyze the relative changes in stress level (section 6.4) and hence it is not strongly dependent on a specific stress model.

6. RESULTS AND IMPLICATIONS

6.1 Overall Data Yield and Trends

We use the framework presented in Section 4 (see Figure 1) to report the overall data yield and its characteristics. Table 2 lists the quantity of data collected in the two studies. It also reports data lost due to various factors defined in Section 4. We also report data yield for ECG and respiration separately. Even though these two sensors are hosted on the same sensor device and share the wireless radio, difference in their attachment requirements may explain the difference in yield observed for these two sensors. Finally, the amount of physiological data affected by physical activity is also reported, since stress assessment may not be accurate during these periods. This data set has produced several important observations and results. In the following, we analyze in greater detail some of the main results from which we can draw lessons and suggest implications for stress assessment in the natural environment to support self-monitoring, self-management, or broadly health research.

6.2 Key Observations for Data Collection with Wireless Physiological Sensors

6.2.1 Feasibility of Long-term Data Collection

We first observe from Table 2 that participants in Study 2 wore the sensors for 10.45 hours, of which we get 7.87 hours of acceptable data on ECG[2], which is 75.3% yield. In Study 1, the yield is 85.7% (i.e., 11.33 hours of acceptable ECG data out of 13.22 hours of sensors on-body). This represents a significant higher yield in

[2]We get an higher yield on respiration, but analyze ECG since it has the lowest yield of all physiological sensors.

Factors	Study 1	Study 2
Study Length (person days)	922	210
Active period	14.57±2.8	13.02±2.01
Phone off	0.78±0.23	2.2±0.31
Phone on	13.73±2.2	10.82±1.6
Sensor off	0.17±0.06	0.37±0.04
Sensor Battery Down	0.03±0.32	0.08±0.88
Sensor off Body	0.34±1.31	0.11±0.19
Sensor on Body	13.22±1.86	10.45±1.55
Packet Loss	0.27±0.01	0.18±0.04
Wireless Disconnection	0.04±0.01	0.15±0.01
ECG		
Delay in Attachment	0.22±0.14	0.3±0.10
Intermittent Loosening	1.17±0.09	1.8±0.04
Improper Attachment	0.19±1.72	0.13±0.71
Acceptable Data	11.33±0.88	7.87±0.31
Activity	2.26±0.18	1.32±0.06
Usable	9.06±0.72	6.55±0.26
RIP		
Delay in Attachment	0.12±0.1	0.05±0.06
Intermittent Loosening	0.72±0.05	0.17±0.01
Improper Attachment	0.20±0.63	0.1±0.51
Acceptable Data	11.84±0.52	9.83±0.55
Activity	2.36±0.11	1.66±0.10
Usable	9.47±0.42	8.18±0.47

Table 2: Overall physiological data yield statistics computed from all person days of both studies using the hierarchical framework proposed in Figure 1. Values (mean±std) are in hours per participant per day. To show the subject variation, standard deviation is added with each mean value.

Study 1. To understand this difference, we examine data yield from Study 1 for each study week individually(as opposed to considering the average of all four weeks).

Filled bars in Figure 3 present an average yield in each of the four weeks of Study 1, and unfilled bar is the average yield in Week 1 of Study 2. We observe that the (78.9%) yield in Week 1 of Study 1 is comparable to the (75.3%) yield in Week 1 of Study 2. Statistically significant improvement (from 78.9% to 84.3%) in data yield occurs in Week 2. We thus hypothesize that one week of participation may suffice to learn how to wear and maintain sensors well. We summarize this finding below:

OBSERVATION 1. *For physiological sensor wearing, a significant learning effect is observed after Week 1.*

We further observe that the improvement, though not statistically significant, continues through all four weeks. This points to potential feasibility of studies longer than four weeks that is reported here. As mentioned in Section 1, the longest wearing episode for physiological sensors reported previously was 45.6 hours / participant. In Study 1, the yield is seen to be increasing (beyond 85%) even after each participant wore the physiological sensors for 317 hours. We, therefore, propose the following hypothesis.

HYPOTHESIS 1. *It may be feasible to obtain a good data yield (e.g., 11+ hours/day) from wireless physiological sensors, even when they are worn for four weeks or longer.*

In the following, we discuss major factors that may explain or improve data yield with wireless physiological sensors.
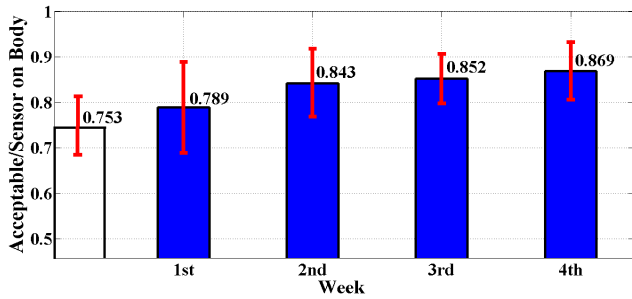
Figure 3: Data yield (the ratio of *acceptable* to *sensor on-body duration*) for Week 1 in Study 2 (unfilled bar) and for all four weeks in Study 1 (filled bars). On average, data yield increased over time. The increase observed from first week to second week in Study 1 is statistically significant (paired t-test, $p = 0.035$).
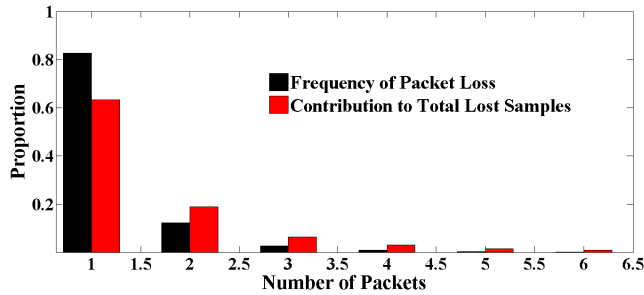

Figure 4: Number of consecutive packets lost between two successive packet arrivals with their relative contribution to overall lost samples. Note that each packet contains only five samples.

6.2.2 Impact of Short Packets on Wireless Data Loss

There have been significant improvements in the design of low-power wireless radios in recent years, making them a mature technology. For example, studies conducted few years ago [18] reported 50% data loss due to various issues, including that due to wireless radio. Similarly, [40] reported 30% losses. It was also reported in [40] that the average length of consecutive valid data was less than 4 minutes. The sensor suite in our study used ANT radio, which is similar to the newly emerging low power Bluetooth radio. In addition to various improvements in antenna design that are better suited to the body, and higher energy efficiency, we discuss a specific design decision that decreased data loss in wireless transmission in our studies.

Shorter packets: Figure 4 shows number of packets that were lost between two successive packet arrivals. We observe that 82% of the time, we lose only one packet in a burst and 10% of the time we lose two packets subsequently. Figure 4 also depicts the contribution of packet loss to the overall data. We note that the use of short packets is quite advantageous for physiological sensors because each packet then constitutes only a small portion of ECG or respiration cycle. Interpolating one, two, and three packets reduces the overall data loss rate due to packet losses from 9.5% to 1.6%, 0.51%, and 0.25% respectively. We only interpolated one packet that constitutes 8% of a cycle for both ECG and respiration using Hermite interpolation (which did not adversely affect signal morphology). As shown in Table 2, data lost due to packet losses or wireless disconnection is less than 2.5% each, which is a significant improvement over the previously reported studies.

We summarize this observation in the following.

OBSERVATION 2. *Using short packets can limit data loss in wireless transmissions.*

6.2.3 Impact of Attachment Burden

A major hurdle in collecting physiological data in daily life over long term is the *stringent attachment* requirement for these sensors. We observe (Table 2) that in Study 2, we lose only 0.32 hours/day of respiration sensor data as compared to 2.23 hours/day for ECG sensor data. In Study 1 (Drug Users), we lose 1.04 hours/day of respiration data as compared to 1.58 hours/day of ECG data[3]. In both studies the difference in the yield of two sensors can be attributed to attachment losses. ECG is sensitive to proper attachment of the gel electrodes to skin at an appropriate location, but for respiration sensor, it suffices to wear a band around mid-chest. Respiration sensor does not require skin contact and it can be worn with flexible positioning of the band from the upper chest position to the abdomen at the level of umbilicus [4] - making it more challenging to wear and maintain ECG electrodes properly than the respiration band. However, even respiration sensor band can become loose and slip over the course of wearing for the entire day.

The major component of data loss from attachment constraints is the one lost due to intermittent loosening. We analyzed the pattern of this loss as the day progresses. We find that intermittent loosening increases as the day progresses ($r = 0.8, p < 0.001$) (see Figure 5). We believe that eliminating or reducing the attachment constraints for physiological sensing may significantly improve data yield in the field environment. We note, however, that our findings are specific to the ECG sensor used in our studies. Newer ECG sensors such as patches or smart fabric electrodes may have lower attachment issues, which can be investigated in future studies. We summarize this effect in the following observation.

OBSERVATION 3. *Attachment constraints is a major source of data loss during everyday wearing of physiological sensors that involve careful attachment.*
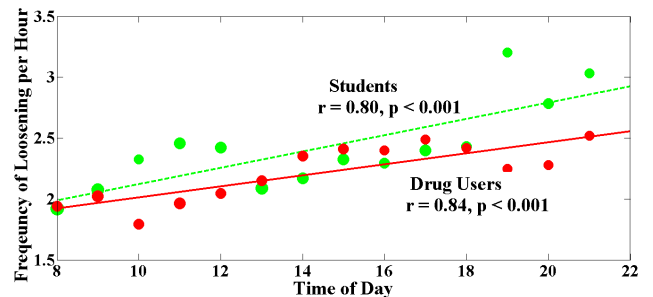

Figure 5: Frequency of intermittent loosening throughout the day. Red bold line is the least square regression for the incidents of intermittent loosening (ECG) for drug users (Study 1) ($r = 0.84, p < 0.001$). Blue dashed line is the regression line for daily smokers (Study 2) ($r = 0.8, p < 0.001$). It shows that intermittent loosening increases with the progression of the time of day for both population.

6.3 Implications for Stress-related Research

In this section, we discuss implications of this study on future prospects of continuous stress assessment in the natural environment for stress research. It has also implications on self-monitoring of stress because these types of systems have potential in assisting

---

[3]As explained earlier, the higher overall yield in ECG occurs in Study 1 due to learning effect over four weeks.

people in monitoring their health conditions. Although wearing the chest band is burdensome for daily use, some users may be motivated, perhaps due to a health condition (e.g., stress can aggravate migraine attack), to use such systems even when there is no direct monetary incentive. Wearable physiological sensors are also becoming popular in the quantified-self community [43].

Here, we analyze how much overall physiological data is available daily for stress assessment, especially around events (or behaviors) of interest. We consider five events of interest, each of which were self-reported by our participants as part of the study protocol. In Study 1, the drug users reported smoking, drug use, craving, and stress events while in Study 2, daily smokers and social drinkers reported each smoking and drinking event. Table 3 shows overall statistics of all types of voluntary self-reports collected.

	Events	# Self-reports	Self-report/day
Study 1	Smoking (Cigarette)	2643	2.87
	Craving for drug	302	0.33
	Illicit Drug Use	142	0.15
	Perceived Stress	108	0.12
Study 2	Smoking (Cigarette)	1520	7.24
	Drinking (Alcohol)	144	0.68

Table 3: Statistics of voluntary self-reports in the field. Smoking self-reports are frequent. Craving, drug use, drinking alcohol, and stress reports are less frequent.

We first analyze the availability of stress measurements preceding events of interest since stress has long been known to play a prominent role in precipitating several adverse behaviors, especially smoking. We then analyze how likely participants are to wear the sensors when such events of interest occur.

6.3.1 Stress Assessment Preceding Events of Interest

Given that smoking is the largest cause of death [34], over $300 million are spent on smoking cessation research by US National Institutes of Health alone. Decades of work in this area, all based on self-reported stress data, found stress rising hours prior to a lapse in addictive behaviors [46]. However, due to a lack of continuous stress measurement, it has not been known what happens minutes prior to a smoking lapse. Although our study did not assess stress prior to a smoking lapse in abstinent smokers, it shows the feasibility of obtaining stress data in minutes prior to a smoking event.

Figure 6 shows the availability of stress assessment in minutes preceding to a self-reported smoking, drinking, and drug craving events. Data for self-reported stress and drug use are similar and hence are omitted for brevity. We report these data in terms of percentage of time (out of the active period per day). This helps normalize these data across the two populations since the absolute numbers are quite different in the two studies.

We observe that it is indeed feasible to obtain stress measurement in each minute preceding an event of interest via self-reports. Up to 65% of the data is available for stress assessment in the neighborhood of the event even after excluding data affected by physical activity. Availability of stress measurement data arbitrarily close to the self-report opens up tremendous new opportunities to not only study the role of stress in these adverse events with significant health impact, but also opens up an opportunity to deliver interventions upon detection of a rise in stress in real time (since stress data are available on mobile phone).

We make several observations from our data. First, data availability in the minute preceding a self-report is higher than usual. This may be due to enhanced consciousness of participants to check sensor attachments prior to reporting an event. Second, we observe

that there is a higher level of activity preceding a smoking event than usual. This may be due to walking out of buildings prior to smoking. We still obtain over 55% usable data in each population. Third, we observe that if the analysis is flexible in using best 3 out of 5 preceding minutes, then the amount of data available for stress increases to around 70%. We now summarize these observations.

OBSERVATION 4. *It is now feasible to assess stress in the minutes preceding self-reported addictive behavior events.*

6.3.2 Likelihood of Capturing Events of Interest

For self-reporting stress and drug use, participants used another digital device that they were asked to carry with them at all times. We use these two reports to estimate how likely the participants are to be wearing the sensors (to enable stress assessment) when these events of interest occur. We observe that on days when participants reported a stress event, the average time of sensor wearing is 11.96 hours/day as compared to 13.32 hours/day for those days when these same participants did not report a stress event (two tailed paired t-test, $p < 0.001$). Similarly, on the days when participants reported a drug use event, the average time of wearing was 12.07 hours/day versus 13.43 hours from these same participants on non-drug days (two tailed paired t-test, $p < 0.001$). We use these observations to hypothesize that on days or times when participants engage in unusual behavior (anticipated stress such as job interviews or planned drug use), they are not as likely to wear the sensors. Yet, there are sufficient instances of these events when it is feasible to capture physiological data for stress assessment.

OBSERVATION 5. *Participants wear fewer hours on days when stress or drug events are reported.*

6.4 Implications for Stress-related Just-in-time Intervention

Now that it is feasible to obtain stress assessment in real-time on mobile phone preceding events of interest, it has several implications for the design and delivery of just-in-time interventions (JITI). In this section, we first present the pattern of stress observed preceding events of interest and draw implications on how they may inform the design of JITIs. We then discuss how the measurement of other contexts from sensors can be used in the design or delivery of JITIs.

Stress Pattern: For smoking, stress, craving, and drinking, we use the self-report from participants as the timing of these events. But, for drug use, self-report is usually provided tens of minutes after the use episode. Hence, self-report is not a reliable marker of the timing of these events. Hence, we use a model we recently developed to detect the timing of illicit drug use from physiology [23].

Figure 7 shows the stress patterns observed preceding the events of interest. We make several observations. First, we observe that stress is seen to be rising in the minutes preceding a self-report of stress. It also confirms that the stress model used here is sensitive enough to capture the rise in self-reported stress. Second, we observe that even though this study did not involve abstinent smokers, stress levels preceding a smoking event is still higher than usual[4]. Further, the stress level is elevated several minutes preceding a smoking report. This is observed in both populations, which may warrant further investigation by smoking researchers to understand and explain the pattern. Stress level preceding other events do not show such clear patterns. We summarize these observations in the following.

---

[4]We note that activity affected data is not considered for stress assessment and hence the rise on stress level is not a result of walking episodes before a smoking event.
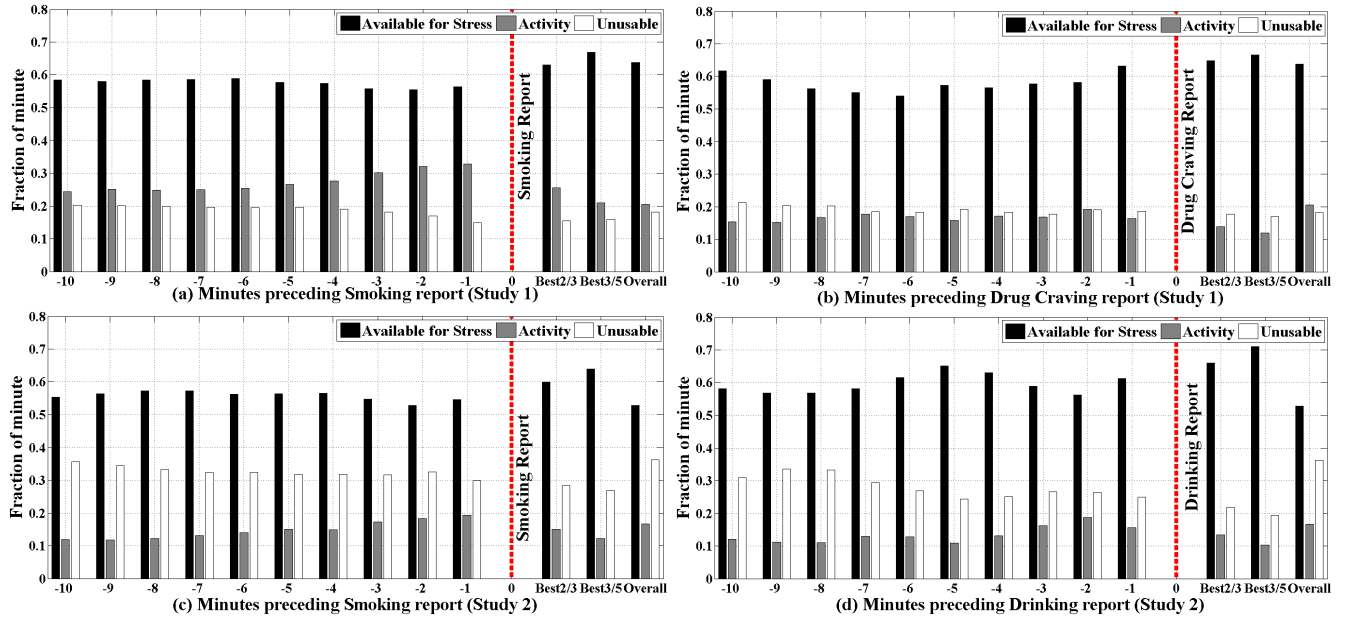
Figure 6: In each group of 3 bars, 1st bar shows the fraction of physiological data available for stress assessment, 2nd bar shows what fraction of data is affected by activity, and 3rd bar shows what fraction becomes unacceptable due to poor data quality. Vertical dashed line indicates the relative time of self-reports. Bars before the vertical dashed lines present the availability of data for stress assessment in preceding individual minutes -1 through -10. After the vertical line, best 2/3 means taking the average availability of two best previous minutes out of three, and best 3/5 means taking the average availability of three best previous minutes out of five. Last group of bars in each figure represents overall minute by minute availability of data for stress assessment. (a), (b) are for drug users ($n = 40$) and (c), (d) are for daily smokers ($n = 30$).

OBSERVATION 6. *Elevated stress from sensors is observed prior to a self-reported smoking or stress event, which can inform the design and delivery of just-in-time interventions.*

Third, we observe a higher stress level among drug users than student smokers. This may be due to lower socio-economic status of drug users, as compared to students. Fourth, within each population, stress is also highly variable across participants. There is also wide within-person variability which is omitted for clarity of presentation. Finally, as the vertical bars show in these figures, there is wide variability in the stress level. These observations suggest that stress provides important information for triggering an intervention to avert an adverse behavior. But, stress measurement may not be sufficient by itself due to wide variability and additional strategies should be employed to make such interventions more accurate. They could include personalization of the models as well as use of other contexts.

Incorporating other contexts to trigger just-in-time intervention: To explore additional context that may be detectable from sensors and can help predict an imminent adverse behavior, we analyze the physical activity levels. In the case of smoking events, activity pattern from both studies show a common pattern (Figure 6(a) and (c)). Physical activity around the smoking self-reports for both the cases is higher at times close to the self-reports. This is because smokers usually smoke cigarette in a designated area. They walk to or from the smoking spot around smoking session. Therefore, it may be possible to use the detection of walking outside a building to improve the prediction of a smoking event which may then be used to trigger an intervention to avert a smoking lapse event.

7. LIMITATIONS AND FUTURE WORK

This work presents several findings that have implications for future monitoring of stress in research as well as for personal health and well-being. However, our work has several limitations that could be addressed in future works.

Generalizability: Both studies were monitored or supervised by professional staff. Participants had a training session where they learnt about proper wearing of devices, which improved the data yield over time (see Figure 3). It remains an open question whether similar or better data yield can be obtained without daily meetings, especially in the absence of micro-incentives to encourage compliance with the protocol, such as when users voluntarily wear sensors for self-monitoring. Nevertheless, this study provides a first evidence of high data yield in a 4-week study.

Sensor Type: This study considered two physiological sensors with specific attachment requirements. Again, it remains an open question whether similar or better yield can be obtained for other sensors that may have different attachment requirements (e.g., smart patch, smart watch).

Temporal precision of self-report: Self-reports and actual events may not always be synchronized. Participants may report an event before, during, or after the actual occurrence. Hence, the stress levels observed preceding self-reported events may not be temporally precise. Automated detection of these events from sensors (e.g., traffic stressors from GPS, exam from calendar) will improve the temporal precision of relating stress to the event of interest. These methods are an active area of research and promise a future where addictive behaviors such as smoking [7], drug use [23], and drinking [27] may be detected accurately from sensors.

8. CONCLUSION

This work demonstrates the feasibility of collecting continuous stress data from wireless physiological sensors in the natural environment for 4 weeks. Stress research that have been traditionally limited to self-reports, can now be conducted with sensors and
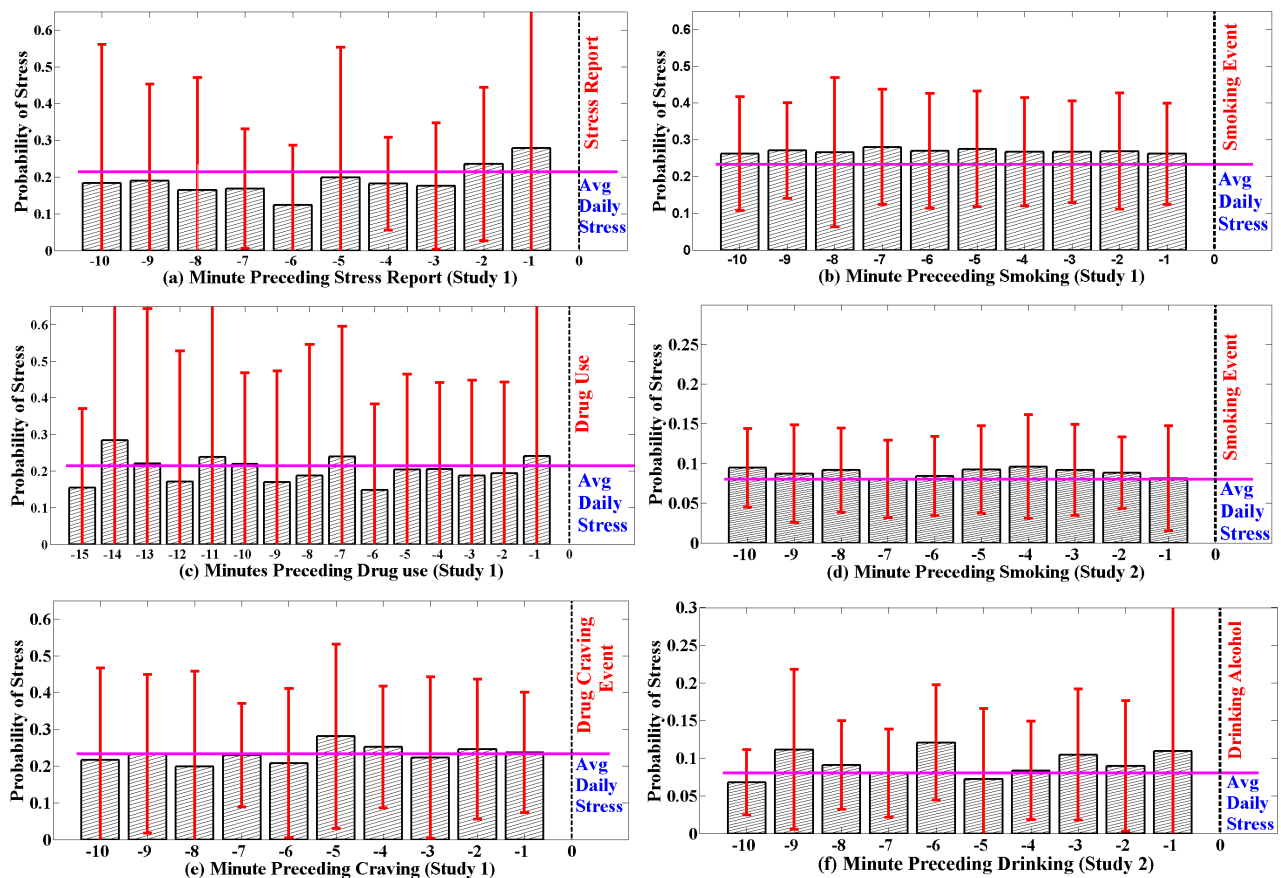
Figure 7: Stress level before self-reported stress (a), smoking (b), drug use (c), and drug craving (e) events in Study 1. Stress level before self-reported smoking (d) and drinking (f) events in Study 2. Vertical dashed line shows the time of the event report, and horizontal bold line shows average daily stress level. Figure (a) shows that the stress level in minutes preceding a voluntary self-report of stress event is higher than average daily stress. Also, stress level in minutes preceding smoking is relatively higher than average daily stress.

provide tremendous visibility into stress dynamics around events of interest. Similarly, interventions that have been limited to self-initiation or pre-selected-time-based can now be triggered based on real-time sensor data. Primary hurdle to further increasing the data yield as well as duration of sensor wearing to several months is the attachment constraints and burden associated with physiological sensing. Development of contact-less sensors may enable the next giant leap in physiological sensing.

9. ACKNOWLEDGMENTS

10. REFERENCES

[1] Alive Technologies. **http://www.alivetec.com/**, Accessed: September 2013.

[2] ANT Radio. **http://www.thisisant.com/**, Accessed: September 2013.

[3] Zephyr Bioharness. **http://www.zephyr-technology.com/bioharness-bt**, Accessed: September 2013.

[4] J. A. Adams. Respiratory inductive plethysmography. *Infant respiratory function testing. New York: Wiley-Liss*, pages 139–64, 1996.

[5] M. Al'Absi. *Stress and addiction: Biological and psychological mechanisms*. Academic Press, 2011.

[6] M. Al'Absi and D. Arnett. Adrenocortical responses to psychological stress and risk for hypertension. *Biomedicine & pharmacotherapy*, 54(5):234–244, 2000.

[7] A. A. Ali, S. M. Hossain, K. Hovsepian, M. M. Rahman, K. Plarre, and S. Kumar. mpuff: automated detection of cigarette smoking puffs from respiration measurements. In *ACM/IEEE IPSN*, 2012.

[8] L. Atallah, B. Lo, R. King, and G.-Z. Yang. Sensor placement for activity detection using wearable accelerometers. In *IEEE BSN*, pages 24–29, 2010.

[9] M. Chan, D. Estève, J.-Y. Fourniols, C. Escriba, and E. Campo. Smart wearable systems: Current status and future challenges. *Artificial intelligence in medicine*, 56(3):137–156, 2012.

[10] G. P. Chrousos and P. W. Gold. The concepts of stress and stress system disorders: overview of physical and behavioral homeostasis. *JAMA*, 267(9):1244–1252, 1992.

[11] B. Cinaz, R. La Marca, B. Arnrich, and G. Tröster. Towards continuous monitoring of mental workload. In *ACM UbiComp*, 2010.

[12] S. E. Cohen, R. C. Kessler, and L. U. E. Gordon. *Measuring stress: a guide for health and social scientists*. Oxford University Press, 1995.

[13] S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, et al. Activity sensing in the wild: a field trial of ubifit garden. In *ACM SIGCHI*, 2008.

[14] M. A. Enoch. Pharmacogenomics of alcohol response and addiction. *American Journal of Pharmacogenomics*, 3(4):217–232, 2003.

[15] M. A. Enoch. Genetic and environmental influences on the development of alcoholism. *Annals of the New York Academy of Sciences*, 1094(1):193–201, 2006.

[16] E. Ertin, N. Stohs, S. Kumar, A. Raij, M. al'Absi, T.Kwon, S. Mitra, S. Shah, , and J. Jeong. AutoSense: Unobtrusively Wearable Sensor Suite for Inferencing of Onset, Causality, and Consequences of Stress in the Field. In *ACM SenSys*, 2011.

[17] C. Gopalsamy, S. Park, R. Rajamanickam, and S. Jayaraman. The wearable motherboard: The first generation of adaptive and responsive textile structures (arts) for medical applications. *Virtual Reality*, 4(3):152–168, 1999.

[18] J. Healey, L. Nachman, S. Subramanian, J. Shahabdeen, and M. Morris. Out of the lab and into the fray: Towards modeling emotion in everyday life. *Pervasive Computing*, pages 156–173, 2010.

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

[20] J. P. Henry. *Stress, neuroendocrine patterns, and emotional response.* John Wiley and Sons, 1990.

[21] N. Holter and J. Generelli. Remote recording of physiological data by radio. *Rocky Mountain medical journal*, 46(9):747, 1949.

[22] J. H. Hong, J. Ramos, and A. K. Dey. Understanding physiological responses to stressors during physical activity. In *ACM UbiComp*, 2012.

[23] S. M. Hossain, A. A. Ali, M. M. Rahman, E. Ertin, D. Epstein, A. Kennedy, K. Preston, A. Umbricht, Y. Chen, and S. Kumar. Identifying drug (cocaine) intake events from acute physiological response in the presence of free-living physical activity. In *ACM/IEEE IPSN*, 2014.

[24] A. K. Johnson and E. A. Anderson. Stress and arousal. 1990.

[25] V. Jones, V. Gay, and P. Leijdekkers. Body sensor networks for mobile health monitoring: Experience in europe and australia. In *IEEE ICDS*, pages 204–209, 2010.

[26] M. Kusserow, O. Amft, and G. Troster. Modeling arousal phases in daily living using wearable sensors. *IEEE Transactions on Affective Computing*, 4(1):93–105, 2013.

[27] T. R. Leffingwell, N. J. Cooney, J. G. Murphy, S. Luczak, G. Rosen, D. M. Dougherty, and N. P. Barnett. Continuous objective monitoring of alcohol use: Twenty-first century measurement using transdermal sensors. *Alcoholism: Clinical and Experimental Research*, 37(1):16–22, 2013.

[28] P. Leijdekkers, V. Gay, and E. Barin. Feasibility study of a non-invasive cardiac rhythm management system. *International Journal of Assistive Robotics and Systems*, 10:5–14, 2010.

[29] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *ACM UbiComp*, 2012.

[30] B. McEwen. Stress, adaptation, and disease: Allostasis and allostatic load. *Annals of the New York Academy of Sciences*, 840(1):33–44, 2006.

[31] B. McEwen. Physiology and neurobiology of stress and adaptation: Central role of the brain. *Physiological Reviews*, 87(3):873–904, 2007.

[32] B. S. McEwen. Protection and damage from acute and chronic stress: allostasis and allostatic overload and relevance to the pathophysiology of psychiatric disorders. *Annals of the New York Academy of Sciences*, 1032(1):1–7, 2004.

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

[34] A. M. Miniño, S. L. Murphy, J. Xu, and K. D. Kochanek. Deaths: final data for 2008. *National vital statistics reports: from the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System*, 59(10):1–126, 2011.

[35] C. Mundt, K. Montgomery, U. Udoh, V. Barker, G. Thonier, A. Tellier, R. Ricks, R. Darling, Y. Cagle, N. Cabrol, et al. A multiparameter wearable physiologic monitoring system for space

and terrestrial applications. *IEEE Transactions on Information Technology in Biomedicine*, 9(3):382–391, 2005.

[36] M. Mustang, A. Raij, D. Ganesan, S. Kumar, and S. Shiffman. Exploring micro-incentive strategies for participant compensation in high burden studies. In *ACM UbiComp*, 2011.

[37] P. Pandian, K. Mohanavelu, K. Safeer, T. Kotresh, D. Shakunthala, P. Gopal, and V. Padaki. Smart vest: Wearable multi-parameter remote physiological monitoring system. *Medical engineering & physics*, 30(4):466–477, 2008.

[38] A. Perkins. Saving money by reducing stress. *Harvard Business Review*, 72(6):12, 1994.

[39] K. Plarre, A. Raij, S. Guha, and S. Kumar. Automated detection of sensor detachments for physiological sensors in the wild. In *ACM Wireless Health*, 2010.

[40] K. Plarre, A. Raij, M. Hossain, A. Ali, M. Nakajima, M. al'Absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, D. Siewiorek, A. Smailagic, and L. Wittmers. Continuous Inference of Psychological Stress from Sensory Measurements Collected in the Natural Environment. In *ACM/IEEE IPSN*, 2011.

[41] R. A. Rabinovich, Z. Louvaris, Y. Raste, D. Langer, H. Van Remoortel, S. Giavedoni, Burtin, et al. Validity of physical activity monitors during daily life in patients with copd. *European Respiratory Journal*, 42(5):1205–1215, 2013.

[42] M. M. Rahman, A. A. Ali, K. Plarre, M. al'Absi, E. Ertin, and S. Kumar. mConverse: Inferring Conversation Episodes from Respiratory Measurements Collected in the Field. In *ACM Wireless Health*, 2011.

[43] J. Rooksby, M. Rost, A. Morrison, and M. Chalmers. Personal tracking as lived informatics. In *ACM SIGCHI*, 2014.

[44] R. Rosmond and P. Björntorp. Endocrine and metabolic aberrations in men with abdominal obesity in relation to anxio-depressive infirmity. *Metabolism*, 47(10):1187–1193, 1998.

[45] R. Rosmond, M. F. Dallman, and P. BjӧĹrntorp. Stress-related cortisol secretion in men: Relationships with abdominal obesity and endocrine, metabolic and hemodynamic abnormalities 1. *Journal of Clinical Endocrinology & Metabolism*, 83(6):1853–1859, 1998.

[46] S. Shiffman, J. A. Paty, M. Gnys, J. A. Kassel, and M. Hickcox. First lapses to smoking: within-subjects analysis of real-time reports. *Journal of consulting and clinical psychology*, 64(2):366, 1996.

[47] A. Steptoe, G. Fieldman, O. Evans, and L. Perry. Cardiovascular risk and responsivity to mental stress: the influence of age, gender and risk factors. *Journal of cardiovascular risk*, 3(1):83–93, 1996.

[48] C. S. S. Tan, J. Schöning, and K. L. K. Coninx. Investigating the effects of using biofeedback as visual stress indicator during video-mediated collaboration. In *ACM CHI*, 2014.

[49] K. Taraldsen, S. F. Chastin, I. I. Riphagen, B. Vereijken, and J. L. Helbostad. Physical activity monitoring by use of accelerometer-based body-worn sensors in older adults: a systematic literature review of current knowledge and applications. *Maturitas*, 71(1):13–19, 2012.

[50] H. Ursin and R. Murison. Classification and description of stress. *Neuroendocrinology and psychiatric disorder*, pages 123–132, 1984.

[51] F. Wilhelm, W. Roth, and M. Sackner. The lifeshirt: an advanced system for ambulatory measurement of respiratory and cardiac function. *Behavior Modification*, 27(5):671–691, 2003.