Dynamic models of stress-smoking responses based on high-frequency sensor data

Sahar Hojjatinia¹, Elyse R. Daly², Timothy Hnat³, Syed Monowar Hossain³, Santosh Kumar³,
Constantino M. Lagoa¹, Inbal Nahum-Shani⁴, Shahin Alan Samiei³, Bonnie Spring², & David E. Conroy^{2,5}

¹ School of Electrical Engineering and Computer Science, The Pennsylvania State University

² Department of Preventive Medicine, Northwestern University Feinberg School of Medicine

³ Department of Computer Science, University of Memphis

⁴ Institute for Social Research, University of Michigan

⁵ Department of Kinesiology, The Pennsylvania State University

Declaration of Competing Interests: None

Word Count: 5838

Corresponding author: David E. Conroy, Ph.D., 266 Rec Hall, The Pennsylvania State University,
University Park, PA 16802 USA. Email: conroy@psu.edu

Abstract

Background and Aims: Self-reports indicate that stress increases risk for smoking; however, intensive data from sensors can provide a more nuanced understanding of stress in the moments leading up to and following smoking events. Identifying personalized dynamical models of stress-smoking responses can improve characterizations of smoking responses following stress; but techniques used to identify these models require intensive longitudinal data. This study leveraged advances in wearable sensing technology and digital markers of stress and smoking to identify person-specific models of stress and smoking system dynamics by considering stress immediately before, during, and after smoking episodes. **Design:** Observational study. **Methods:** Adult smokers (n=45) wore the AutoSense chestband (respiration-inductive plethysmograph, electrocardiogram, accelerometer) with MotionSense (accelerometers, gyroscopes) on each wrist for three days prior to a quit attempt. The odds of minute-level smoking events were regressed on minute-level stress probabilities to identify person-specific dynamic models of smoking responses to stress. **Findings:** Simulated pulse responses to a continuous stress episode revealed a consistent pattern of increased odds of smoking either shortly after the beginning of the simulated stress episode or with a delay, for all participants. This pattern is followed by a dramatic reduction in the probability of smoking thereafter, for about half of the participants (49%). **Conclusions:** Sensor-detected stress probabilities indicate a vulnerability for smoking that may be used as a tailoring variable for just-in-time interventions to support quit attempts.

Keywords: System Dynamics Analysis, Mobile Health, Patient-Specific Modeling, Smoking Urge, Remote Sensing Technology, System Identification.

Dynamic models of stress-smoking responses based on high-frequency sensor data

Cigarette smoking is the leading preventable cause of death in the United States, responsible for about one in five deaths annually (1). The costs of direct medical care for adult smokers, lost economic productivity caused by smoking-related disability and premature mortality, and secondhand smoking exposure exceed \$300 billion per year in the United States (2,3). Comprehensive smoking control programs require a clear understanding of the dynamics of smoking triggers and behaviors; however, few studies have attempted to describe those systems. This study used high-frequency sensors and novel markers of both stress and smoking to describe system dynamics and then simulate how stress influences subsequent risk for smoking.

Stress Influences Smoking

Stress and negative emotions have long been thought to be motives for smoking (4–7). Stress represents the non-specific response of the body to a demand (8). It manifests as distress and negative affect when the perceived demand exceeds the individual’s coping potential (9). Smokers identify stress as a reason for smoking (4–6). One reason may be that they learn that nicotine self-administration can dispel negative affective states that are caused by withdrawal and produce acute psychological benefits such as improved mood and concentration (10–12).

Evidence based on a variety of methods supports the concept that stress-induced negative affect is positively associated with smoking (13–18). For example, the acute stress of a terrorist attack in New York City increased smoking rates among residents (19). On a faster time scale, Shiffman et al. found a positive association between repeated assessments of daily stress and smoking (14). Affect dysregulation, particularly negative mood variation, has also emerged as a risk factor for future smoking escalation (15). These findings point to stress responses as a target

for behavioral interventions to reduce smoking and as a potential tailoring variable for interventions to reduce stress-related smoking.

The stress and smoking literature has three gaps that need to be addressed to inform intervention development efforts. First, most of the literature assessing links between stress or negative affect, on one hand, and smoking on the other, involves data collected during a quit attempt (20–24). Assessments of stress taken after quitting smoking reflect both physiological arousal due to daily life experiences and distress due to nicotine withdrawal. Stress-smoking dynamics may be able to be characterized more consistently prior to a quit attempt, before prolonged nicotine withdrawal disrupts baseline stress – smoking dynamics. These dynamics can expose the duration of vulnerability between the onset of a stress response and smoking behavior. That information can be used to tailor decision rules (i.e. person-specific algorithms) for adaptive smoking cessation interventions in the future.

Second, available analyses have relied on periodic measurement of stress and smoking based on self-reports. However, these self-report assessments are made too infrequently to capture sufficient granularity in continuous stress dynamics leading to and following smoking events. Additionally, self-reports are vulnerable to bias from memory errors, psychiatric limitations and self-presentational concerns (13,25). The present study uses sensor-based markers of both stress and smoking to obtain more intensive data that can reveal the dynamics of stress and smoking.

Third, no attempts have been made to study the dynamics of stress in the moments immediately prior to and following smoking events. To date, stress has not been collected continuously and data have typically been limited to a single data point about stress prior to a smoking episode. The availability of sensors and digital markers of stress creates new

possibilities for monitoring stress continuously so a time-series of stress measurements can be used to predict smoking events and to characterize changes in stress after smoking. In this study, intensive longitudinal digital assessments afforded sufficiently continuous, granular portrayal of stress dynamics to develop a model informed by stress immediately before, during, and after smoking episodes. This approach can provide the foundation for predicting the temporal response of individuals to stress as a basis for developing evidence-based, just-in-time smoking cessation interventions.

mHealth Tools May Reveal Dynamics of Self-Medication Smoking Effect

Recent advances in wearable sensors and mHealth markers enable continuous monitoring and detection of stress states and smoking behavior. Useful measurements may include electrical activity of the heart, breathing dynamics, hand movements, and physical activity. Continuously-measured stress states (i.e., inferred from physiological arousal in the absence of physical activity) and smoking behavior sampled at high frequencies (e.g., minute-level) via wearable devices require different tools for their analysis than more sparsely-distributed periodic data from ecological momentary assessments to characterize system dynamics. Understanding these dynamics will provide a foundation for adaptive treatments on the individual level so interventions can be delivered at the moments of greatest need. High-frequency data indicators of an outcome (e.g., minute-level smoking behavior) are not only dependent on the present moment of the candidate predictor (e.g., stress), but are also correlated with past values of that predictor. In other words, a stress episode can affect both the probability of smoking immediately, and increase the smoking likelihood in the minutes following the stress episode. Therefore, data signifying present moments of smoking need to be connected to data from immediately prior values of stress to model the stress-smoking responses. Dynamical systems provide the necessary

tools to model this so-called “memory effect”: that is, the effect of past and present values of the predictor on the future outcome. In this context, the system is a process that connects past and present values of a predictor (stress) to an outcome (smoking).

Dynamical Systems

Smoking is a complex, dynamic behavior. It can be influenced by a variety of factors over time, and is not necessarily the instantaneous result of stress (or other causes). Given the potential for a delay of unknown duration between stress and smoking, existing analytic tools for behavioral studies mentioned above are not well-suited to using all of the information in high-frequency intensive longitudinal data and providing a comprehensive description of the relation between system variables (26,27).

Control systems engineering tools have been applied to regulate complex, dynamic systems (e.g., air conditioning in homes, cruise control in cars, flight in aircraft). In general, a system is a process that produces repeated outcomes in response to repeated predictors. Identifying a model of the relations between variables (i.e., predictors and outcomes) that unfold over time (hereafter, a system) is the starting point for developing algorithms to make informed decisions for controlling the system. System identification techniques are a set of tools that can be used to characterize the relation between predictors and outcomes over time. Often the identified models need to be tuned for the specific system (e.g., for a specific home, vehicle or aircraft) and this tuning is particularly important for systems describing human behavior which have a high level of uncertainty due to multiple causes. We propose to leverage individual intensive longitudinal data to identify person-specific models that describe dynamic relations between stress and smoking for each individual. Models obtained by these techniques provide the parameters needed to potentially develop person-specific algorithms (i.e., decision rules) for

adapting the delivery of a just-in-time intervention based on person-specific vulnerabilities. Such models could realize the potential for personalized behavior medicine to provide people with the right treatment only at the moment when they need it, based on the person-specific dynamics. This approach allows for greater flexibility in designing just-in-time interventions because it can provide evidence-based, tailored – even person-specific – decision rules when stress-smoking systems differ between people (e.g., due to varying model complexity or input salience). In this paper, we focus on stress as a potential tailoring variable for just-in-time interventions to prevent smoking behavior and model stress-smoking systems. Previously, control systems engineering tools were used to model and design interventions for gestational weight gain and physical activity (28–30), but we apply them here to characterize the stress-smoking responses dynamics.

System identification methods – techniques to fit a dynamic model from longitudinal measurements of a predictor to an outcome – have been used to evaluate smoking cessation treatments (31). These methods have been combined with model predictive controllers – controllers (decision rules) that rely on dynamic models of a process obtained by system identification to determine the “right” inputs to the process – to provide a foundation for designing adaptive intensive smoking interventions (32,33). Simulations by Lagoa et al. revealed that treatment outcomes could be improved by delivering real-time interventions when and where needed (33). Bekiroglu et al. also represented the relations between tobacco withdrawal-related processes over time using dynamical models (34). These works demonstrated the feasibility of applying system identification tools to model problems involving smoking. Moreover, they provided preliminary results on how the identified models can be used to develop just-in-time interventions. This study will provide the first application of system identification tools to model person-specific stress-smoking systems and will provide the basis

for applying model predictive control tools to design an adaptive intensive, smoking intervention tailored on stress.

These prior smoking-related applications involved continuous outcome measures, leading to relatively straightforward modeling procedures. However, in our system, the predictor (stress) is represented by a continuous score and the outcome (smoking event) is represented by a binary value. Therefore, another contribution of this work is the development of a modeling technique to describe relations between our continuous predictor and a binary outcome in an efficient way. Model identification of a connection of a linear system with a hard nonlinearity is known to be a complex problem because the resulting optimization problem is non-convex and, hence, hard to solve numerically. In this paper, we provide computationally efficient convex relaxations aimed at obtaining the lowest complexity model compatible with the data available. Furthermore, to the best of our knowledge, system identification methods have never been applied to describe the person-specific dynamics of stress-smoking responses with a binary outcome.

The Present Study

This study used wearable devices to collect intensive longitudinal data and applied algorithms to estimate the probability of stress (cStress) at every minute and detect smoking episodes (puffMarker) (35,36). These two measures have been used to evaluate stress and smoking lapses in another study conducted by some of current paper's authors (37). We applied system identification techniques from control systems engineering to model person-specific dynamic relations between stress and smoking. In other words, using only data recorded from each individual, we modeled their individual behavior and, therefore, determined a different system/model for each participant. Analyzing stress-smoking relations especially at the moments preceding and following a smoking episode and generating person-specific model parameters in

the present study provides the basis for designing person-specific, tailored smoking cessation interventions and treatments from the parameters linking vulnerable moments (when stress is likely to lead to smoking) to intervention decisions in the future works.

Method

Design

This study is a secondary analysis of data from the Sense2Stop trial conducted through the MD2K Center of Excellence (38). The Sense2Stop trial involved 15 days of data collection: 3 days prior to a quit attempt, a quit day, and 11 post-quit days. This study used data from the 3 pre-quit days to develop dynamic models of stress and smoking.

Participants

Flyers posted in the community were used to recruit adult smokers interested in quitting smoking for the Sense2Stop trial. Prospective participants ($N = 371$) were screened for eligibility and 75 eligible participants enrolled. Participants were between 18 and 65 years, lived in the Chicago area, smoked more than one cigarette per day on average in the last year, were willing to abstain from non-cigarette tobacco products for the course of the study, willing to try to quit smoking for at least 48 hours (the parent trial focused on smoking cessation but the present study was limited to data from the pre-quit period), not taking any medications for smoking cessation currently or planned in the next 30 days, not pregnant or trying to get pregnant, and willing to provide an emergency contact, social security number, and address for study payment.

Procedures

Participants were provided with four devices and asked to wear them for up to 16 hours/day: two MotionSense wristbands, one Autosense chestband, and an Android smartphone (39,40). The Autosense chestband supports a wireless sensor suite that records continuous

physiological measurements consisting of an electrocardiogram (ECG, 64 Hz), respiratory inductive plethysmograph (RIP, 21.3 Hz) , and 3-axis accelerometer (10.7 Hz) (39). Each MotionSense wristband contained a 3-axis accelerometer (16 Hz) and a 3-axis gyroscope (16 Hz) to record arm motion (36,40). The open-source mCerebrum software has been used to support continuous sensor data transfer from the chestband and wristbands to the smartphone (41). Participants were compensated if they wore the devices above 70% of the requested wear time during the study period (i.e., 11.2 h/day); however, due to either battery issues or participants not wearing the devices correctly, the wear time was less than that for many participants. Full procedures for the Sense2Stop trial are available in (BLINDED FOR REVIEW).

Measures

Stress was operationalized as the minute-level probability of stress determined using the cStress algorithm (35). cStress marker uses data from a 2-lead electrocardiograph and a respiration inductive plethysmograph collected by chest-worn AutoSense sensors to estimate the probability of stress at each minute. Specific features used to construct minute-level stress probabilities from those sensors included the 80th percentile and mean of interbeat interval, and mean and median of the ratio between inspiration and expiration duration. Stress probabilities were calculated every minute except when accelerometers in the chestband recorded physical activity (due to confounding factors from the physiological responses to activity) or when data were unavailable due to poor sensor attachments or other noise. This cStress model previously demonstrated a true positive rate of 88.6% and false positive rate of 4.65% on a test dataset from the lab. The cStress model achieved a median accuracy of 90% and 72% with self-reports from the lab and the field, respectively (35).

Smoking behavior was operationalized every minute as a binary score from the puffMarker algorithm (36). This algorithm uses two sources of data: breathing dynamics collected from the respiration inductive plethysmograph in the chestband, and arm movements gathered from inertial sensors on each wrist. Puffs were identified from patterns of hand-to-mouth gestures and respiration cycles. Features extracted consisted of inhalation and exhalation duration, respiration duration, minimum and maximum of the rate of change signal (first derivation of the respiration signal), and mean, median, standard deviation, and quartile deviation of magnitude of gyroscope, pitch, and roll. Excluding isolated puffs, smoking episodes were marked if a minimum of four smoking puffs was detected. The puffMarker algorithm achieved a true positive rate of 96.9% and a false positive rate of 1.1% on the training dataset (36).

Analyses

Data quality screening. Data were recorded from 75 participants, but participants were excluded if they were part of the pilot test ($n = 5$) or had excessive missing sensor data due to physical activity, improper attachment, non-wear, sensor failures, or insufficient data (i.e. less than five minutes of continuous data). After accounting for these problems ($n = 25$), the analytic dataset comprised 45 participants (60%).

Modeling strategy. In the following we use the term “system” to represent a longitudinal process that produces repeated outcomes in response to repeated predictors. For these analyses, stress probability will be considered as the predictor and smoking will be considered as the outcome. The modeling objective is to identify a system that represents the relation between these two longitudinal signals. Figure 1 represents the configuration of the system to be identified. Note that, in the proposed approach, we aim at modeling individual behavior and, therefore, a different system or model is identified for each participant.

The first part of the model is a linear system with stress probability as the predictor and “response of the system,” denoted by y_{system} , as the outcome (42). See equation **Error!**

Reference source not found. below for how y_{system} is related to the binary outcome of smoking. The linear system response, y_{system} , as indicated in Figure 1, is the summation of the behavior of the system which is directly connected to stress, y_{cause} , and the intrinsic behavior of the system, $y_{intrinsic}$. In the context of dynamical model of the stress-smoking responses, intrinsic response, $y_{intrinsic}$, can be thought as the estimate of smoking odds when the participant is not experiencing stress.

The right-hand side block in the system configuration depicted in Figure 1, which is called “sign nonlinearity”, provides the binary outcome indicating whether, at a specific time, the participant is smoking or not. The response of identified system at a specific time t , $y_{system}(t)$, has a nonlinear relation with smoking which is described as

$$\begin{cases} smoking(t) = 1 & \text{if } y_{system}(t) \geq 1 \\ smoking(t) = 0 & \text{if } y_{system}(t) < 1 \end{cases} \quad (1)$$

In equation (1), the use of a threshold equal to 1 is arbitrary; the choice of the threshold does not impact the overall results.¹

Data processing. Data pre-processing consisted of five general stages: correcting sample times, handling small intervals of missing data via linear interpolation, determining ending time of smoking episodes, dividing data into continuous pieces called data chunks, and excluding data chunks with less information is done. We now describe each step in detail.

¹ Scaling the threshold will lead to an inverse scaling of the output of the linear system. Therefore, since we are dealing with linear systems, it does not change the identification problem.

First, the timing for each stress probability and smoking event is provided in timestamp format. Therefore, the primary step in processing the data is to convert timestamps to datetime. Datetime version represents the time and date information of the events more intuitively, making the results and their display more easily interpretable. The sampling rate in this study is approximately one minute; however, the timing of some samples violated this criterion, such as when two probabilities were provided for a minute (e.g., with time difference of 59 seconds). To address this matter, the approach taken is to round time to the nearest minute. Therefore, in this context, a *sample* means one minute of information, and a *chunk* is a continuous series of samples.

Second, different durations of missing data occurred in the stress probability measure, ranging from one to about 120 minutes. If intervals of missing data were brief (≤ 2 min), linear interpolation technique was used. The reason for this choice is that, given our dataset, two interpolated samples can still provide good consistency as well as maintaining the accuracy of the provided data. Also, these short intervals of missing data would mostly occur as the result of mis-attached sensors on the body rather than any kind of confounding activities which can interfere with stress probability estimation. If intervals were large (> 2 min), there was not enough information to make any prediction/estimation of missing stress probabilities at those time intervals.

Third, the puffMarker algorithm identified the start timing of each identified smoking episode and the timing of all detected smoking puffs, but did not identify when smoking episodes ended. A minimum of four puffs in a five-minute time interval was required to qualify as a smoking episode. To determine the duration of the smoking episodes, a sliding five-minute window was used beginning at the start of the episode. As long as the criterion of four puffs

inside the considered window was met, the smoking episode was extended. Smoking episode durations for participants ranged from two to ten minutes.

Finally, given the fragmented nature of the stress probability dataset and the fact that no information was available to understand relations between stress and smoking during time gaps, we considered each continuous interval of data samples as one “chunk” of data. Some data chunks in our dataset consisted of very small number of samples, as low as just one sample, which makes it hard for the system to understand the behavior and identify the meaningful relation between stress and smoking. To extract relevant information, chunks with less than five minutes of information were excluded. Removing these small data chunks from the analytic dataset had minimal impact on the number of minutes that could be modeled (on average, 68 minutes/participant). Mathematical details of the system identification methods used are provided in Appendix 1.

Results

Participant characteristics are summarized in Table 1. A summary of pre-processed and analyzed data for all participants can be found in Supplementary Table 1 and Supplementary Table 2. These quantities include the numbers of samples in the original dataset, samples after interpolation (with missing samples ≤ 2), chunks based on the interpolated dataset, chunks after removing the small chunks (size < 5 samples), and samples after excluding the small chunks. Prior to interpolation, the average total duration of data from each participant sampled in the dataset was 723.1 min (SD = 323.6; range = 79 - 1452) over 3 days; after interpolation, the average total duration of data from each participant increased to 887.9 min (SD = 385.5; range = 90 - 1776) over 3 days (wear time for many participants was less than 11.2 h/day). The interpolated samples were divided into chunks with a mean number of 76.2 chunks/participant

(SD = 33.1; range = 3 - 150). After removing data chunks less than 5 minutes long, the mean number of data chunks/participant used for modeling was 40.5 (SD = 16.45; range = 3 – 80). The mean total duration of samples used for system identification was 819.9 minutes (SD = 375.8; range = 90 – 1716).

The modeling strategy used in this paper was person-specific dynamical modeling. Below, we report a single dynamic model estimated for a single participant to demonstrate the process of linking predictor and outcome through an identified system consisting of a response to the predictor, intrinsic response, and system response. This modeling strategy was then deployed on usable data from each participant separately. Next, we simulated how each participant would respond to a specific continuous stress episode lasting 10 minutes (an arbitrary value to standardize the comparison among participants), to illustrate how these response patterns vary across participants.

Deriving Person-Specific Model of Smoking

Data from a single participant are used to illustrate components of the dynamical system. Figure 2 and Figure 3 show changes in the candidate predictor, outcome, and responses of the identified system for one data chunk consisting of 15 samples. In all figures, the horizontal axis shows time with a minute scale. Figure 2 (left) represents the candidate predictor (stress probability) and Figure 2 (right) shows the outcome (smoking) of the system for every minute. Stress probabilities in Figure 2 (left) ranged from 0 to 1; smoking values of 0 and 1 in Figure 2 (right) indicate non-smoking and smoking events at that minute. Respectively, Figure 3 (a) to (c) represent the predictor response, intrinsic response, and the overall response of the system for this data chunk, described in Figure 1. As mentioned earlier, *predictor response* is a behavior of the system which is directly connected to stress (y_{cause} , Figure 3 (a)), *intrinsic response*

represents the odds of smoking when stress is zero ($y_{intrinsic}$, Figure 3 (b)), and the *system response* plot represents the summation of the response of the system to stress and intrinsic response (y_{system} , Figure 3 (c)). The concept represented in equation (1) is also depicted in (Figure 3 (c)). According to this criterion, whenever the value of $y_{system}(t)$ equals or exceeds the threshold of 1, there is a smoking minute in the outcome (Figure 2 (right)), and if $y_{system}(t) < 1$, there is no smoking occurrence.

Simulated Smoking Responses to Stress

Another response of a system, *pulse response*, is used here to show the relation between stress and smoking. Pulse response is the response to a continuous stress episode for a given duration. For standardization, we simulate the responses to a presumed stress episode of length 10 minutes. This duration is arbitrary and allows us to standardize the comparison among participants. Figure 4 shows the stress episode of length 10 minutes. The same stress episode is used to achieve the pulse response for all participants.

According to their respective pulse responses, results from the modeling of smoking behavior due to stress for all participants can be divided into five clusters. Due to the limited sample size of this study, visual clustering is applied based on the shape of observed participants responses. Figure 5 presents the simulated responses of a representative participant from each cluster to a 10-minute stress episode and Figure 6 to Figure 10 represent the pulse responses of individual participants in each cluster to the stress episode, respectively. Cluster 1 was characterized by instantaneous or relatively rapid responses to stress as indicated by an immediate increase in the odds of smoking following by a sharp decrease in odds of smoking. Cluster 2 was characterized by a delayed response to stress, approximately 7 to 12 minutes after the start of stress episode. Participants in this cluster had shown an increase and subsequently a

decrease in their odds of smoking. Cluster 3 was characterized by an instant increase followed by a second, delayed increase in the odds of smoking due to stress. Cluster 4 was characterized by a single delayed increase in response to stress without a subsequent sharp decrease in the odds of smoking. Finally, Cluster 5 was characterized by an instant decrease in the odds of smoking following by a sharp delayed increase in its odds. Table 2 summarizes the number of participants belonging to each cluster. It should be noted that the higher magnitudes of the pulse response do not necessarily imply a higher likelihood of smoking. In fact, the magnitude of the pulse response represents the estimated changes of propensity for smoking when stress happens, as a function of time. If the magnitude is positive, it represents an increase in the likelihood of smoking and if it is negative, it shows a decrease in the likelihood of smoking.

As shown by the results, all pulse responses to a continuous stress episode demonstrated an increase in the odds of smoking either immediately or with some delay. For the majority of participants ($n = 25$), the increase is followed by a sharp decrease in the tendency to smoke, down to negative probability. This can be interpreted as smoking due to stress leads to reducing the probability of further smoking in the near future for these participants. However, some participants ($n = 12$) did not follow the same trend, i.e., the increase in the odds of smoking did not proceed with a decrease in the smoking likelihood in the near future. For a few individuals ($n = 3$), the early increase in the odds of smoking is followed by a second increase in its odds which shows that smoking in response to stress elevates the odds for further smoking for some people. For some participants ($n = 5$), the behavior is quite different. They represent a negative response to stress at the beginning which can be interpreted as a decrease in the chance of smoking at first moments of being under stress. The trend is followed by a sharp delayed increase in the odds of

smoking. This can be an indicator of the effort of participants to prevent smoking at first; but eventually stress causes smoking to happen.

Discussion

This work makes two key contributions toward understanding relations between stress and smoking. First, it provides proof-of-concept for using system identification tools from control systems engineering to identify person-specific dynamic models of stress-smoking responses for participants. In doing so, it introduced a method for combining a linear system with a sign nonlinearity (Figure 1) to enable modeling of a continuous predictor (stress) and a binary outcome (smoking). Second, results of simulations based on person-specific models demonstrated heterogeneity in smoking responses to stress. For about half of the participants (49%), following an initial stress-induced increase in the likelihood of smoking, we expect to see an eventual reduction in the odds of smoking.

This paper demonstrated how dynamical systems modeling techniques can be applied to characterize the dynamics of stress and smoking. This modeling approach accounted for the influence of both recent stress and smoking to predict the odds of smoking in response to future stress. Prior efforts to link stress and smoking have been based on infrequent stress and smoking assessments for each smoking assessment (10,16,43,44). As sensing technology and digital markers improve, intensive longitudinal data will become more readily available to and important for researchers. Dynamical systems models can incorporate that information to predict the dynamics of smoking systems as a function of stress and other predictor variables.

Comparison of approaches for modeling stress-smoking dynamics

Another tool for modeling longitudinal data is the time-varying effect model (TVEM) (45–51). This model has two noteworthy limitations. First, the TVEM cannot determine complex

underlying within-time associations of variables (52), so it is not useful for investigating the long-term effect of predictors on an outcome. TVEM is limited to contemporaneous relations in a single model without considering the effect of delay in that association. That is, TVEM is not capable of exploring the memory effect of stress on smoking. Tools such as dynamical system modeling techniques from control systems engineering can fill this gap when investigators seek to understand the delayed effect of stress on smoking. Second, the TVEM represents bivariate associations as a function of time but requires an assumption that those time-varying associations are equivalent across people. The dynamical modeling technique applied in this paper was used to generate person-specific models, and the results obtained would not support assumptions of equivalence across participants.

It is worth noting that the dynamical modeling approach applied here was a powerful technique considering the relatively limited amount of time series data, three days in this dataset. Artificial intelligence techniques, including machine learning (53,54) and deep learning methods (55), or statistical techniques such as Bayesian approaches and stochastic approximation expectation-maximization algorithms (56,57) require a large volume of data. Additionally, the majority of these methods only provide the ultimate result not the dynamics of the system, that is, predicting whether the participant smokes or not by the end of the stress episode. With the three days of data available in this study, these alternative methods were not possible. Thus, control systems engineering tools strike a balance between capitalizing on intensive longitudinal data to reveal dynamics and being efficient in the volume of data required.

The person-specific models describing stress-smoking systems were used to simulate smoking dynamics following a 10-minute stress response. In those simulations, stress consistently increased the subsequent likelihood of smoking but there was considerable variation

in how long it took for the likelihood of smoking to increase. In other words, although some participants did not show an immediate increase in smoking likelihood in response to stress, all participants were observed to increase smoking within some minutes of the stress episode. These findings are consistent with the hypothesis that some people use smoking to regulate stress (10,16,43,44). Although stress consistently increased the likelihood of smoking, participants varied in the delay between the onset of stress and observed smoking behavior. The latency of stress effects on smoking is an important contribution about which theories have been silent. For some individuals, stress responses have a relatively rapid effect on smoking behavior that may preclude intervention; for others, stress responses have a more delayed effect that represents a period of vulnerability when an intervention could be delivered. Data from this study do not indicate that participants learned to self-administer nicotine to regulate stress responses. Nevertheless, the approach demonstrated here can be extended to multiple time scales to test self-medication hypotheses – models on faster time scales can characterize stress effects on smoking and models on slower time scales can characterize how stress responses to smoking affect future stress-smoking dynamics (i.e., learning to self-administer nicotine to regulate stress (see (58,59))). Future research could also use n-of-1 or single-case research designs to compare system dynamics (e.g., latency of smoking following a stressor) following the introduction and removal of stress management interventions (60,61). These approaches would capitalize on the idiographic nature of the systems but allow quasi-experimental comparisons of clinically-meaningful outcomes before and after introduction of an intervention. For example, n-of-1 trials could be applied to determine system properties that are amenable or hostile to just-in-time stress-management prompts (e.g., how long of a delay between stress and smoking is required to intervene?).

Limitations

Participants in this study were smokers from large metropolitan area in a Western country. The stress-smoking systems identified here may not generalize to other contexts or populations. Additionally, the smokers in this study were preparing for a quit attempt. Stress-smoking system dynamics may differ for smokers who are not intending to quit or during a quit attempt when withdrawal symptoms present an additional demand. The small sample size was suitable for establishing proof-of-concept for modeling person-specific changes in smoking-responses to stress that can be used to develop person-specific decision rules for just-in-time stress management interventions; however, the sample size was insufficient to permit empirical clustering of participants based on their responses to stress. Future studies with larger samples can apply clustering algorithms and machine learning techniques like Shapelets to establish a robust typology of stress-smoking systems (62,63).

The volume of missing data in the datasets was another limitation of this study. Some missing data was anticipated because the cStress marker cannot be estimated when the participant is moving. Other missing data was caused by human factors or device issues, such as device non-wear or poor attachment of sensors to the skin. One consequence of the missing data was that each participant's data was a collection of shorter chunks of data instead of a single long time series. This problem was partially mitigated in the modeling by considering different intrinsic responses for each chunk according to its own past; however, conclusions cannot be drawn about system dynamics between the available chunks, where data were not available.

A final modeling-related limitation was that we assumed the system is stationary. The dataset did not include information on situational factors that might alter dynamics (e.g., home versus work; weekday versus weekend). Given the relatively short period of three days of pre-

quit data, we believe it was reasonable to assume that participant behavior would not change dramatically during this period. Accessing and analyzing more data including situational and contextual factors and exploring how they impact smoking should be a priority for the future studies.

Conclusion

This study provided proof-of-concept for using system identification tools from control systems engineering to identify relations between stress and smoking for regular smokers over time. The study made both substantive and methodological contributions to the addiction literature. From a substantive perspective, the identified person-specific dynamical models revealed a direct correlation between stress and subsequent increases in smoking risk. This general pattern was uniform but the latency between stress and smoking events varied from instantaneous to delayed. Additionally, for about half of the participants, smoking in response to stress was followed by a marked reduction in the odds of smoking in the near future. From a methodological perspective, the system identification method yielded person-specific dynamical models that might have relatively consistent pattern for a given individual but different patterns for different people. This method provides a new approach for both (a) characterizing the dynamics of smoking risk at different stages (e.g., prior to a quit attempt), and (b) developing person-specific decision rules for just-in-time interventions that support smoking cessation.

References

1. Centers for Disease Control and Prevention National Center for Chronic Disease Prevention and Health Promotion, Smoking and Tobacco Use, Fast Facts [Internet]. Accessed June 22, 2020. Available from: https://www.cdc.gov/tobacco/data_statistics/fact_sheets/fast_facts/.
2. Xu X, Bishop E, Kennedy S, Simpson S, Pechacek T. Annual Healthcare Spending Attributable to Cigarette Smoking: An Update. *Am J Prev Med*. 2015;48(3):326–333.
3. Hall W, Doran C. How Much Can the USA Reduce Health Care Costs by Reducing Smoking? *PLoS Med*. 2016;13(5):e1002021.
4. McEwen A, West R, McRobbie H. Motives for Smoking and their Correlates in Clients Attending Stop Smoking Treatment Services. *Nicotine Tob Res*. 2008;10(5):843–50.
5. Byrne DG, Byrne AE, Reinhart MI. Personality, Stress and the Decision to Commence Cigarette Smoking in Adolescence. *J Psychosom Res*. 1995;39(1):53–62.
6. Torres OV, O'Dell LE. Stress Is a Principal Factor that Promotes Tobacco Use in Females. *Prog Neuro-Psychopharmacology Biol Psychiatry*. 2016;65:260–8.
7. Pomerleau OF, Pomerleau CS. Research on Stress and Smoking: Progress and Problems. *Br J Addict*. 1991;86(5):599–603.
8. Selye H. *The Stress of Life*. New York, McGraw-Hill. 1956.
9. Lazarus RS, Folkman S. *Stress, Appraisal and Coping*. New York: Springer. 1984.
10. Khantzian EJ. The Self-Medication Hypothesis of Substance Use Disorders: A Reconsideration and Recent Applications. *Harv Rev Psychiatry*. 1997;4(5):231–44.
11. Kassel JD, Stroud LR, Paronis CA. Smoking, Stress, and Negative Affect: Correlation, Causation, and Context Across Stages of Smoking. *Psychol Bull*. 2003;129(2):270–304.
12. Benowitz NL. Clinical Pharmacology of Nicotine: Implications for Understanding, Preventing, and Treating Tobacco Addiction. *Clin Pharmacol Ther*. 2008;83(4):531–41.
13. Delfino RJ, Jamner LD, Whalen CK. Temporal Analysis of the Relationship of Smoking Behavior and Urges to Mood States in Men versus Women. *Nicotine Tob Res*. 2001;3(3):235–48.

14. Jahnel T, Ferguson SG, Shiffman S, Schüz B. Daily Stress as Link Between Disadvantage and Smoking: An Ecological Momentary Assessment Study. *BMC Public Health*. 2019;19(1):1–8.
15. Weinstein SM, Mermelstein R, Shiffman S, Flay B. Mood Variability and Cigarette Smoking Escalation Among Adolescents. *Psychol Addict Behav*. 2008;22(4):504–13.
16. Pesko MF, Baum CF. The Self-Medication Hypothesis: Evidence from Terrorism and Cigarette Accessibility. *Econ Hum Biol*. 2016;22:94–102.
17. Aronson KR, Almeida DM, Stawski RS, Klein LC, Kozlowski LT. Smoking Is Associated with Worse Mood on Stressful Days: Results from a National Diary Study. *Ann Behav Med*. 2008;36(3):259–69.
18. Savoy EJ, Businelle MS, Nguyen N, Chen TA, Neighbors C, Norton PJ, Taing M, Reitzel LR. Examining Moment to Moment Affective Determinants of Smoking Rate Following a Quit Attempt Among Homeless Daily Smokers. *Addict Behav*. 2021;115:106788.
19. Pesko MF. Stress and Smoking: Associations with Terrorism and Causal Impact. *Contemp Econ Policy*. 2014;32(2):351–71.
20. Hughes JR. Effects of Abstinence from Tobacco: Valid Symptoms and Time Course. *Nicotine Tob Res*. 2007;9(3):315–27.
21. Hughes JR. Tobacco Withdrawal in Self-Quitters. *J Consult Clin Psychol*. 1992;60(5):689–97.
22. Cameron A, Reed KP, Ninnemann A. Reactivity to Negative Affect in Smokers: The Role of Implicit Associations and Distress Tolerance in Smoking Cessation. *Addict Behav*. 2013;38(12):2905–12.
23. Shiffman S. Dynamic Influences on Smoking Relapse Process. *J Pers*. 2005;73(6):1715–48.
24. McKee SA, Sinha R, Weinberger AH, Sofuoglu M, Harrison EL, Lavery M, Wanzer J. Stress Decreases the Ability to Resist Smoking and Potentiates Smoking Intensity and Reward. *J Psychopharmacol*. 2011;25(4):490–502.
25. Devaux M, Sassi F. Social Disparities in Hazardous Alcohol Use: Self-Report Bias May Lead to Incorrect Estimates. *Eur J Public Health*. 2016;26(1):129–34.

26. Conroy DE, Lagoa CM, Hekler EB, Rivera DE. Engineering Person-Specific Behavioral Interventions to Promote Physical Activity. *Exerc Sport Sci Rev.* 2020;48(4):170–9.
27. Dunton GF, Rothman AJ, Leventhal AM, Intille SS. How Intensive Longitudinal Data Can Stimulate Advances in Health Behavior Maintenance Theories and Interventions. *Transl Behav Med.* 2021;11(1):281–6.
28. Conroy DE, Hojjatinia S, Lagoa CM, Yang CH, Lanza ST, Smyth JM. Personalized Models of Physical Activity Responses to Text Message Micro-Interventions: A Proof-of-Concept Application of Control Systems Engineering Methods. *Psychol Sport Exerc.* 2019;41:172–80.
29. Hojjatinia S, Lagoa CM, West AB, Hojjatinia S, Turrisi T, Brunke-Reese D, Conroy DE. Abstract P196: Personalized Dynamical System Models of Individual Text Message Effects on Changes in Physical Activity. *Circulation.* 2020;141(Suppl_1):AP196.
30. Guo P, Rivera DE, Downs DS, Savage JS. Semi-Physical Identification and State Estimation of Energy Intake for Interventions to Manage Gestational Weight Gain. 2016 Am Control Conf. 2016;1271–6.
31. Timms KP, Rivera DE, Collins LM, Piper ME. Control Systems Engineering for Understanding and Optimizing Smoking Cessation Interventions. 2013 Am Control Conf. 2013;1964–9.
32. Rawlings JB. Tutorial Overview of Model Predictive Control. *IEEE Control Syst Mag.* 2000;20(3):38–52.
33. Lagoa CM, Bekiroglu K, Lanza ST, Murphy SA. Designing Adaptive Intensive Interventions Using Methods from Engineering. *J Consult Clin Psychol.* 2014;82(5):868–78.
34. Bekiroglu K, Russell MA, Lagoa CM, Lanza ST, Piper ME. Evaluating the Effect of Smoking Cessation Treatment on a Complex Dynamical System. *Drug Alcohol Depend.* 2017;180:215–22.
35. Hovsepian K, Al'Absi M, Ertin E, Kamarck T, Nakajima M, Kumar S. cStress: Towards a Gold Standard for Continuous Stress Assessment in the Mobile Environment. *UbiComp 2015 - Proc 2015 ACM Int Jt Conf Pervasive Ubiquitous Comput.* 2015;493–504.
36. Saleheen N, Ali AA, Hossain SM, Sarker H, Chatterjee S, Marlin B, et al. puffMarker: A Multi-

- Sensor Approach for Pinpointing the Timing of First Lapse in Smoking Cessation. UbiComp 2015 - Proc 2015 ACM Int Jt Conf Pervasive Ubiquitous Comput. 2015;999–1010.
37. Nakajima M, Lemieux AM, Fiecas M, Chatterjee S, Sarker H, Saleheen N, Ertin E, Kumar S, al'Absi M. Using Novel Mobile Sensors to Assess Stress and Smoking Lapse. *Int J Psychophysiol.* 2020;158:411–8.
38. Kumar S, Abowd GD, Abraham WT, Al'Absi M, Gayle Beck J, Chau DH, et al. Center of Excellence for Mobile Sensor Data-to-Knowledge (MD2K). *J Am Med Informatics Assoc.* 2015;22(6):1137–42.
39. Ertin E, Raij A, Stohs N, Al'Absi M, Kumar S, Mitra S. AutoSense: Unobtrusively Wearable Sensor Suite for Inferring the Onset, Causality, and Consequences of Stress in the Field. *SenSys 2011 - Proc 9th ACM Conf Embed Networked Sens Syst.* 2011;274–87.
40. Holtyn AF, Bosworth E, Marsch LA, McLeman B, Meier A, Saunders EC, et al. Towards Detecting Cocaine Use Using Smartwatches in the NIDA Clinical Trials Network: Design, Rationale, and Methodology. *Contemp Clin Trials Commun.* 2019;15:100392.
41. Hossain SM, Hnat T, Saleheen N, Nasrin NJ, Noor J, Ho BJ, et al. mCerebrum: A Mobile Sensing Software Platform for Development and Validation of Digital Biomarkers and Interventions. *SenSys 2017 - Proc 15th ACM Conf Embed Networked Sens Syst.* 2017;1–14.
42. Lathi BP, Green RA. *Linear Systems and Signals*. New York: Oxford University Press; 2005.
43. Shiffman S, Wills TA. *Coping and Substance Use*. New York: Academic Press; 1985. 3–24 p.
44. Weinstein SM, Mermelstein RJ. Dynamic Associations of Negative Mood and Smoking Across the Development of Smoking in Adolescence. *J Clin Child Adolesc Psychol.* 2013;42(5):629–42.
45. Lanza ST, Vasilenko SA, Liu X, Li R, Piper ME. Advancing the Understanding of Craving During Smoking Cessation Attempts: A Demonstration of the Time-Varying Effect Model. *Nicotine Tob Res.* 2014;16(Suppl 2):S127–34.
46. Shiyko MP, Lanza ST, Tan X, Li R, Shiffman S. Using the Time-Varying Effect Model (TVEM) to Examine Dynamic Associations between Negative Affect and Self Confidence on Smoking

- Urges: Differences between Successful Quitters and Relapsers. *Prev Sci.* 2012;13(3):288–99.
47. Lanza ST, Vasilenko SA. New Methods Shed Light on Age of Onset as a Risk Factor for Nicotine Dependence. *Addict Behav.* 2015;50:161–4.
48. Merrill JE, Kenney SR, Barnett NP. A Time-Varying Effect Model of the Dynamic Association Between Alcohol Use and Consequences Over the First Two Years of College. *Addict Behav.* 2017;73:57–62.
49. Snippe E, Dziak JJ, Lanza ST, Nyklíček I, Wichers M. The Shape of Change in Perceived Stress, Negative Affect, and Stress Sensitivity During Mindfulness-Based Stress Reduction. *Mindfulness.* 2017;8(3):728–36.
50. Tan X, Shiyko MP, Li R, Li Y, Dierker L. A Time-Varying Effect Model for Intensive Longitudinal Data. *Psychol Methods.* 2012;17(1):61–77.
51. Koslovsky MD, Hébert ET, Businelle MS, Vannucci M. A Bayesian Time-Varying Effect Model for Behavioral mHealth Data. *Ann Appl Stat.* 2020;14(4):1878–902.
52. Lanza ST, Vasilenko SA, Russell MA. Time-Varying Effect Modeling to Address New Questions in Behavioral Research: Examples in Marijuana Use. *Psychol Addict Behav.* 2016;30(8):939–54.
53. Amaral JLM, Lopes AJ, Jansen JM, Faria ACD, Melo PL. An Improved Method of Early Diagnosis of Smoking-Induced Respiratory Changes Using Machine Learning Algorithms. *Comput Methods Programs Biomed.* 2013;112(3):441–54.
54. Senyurek VY, Imtiaz MH, Belsare P, Tiffany S, Sazonov E. A CNN-LSTM Neural Network for Recognition of Puffing in Smoking Episodes Using Wearable Sensors. *Biomed Eng Lett.* 2020;10(2):195–203.
55. Wang S, Zhang R, Deng Y, Chen K, Xiao D, Peng P, Jiang T. Discrimination of Smoking Status by MRI Based on Deep Learning Method. *Quant Imaging Med Surg.* 2018;8(11):1113–20.
56. Lu ZH, Chow SM, Sherwood A, Zhu H. Bayesian Analysis of Ambulatory Blood Pressure Dynamics with Application to Irregularly Spaced Sparse Data. *Ann Appl Stat.* 2015;9(3):1601–20.
57. Chow SM, Lu Z, Sherwood A, Zhu H. Fitting Nonlinear Ordinary Differential Equation Models

- p>with Random Effects and Unknown Initial Conditions Using the Stochastic Approximation Expectation–Maximization (SAEM) Algorithm.
- Psychometrika*
- . 2016;81(1):102–34.
58. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building Health Behavior Models to Guide the Development of Just-in-Time Adaptive Interventions: A Pragmatic Framework. *Heal Psychol*. 2015;34:1209–19.
 59. Spruijt-Metz D, Hekler E, Saranummi N, Intille S, Korhonen I, Nilsen W, et al. Building New Computational Models to Support Health Behavior Change and Maintenance: New Opportunities in Behavioral Research. *Transl Behav Med*. 2015;5(3):335–46.
 60. Dallery J, Cassidy RN, Raiff BR. Single-Case Experimental Designs to Evaluate Novel Technology-Based Health Interventions. *J Med Internet Res*. 2013;15(2):e22.
 61. McDonald S, Quinn F, Vieira R, O'Brien N, White M, Johnston DW, Sniehotta FF. The State of the Art and Future Opportunities for Using Longitudinal n-of-1 Methods in Health Behaviour Research: A Systematic Literature Overview. *Health Psychol Rev*. 2017;11(4):307–23.
 62. Dumortier A, Beckjord E, Shiffman S, Sejdić E. Classifying Smoking Urges via Machine Learning. *Comput Methods Programs Biomed*. 2016;137:203–13.
 63. Ye L, Keogh E. Time Series Shapelets: A New Primitive for Data Mining. *Proc ACM SIGKDD Int Conf Knowl Discov Data Min*. 2009;947–55.
 64. Mohan K, Fazel M. Reweighted Nuclear Norm Minimization with Application to System Identification. *Proc 2010 Am Control Conf ACC 2010*. 2010;(5):2953–9.
 65. Liu Z, Vandenberghe L. Interior-Point Method for Nuclear Norm Approximation with Application to System Identification. *SIAM J Matrix Anal Appl*. 2009;31(3):1235–56.
 66. Shah P, Bhaskar BN, Tang G, Recht B. Linear System Identification via Atomic Norm Regularization. *Proc 51th IEEE CDC*. 2012;6265–6270.
 67. Chandrasekaran V, Recht B, Parrilo PA, Willsky AS. The Convex Geometry of Linear Inverse Problems. *Found Comput Math*. 2012;12(6):805–49.
 68. Yilmaz B, Bekiroglu K, Lagoa C, Szaier M. A Randomized Algorithm for Parsimonious Model

- Identification. IEEE Trans Automat Contr. 2018;63(2):532–9.
69. Hojjatinia S, Bekiroglu K, Lagoa CM. Parsimonious Volterra System Identification. 2018 Am Control Conf. 2018;1933–8.
70. Grant M, Boyd S. CVX: Matlab Software for Disciplined Convex Programming, version 2.1. 2014.

Appendix

Solving the problem of identifying parsimonious models (i.e., here estimating the model coefficients describing the dynamics of stress and smoking) from experimental data has led to much recent work. The nonconvex nature of this problem has directed the proceeding approaches (64,65) to consider relaxations such as Group Lasso or nuclear norm minimization. Although these approaches perform well, they might not be computationally efficient. This difficulty has led to a new approach called atomic norm minimization (66). This technique represents the response of a linear time-invariant system as a linear combination of suitably chosen objects or atoms (67–69). Atomic norm minimization results in efficient identification of sparse models from experimental data. In this paper, we developed the identification algorithm using the atomic norm method. To identify the system (again, to estimate the model coefficients describing the dynamics of stress and smoking), we rely on the idea that the response of a linear time-invariant system can be represented as a linear combination of atoms.

Discrete-time linear system modeling

Note that since we are applying the tools from control systems engineering, in this appendix, we will use the common terms using in this field. Therefore, from now on “input” refers to predictor, “output” refers to outcome, and “ $y_{Natural}$ ” represents the intrinsic response.

The response of linear system, $y(k)$, is the sum of the response of the system to the input (i.e., stress probability) and the natural response. It can be represented as follows

$$\begin{aligned} y_{System} &= y_{Stress} + y_{Natural} \\ y(k) &= (h * u)(k) + y_N(k) \end{aligned} \tag{2}$$

where k is the time instant, u and h are the input and impulse response of the system, respectively. $h * u$ is the response of system to the input when system is starting at rest and y_N is

the response of the system while input is considered to be zero. The operator “*” denotes convolution and can be expanded as below

$$(h * u)(k) = \sum_{t=0}^k h(t)u(k - t). \quad (3)$$

Any impulse response of a linear time-invariant system can be represented as

$$h(k) = \sum_{i=0}^{N_p} C_i \alpha_i p_i^k \quad (4)$$

where C_i ’s are the unknown coefficient of system related to input response, p ’s are “poles” of the system inside the unit circle, N_p is the number of poles, and α_p is a scaling factor. The scaling term can be represented as

$$\alpha_p = \frac{1 - |p|^2}{1 - |p|^{N_l+2}} \quad (5)$$

where N_l is the length of measurements vector and $|\cdot|$ returns the absolute value. The size of vector α_p is the same as vector p , consisting poles of the system. Detail of this choice of scaling factor is provided in (68).

Remark: The number of poles used here is equal to the order of system or the complexity of system.

The response of system in equation (2) can be rewritten as

$$y(k) = \left(\left(\sum_p C_p \alpha_p p^k \right) * u \right) (k) + \sum_p C_p^N \alpha_p p^k \quad (6)$$

where $\alpha_p p^k$ are the atoms and C_p and C_p^N are the unknown coefficients of system related to input response and natural response, respectively. A precise description of linear systems concepts is represented in (42).

Data Chunks

Considering the data set as data chunks results in having different intrinsic behaviors with respect to each of data chunks. The equation (6) can be rewritten for each data chunk as follows

$$y_j(m) = \left(\left(\sum_p C_p \alpha_p p^m \right) * u_j \right) (m) + \sum_p C_{jp}^N \alpha_p p^m \quad (7)$$

where j represents data chunk j , m is the time instant in data chunk j and y_j , u_j , and C_{jp}^N are the system response, input and coefficients related to natural response for data chunk j .

It should be noted that as our dataset is the integration of data chunks with various lengths, the value N_l in equation (5) is considered to be the average length of all data chunks for each participant.

We should mention that the value of $y(k)$ is not provided, but the output of system which is smoking is available and has a nonlinear relation with $y(k)$ which is described as

$$\begin{cases} y(k) \geq 1 & \text{if and only if } \text{smoking}(k) = 1 \\ y(k) < 1 & \text{if and only if } \text{smoking}(k) = 0 \end{cases} \quad (8)$$

Therefore, we use input stress probability and output smoking to identify the system.

Sparsity

The objective here is to identify a sparse model in order to minimize the complexity. In other words, we seek to minimize the number of nonzero coefficients in the system, C_p 's, and correspondingly maximize the sparsity of the identified system. The literature on sparsity can be found in (66).

For the poles, the unit circle has been gridded uniformly and random poles were picked to check as candidates for the system. Supplementary Figure 1 represents an example of gridded unit circle.

Supplementary Figure 1: A sample of gridding the unit circle (69).

System identification formulation

The identification problem is formulated as

$$\begin{aligned}
 & \min_t \|t\|_0 \\
 & \text{subject to} \\
 & |C_p| \leq t_p \quad \text{for all } p \in \text{Grid} \\
 & |C_p^{N,l}| \leq t_p \quad l = 1, 2, \dots, N_{ch}; \quad \text{for all } p \in \text{Grid} \\
 & C_p = C_p^* \\
 & C_p^N = C_p^{N*} \quad \text{for all data chunks} \\
 & \begin{cases} y(k) - 1 \geq \varepsilon & \text{if } s(k) = 1 \\ y(k) - 1 < \varepsilon & \text{if } s(k) = 0 \end{cases} \\
 & y(k) = \left(\left(\sum_p C_p \alpha_p p^k \right) * u \right)(k) + \sum_p C_p^N \alpha_p p^k
 \end{aligned} \tag{9}$$

where N_{ch} is the number of data chunks and $\|t\|_0$ is l_0 norm which is the number of non-zero elements in vector t . Minimizing $\|t\|_0$ subject to constraints is a computationally complex

optimization problem. Therefore, the system identification problem can be solved using l_1 norm relaxation as follows

$$\begin{aligned}
 & \min_t \|t\|_1 \\
 & \text{subject to} \\
 & |C_p| \leq t_p \quad \text{for all } p \in \text{Grid} \\
 & |C_p^{N,l}| \leq t_p \quad l = 1, 2, \dots, N_{ch}; \quad \text{for all } p \in \text{Grid} \\
 & C_p = C_p^* \\
 & C_p^N = C_p^{N*} \quad \text{for all data chunks} \\
 & \begin{cases} y(k) - 1 \geq \varepsilon & \text{if } s(k) = 1 \\ y(k) - 1 < \varepsilon & \text{if } s(k) = 0 \end{cases} \\
 & y(k) = \left(\left(\sum_p C_p \alpha_p p^k \right)^* u \right)(k) + \sum_p C_p^N \alpha_p p^k
 \end{aligned} \tag{10}$$

where $\|t\|_1$ is the sum of the absolute values of elements in vector t , s is smoking, p^* is the complex conjugate of p , and C_p^* and C_p^{N*} are the complex conjugate of C_p and C_p^N . ε is the small margin to make the optimization problem well posed and is considered 10^{-5} in the implementation. MATLAB software for disciplined convex programming, CVX, is used as the convex optimization toolbox (70).

Figure 1: System configuration.

Figure 2: Predictor (stress probability)(left) and outcome (smoking)(right) of the system for one data chunk.

(a)

(b)

(c)

Figure 3: Different responses of the identified system for one data chunk, described in Figure 1: (a) Response of the system to predictor (stress), (b) Intrinsic response of system, (c) System response which is sum of predictor response and intrinsic response.

Figure 4: Stress episode of length 10 minutes used to stimulate person-specific responses. In the context of stress probability, the probability of individuals being under stress the first 10 minutes is 1 and for the rest of time is 0.

Figure 5: Pulse response of a representative of all clusters. The vertical line at the 10-min mark on the x-axis reflects the time when the stress probability returned to zero.

Figure 6: Pulse response for participants in cluster 1, instant increase following by a sharp decrease in smoking odds. The vertical line at the 10-min mark on the x-axis reflects the time when the stress probability returned to zero.

Figure 7: Pulse response for participants in cluster 2, a delayed increase following by a sharp decrease in smoking odds. The vertical line at the 10-min mark on the x-axis reflects the time when the stress probability returned to zero.

Figure 8: Pulse response for participants in cluster 3, two rounds of increases in smoking odds. The vertical line at the 10-min mark on the x-axis reflects the time when the stress probability returned to zero.

Figure 9: Pulse response for participants in cluster 4, just one delayed increase in smoking odds in response to stress. The vertical line at the 10-min mark on the x-axis reflects the time when the stress probability returned to zero.

Figure 10: Pulse response for participants in cluster 5, instant decrease in the odds of smoking following by a sharp delayed increase. The vertical line at the 10-min mark on the x-axis reflects the time when the stress probability returned to zero.

Table 1: Characteristics of the participants.

Demographic characteristic	n (%)	M (SD)	Observed Range
Age		42.6 (14)	20 - 64
Sex			
Male	23 (51%)		
Female	22 (49%)		
Race			
American Indian/ Alaska Native	0 (0%)		
Asian	1 (2%)		
Native Hawaiian/ Other Pacific Islander	0 (0%)		
Black/African American	23 (51%)		
White	14 (31%)		
Two or more races	5 (11%)		
Other	2 (4%)		
Ethnicity			
Not Hispanic/Latino	40 (89%)		
Hispanic/Latino	5 (11%)		
Number of cigarettes per day			
10 or less	28 (62%)		
11-20	12 (27%)		
21-30	3 (7%)		
31 or more	2 (4%)		
Age of starting to smoke?		18.3 (6.5)	7 - 45

Table 2: Number of participants and the percentage for each detected cluster.

Cluster	Number of Participants	Percentage
Cluster 1: Immediate increase following by a decrease in odds of smoking	12	26.7%
Cluster 2: A delayed response to stress, an increase following by a decrease in odds of smoking	10	22.2%
Cluster 3: Two rounds of increase; an instant increase following by a delayed increase in odds of smoking	5	11.1%
Cluster 4: Just a delayed increase in odds of smoking	13	28.9%
Cluster 5: An instant decrease following by a sharp delayed increase in odds of smoking	5	11.1%

Supplementary Table 1: Data information for all the participants.

Participant ID	Original data samples (in minutes)	Data samples after interpolation (in minutes)	Number of original data chunks	Number of data chunks after discarding chunks with size < 5	Data samples after discarding data with chunk size < 5 (in minutes)
Participant 201	529	632	30	20	607
Participant 202	513	642	63	38	591
Participant 203	650	775	46	22	736
Participant 204	434	554	65	30	473
Participant 205	706	878	81	40	792
Participant 208	894	1169	121	64	1057
Participant 209	538	625	38	26	601
Participant 210	79	90	3	3	90
Participant 211	953	1182	85	51	1114
Participant 212	911	1076	89	45	989
Participant 213	782	913	63	33	860
Participant 214	768	976	113	50	851
Participant 215	1176	1380	51	31	1346
Participant 216	706	889	62	40	846
Participant 217	852	1043	90	48	957
Participant 219	766	933	55	37	898
Participant 220	553	673	66	33	598
Participant 222	625	770	89	47	699
Participant 226	1014	1253	101	67	1186
Participant 228	843	1063	90	50	992
Participant 229	1452	1776	90	60	1716
Participant 230	1077	1329	112	60	1237
Participant 231	1132	1402	148	80	1280
Participant 233	803	992	98	51	909
Participant 234	392	550	69	39	496
Participant 235	802	995	75	36	916
Participant 240	766	960	120	58	842
Participant 242	318	458	94	37	355
Participant 244	356	433	36	19	399
Participant 245	500	620	74	29	538
Participant 250	1169	1401	85	52	1345
Participant 251	1129	1336	95	52	1250
Participant 252	1221	1459	96	51	1365
Participant 253	1180	1408	78	56	1366
Participant 255	611	789	99	47	688
Participant 257	351	420	12	10	412
Participant 258	194	236	22	14	221
Participant 259	310	365	31	17	340
Participant 260	511	629	64	35	582
Participant 261	918	1103	56	37	1067
Participant 262	823	1023	111	60	914
Participant 264	377	529	150	40	335
Participant 265	1152	1353	88	55	1293
Participant 268	590	717	97	40	602
Participant 270	115	157	26	14	143

Supplementary Table 2: Statistics of data used in this study.

Data Characteristic	Mean	Standard Deviation	Median	Minimum	Maximum
Original data samples (in minutes)	723.13	323.57	766	79	1452
Data samples after interpolation (in minutes)	887.91	385.48	913	90	1776
Number of original data chunks	76.15	33.12	81	3	150
Number of data chunks after discarding chunks with size < 5	40.53	16.46	40	3	80
Data samples after discarding data with chunk size < 5 (in minutes)	819.86	375.81	846	90	1716