BMJ Open

Ambulatory Smoking Habits Investigation based on Physiology and Context (ASSIST) using wearable sensors and mobile phones: protocol for an observational study

Donghui Zhai,1,2 Giuseppina Schiavone,3 Ilse Van Diest,4 Elske Vrieze,5 Walter DeRaedt,1 Chris Van Hoof1,2

ABSTRACT

Introduction Smoking prevalence continues to be high over the world and smoking-induced diseases impose a heavy burden on the medical care system. As believed by many researchers, a promising way to promote healthcare and well-being at low cost for the large vulnerable smoking population is through eHealth solutions by providing self-help information about smoking cessation. But in the absence of first-hand knowledge about smoking habits in daily life settings, systems built on these methods often fail to deliver proactive and tailored interventions for different users and situations over time, thus resulting in low efficacy. To fill the gap, an observational study has been developed on the theme of objective and non-biased monitoring of smoking habits in a longitudinal and ambulatory mode. This paper presents the study protocol. The primary objective of the study is to reveal the contextual and physiological pattern of different smoking behaviours using wearable sensors and mobile phones. The secondary objectives are to (1) analyse cue factors and contextual situations of smoking events; (2) describe smoking types with regard to users’ characteristics and (3) compare smoking types between and within subjects.

Methods and analyses This is an observational study aimed at reaching 100 participants. Inclusion criteria are adults aged between 18 and 65 years, current smoker and office worker. The primary outcome is a collection of a diverse and inclusive data set representing the daily smoking habits of the general smoking population from similar social context. Data analysis will revolve around our primary and secondary objectives. First, linear regression and linear mixed model will be used to estimate whether a factor or pattern have consistent (p value<0.05) correlation with smoking. Furthermore, multivariate multilevel analysis will be used to examine the influence of smokers’ characteristics (sex, age, education, socioeconomic status, nicotine dependence, attitudes towards smoking, quit attempts, etc), contextual factors, and physical and emotional statuses on their smoking habits. Most recent machine learning techniques will also be explored to combine heterogeneous data for classification of smoking events and prediction of craving.

Ethics and dissemination The study was designed together by an interdisciplary group of researchers, including psychologist, psychiatrist, engineer and user involvement coordinator. The protocol was reviewed and approved by the ethical review board of UZ Leuven on 18 April 2016, with an approval number S60078. The study will allow us to characterise the types of smokers and triggering events. These findings will be disseminated through peer-reviewed articles.

Strengths and limitations of this study

► This longitudinal study in real-life setting will lead to more representative and informative knowledge about smokers with office occupations.
► Combining ecological momentary assessment survey with objective and quantitative monitoring of smoking behaviour using non-invasive wearable sensors and mobile phones.
► Compared with old-fashioned research methods, less burden and higher compliance can be expected by leveraging the latest automatic data acquisition and smoking detection techniques.
► This is an observational study using routinely collected data, and therefore residual confounding factors may still be present or unaccounted for.

INTRODUCTION

Globally, there are 1.1 billion current smokers and cigarette smoking causes one in ten death.3 Meanwhile, smoking continues to be the leading cause of preventable morbidity and mortality in Europe and North America, and it is also one of the most significant contributing factors to inequalities in health and life expectancy.2 3 Estimated annual smoking-attributable economic cost of direct medical care and lost productivity is around US$300 billion in the USA, which imposes a heavy burden on the healthcare system, economy and society.4 Due to the slow decline in the prevalence of current smoking,
the burden can be expected to remain at high levels for decades into the future.1–4

Primary healthcare used to play an essential role in smoking cessation. But to promote healthcare and well-being of these large at-risk smoking populations at an affordable expense, more and more researchers start looking for solutions through eHealth and mobile health systems, which can deliver tailored counselling and behavioural change treatment in a scalable way.5–8 In the past decades, technology-enabled systems, including short message service (SMS) text messages, self-help websites, self-tracking smartphone apps and wearable sensors, have been exploited for smoking cessation, and they are deemed by researchers as a very promising and low-cost way to facilitate and scale-up end users’ access to valuable information as they can offer service 24/7 without any distance constraints.9–11 Nonetheless, most often these applications turn out to be not as effective as they are expected to be and fails to deliver the right kind of services at the appropriate time or when they are most needed.

This incapability can be mainly attributed to several reasons: first, these systems are built on general information about smoking habits obtained through old-fashioned methods, such as face-to-face interviews, questionnaires, post reflections or retrospective self-reports, which require a substantial effort from participants and are subject to their recall biases12 13; second, the delivering of these digital services (eg, regular text/email reminder, benefits calculator and cigarette tracker in the app) either relies on user’s proactiveness or only uses simplistic mechanisms which don’t take the contextual information into consideration to adapt to users’ needs over time and fall short of providing tailored feedback.14–17 Though a pioneer study tried to provide location-aware smoking cessation, but its principle is still too simple and naïve to account for all the smoking causes.18 In fact, current evidence suggested that a wide range of cues, such as physical environments, situations and feelings, can act as triggers for smoking.19 20 Specifically, cravings induced by cues from the environment are implicated in almost half of lapses to smoking.21

Rationale
To overcome the above drawbacks and lay the foundation for a truly personalised context-aware counselling and treatment system for smoking cessation in the future, a feasible way is to follow the traditional practice of data scientists who would emphasise the importance of collecting the right data to be able to draw any conclusion useful about the world. In this respect, an observational study to objectively monitor factors and routines associated with everyday smoking over a long period is a clear gap not being filled yet.

Fortunately, nowadays researchers have several advantageous tools which were not available in the previous decade enabling us to collect high-quality data from smokers in an unprecedentedly comprehensive and quantitative way. Among many other endeavours to facilitate this kind of study, a significant effort has been devoted to the development of novel systems for the monitoring and autodetection of smoking events in daily life settings by researchers around the world. For example, a personal automatic cigarette tracker system, comprising of an accelerometer, a respiratory inductive plethysmograph band and RF sensors, an instrumented lighter and also a global positioning system (GPS), was successfully developed and validated by Tiffany’s group for automatic smoking recognition based on hand-to-mouth gestures and respiratory signals.22–25 Similarly, a puffMarker system developed by Kumar’s group also employed respiration and arm movement for smoking detection.26 Besides, acoustic sensor and breath carbon monoxide (CO) sensor were also used as a way to identify smoking by Echebarria and Valencia, respectively.27 28 In addition, less cumbersome and more naturalistic systems with smaller form factors such as augmented lighters and a wrist-worn RisQ system, were used for capturing smoking events.29–32 These sensors can avoid users’ subjective bias and significantly alleviate their burden through automatic detection.

With wearable sensors and smartphones becoming ubiquitous and pervasive in our lives, it has opened new avenues for unobtrusive measurement in natural and real-life settings.19 33 On the one hand, wearable sensors can incessantly acquire both physiological, such as electrocardiography (ECG), galvanic skin response (GSR) and photoplethysmography, and inertial signals, such as acceleration and angular velocity.34 35 From these signals, we can assess a subject’s sleep quality, mobility pattern, heart rate variability (HRV) and stress arousal quantitatively, which could help us form a more accurate picture of their psychophysiological statuses36, while on the other, smartphones often embedded with many sensors can not only record time, location, weather, ambient light and sound related to smoking events, but also trigger ecological momentary assessment (EMA) surveys at the opportunistic moments. These features will automate the acquisition of multimodal information, which, once fused, will enable researchers to infer the underlying causes and contexts of smoking.37 38 On the whole, combining traditional research methods with wearable and mobile technologies will empower researchers to conduct a quantified and non-biased study of smoking habits outside laboratory settings.

OBJECTIVES
The primary objective is to uncover underlying patterns of smoking habits from these collected physiological and contextual data by correlating them with subjective EMA surveys. The secondary objectives are to (1) find out the triggering factors of different smoking events based on concurrent contexts, smokers’ psychophysiological statuses and characteristics; (2) cluster smokers into different groups based on their smoking pattern
and triggering factors and (3) compare smoking patterns within and between subjects.

**METHODS AND ANALYSIS**

**Study design**

As smokers usually have various and sometimes disparate backgrounds, we narrow down our research subjects to a cohort group with similar social contexts. Given that most adults in current society spend approximately one-third of their day working, this study is designed to only focus on adult smokers with occupations. Additionally, to facilitate data collection and minimise artefacts caused by intense physical activity, the main part of participants’ jobs should be done indoors in a physical non-active way. Another reason in favour of the decision is that the workplace is an important and useful setting through which large groups of employees can be reached by public health and health promotion initiatives, thus making it a good scenario for implementing smoking cessation services in the future. 16, 39

The monitoring is not confined to just working hours but goes beyond. Compared with the working time, off-work time is of equal importance, if not more, as it constitutes a larger part of the day especially during the weekends. What makes this study differ from most of the past ones is that researchers in this study have exploited the latest mobile technologies and wearable sensors to automatically monitor both physiology and context of smokers while they’re smoking in the scene. We believe this will bring us several advantages: first, through comprehensive coverage and objective measurement, confounders and outliers will be more likely pinpointed and accounted for, thus making conclusions more general and robust; second, it will enable us to compare smoking habits during working time with non-working time, which may vary or interact with each other in different ways.

**Study setting**

This study is conducted in the ambulatory environment within Dutch-speaking zones in the Flanders areas of Belgium, where smokers are recruited and monitored during their normal daily life for a period of 4 weeks. A feasibility study intended for an internal test and validation was finished at the end of 2017. After the initial assessment, the study is now scaled up and currently undergoing. It is due to finish by the end of 2020.

**Sample size**

As this is an observational study, there are no formal sample size calculations. Based on the assessment of logistical factors in the feasibility test, the aim of this study is to recruit 50 smokers at the least, and to reach a total of 100 participants if possible.

**Eligibility criteria**

This study will focus on daily smokers with office jobs. Daily smokers are defined, here in our protocol, as subjects who smoke at least three cigarettes a day and have smoked daily at least for a year before the study. Meanwhile, the participants should work in office settings where jobs are done in a less physical intensive way compared with arduous tasks, such as construction, mining and farming, or athletic activities, such as playing basketball and running. This is merely for the sake of good measurement of physiological signals during the main course of the experiment. It doesn’t restrict participants from performing their normal daily activities, nor does it refrain them from recording the smoking events when they are in active modes (eg, walking fast and talking).

Other inclusion criteria are: adult smokers aged between 18 and 65 years, which is also the legal working age in the country of data collection, and only smoke boxed or roll-your-own cigarettes. Subjects with the following conditions are excluded from the study: pregnancy, psychological diseases history, cardiac and respiratory problems, and other physical diseases, which make them stay at home or temporally stop smoking. Only participants who comply with these criteria will be enrolled for the study. Compliance is verified through an online screening questionnaire at the beginning of the study.

**Recruitment**

To recruit volunteers, flyers and posters about the study are disseminated among candidate participants online through different channels maintained by an imec department called living labs, which is dedicated to managing user involvement for various studies. 40 In these communications, the aim, procedure and methods of this research are explained to candidates. Besides, a website designed for this study is used to provide detailed information, register participants and conduct baseline surveys. 41

As the procedures depicted in figure 1, once the participant is enrolled in the study, two face-to-face appointments will be immediately scheduled at the beginning and end of the study, respectively. During the first meeting, researchers will give instructions on the usage
of the sensors and app. The whole period of the study will last for 4 weeks, during which participants will wear a chest patch in the first 5 days and a wristband over the entire course for physiological measurement. Besides, an electric lighter and a smartphone are used for contextual information collection. The detailed information about the questionnaires and sensors are explained in the following section. Participants who have completed the study will receive a restaurant voucher worth around €50.

Procedures and instruments

Questionnaires

In this study, four kinds of questionnaires, namely baseline profile questionnaire, intake questionnaire, EMA surveys and feedback questionnaires, are used, and all information and questionnaires are available in Dutch and English to facilitate the participation of expats and subjects from different countries. In detail, the profile and intake questionnaires are presented at the beginning, and feedback questionnaires at the end of the study via the website shown in figure 2. The EMA surveys are repeated multiple times a day over the course of the study via the smartphone app.

The baseline profile questionnaire is required to fill up when the participants pass the screening phase and register for this study. The questions are about personal background information, such as age, gender, education, ethnicity, self-assessment of health, daily consumed cigarettes and so on. Then the participants are instructed to fill out an intake questionnaire. This questionnaire collects general information about participants’ smoking, such as smoking ban policy at home and workplace, alcohol and coffee consumption frequency, smoking initiation age, quitting attempts, attitudes towards smoking and so on. Some question items in this questionnaire are excerpted from standard questionnaires, such as the Attitudes Towards Smoking Scale and the Fagerström Test for Nicotine Dependence.42

In the process of this study, participants receive prompts from the app installed on their smartphone reminding them to make annotations about their emotions, social context, activity, smoking urges and so on.43 These prompts are defined to be primarily triggered by smoking events which are monitored through a Bluetooth-enabled lighter. However, when the Bluetooth connexion is down, the triggering can fall back on predefined randomisation mechanism. In such cases, users will receive at most five randomised surveys a day. To prevent users, especially heavy smokers, from being overloaded, this survey is designed to always have at least a 45 min interval in-between. Besides, to relieve boredom that could be caused by the repeat of the same few questions and to maximise the diversity of collected information, a randomised subset from a pool of total 13 questions rotates, resulting in different combination of questions in different pop-ups. Nonetheless, each survey should still consist of 7–8 short questions on average.

Sensors

In total, four sensors comprised of a wristband, a chest patch, an electric lighter and an exhale CO level tester are used in the study. The first three sensors are given to the user during the study, while the last one is only used by the researcher to validate participants’ smoking status.

There are two wearable sensors used to measure physiological signal from the participants as shown in figure 3. The first one is a wrist-worn device called Chillband. It is developed by imec and used to measure GSR, temperature and acceleration. The GSR will be sampled at 256 Hz, temperature at 1 Hz and acceleration at 32 Hz. Participants should wear the sensor the entire day, but to increase comfort, they can take it off at night. As the sensor is not waterproof, participants also need to take it off while taking a shower or washing dishes. Besides, it needs to be charged every 7 days. The second sensor is a chest patch for the measurement of ECG and acceleration. It consists of a patch and a sensor node which is designed to monitor ECG at 256 Hz and acceleration at 32 Hz continuously for 7 successive days. In this study, it is only used in the first 5 entire days and nights of the experiment. Before using it, users first need to switch on the sensor by clicking it into the socket of the patch, which has another sticky side embedded with electrodes. These electrodes are used to connect the sensor with human skin when the patch is applied onto users’ chest. Additionally, the patch, officially called as ePatch, is disposable so that users can take it off and change for a new one when necessary. It is a commercial product manufactured by BioTelemetry, which has received Conformité Européenne (CE) marking and Food and Drug Administration approval.44 For both wearable sensors, data are recorded and stored on their internal secure digital cards, during the experiment. At the end of the experiment, these data can be retrieved through a universal serial bus interface, which is used to charge the devices as well.

The third sensor is a smart lighter called Quibit, shown in figure 4.45 It is an electric lighter powered by a battery. By pressing the side button, it heats coil to light up cigarettes. It records the time when it’s used and syncs to the smartphone wirelessly through Bluetooth. Its battery lasts 1 week on a single charge. All participants are required to only use this lighter every time they smoke throughout the study. Sharing their lighter with other smokers or borrowing others’ lighter are both prohibited. This is
to ensure that all smoking events during the study are recorded correctly and the EMA survey in the app can be activated. Because according to the mechanism of the design, the app is wirelessly connected with the lighter, and it will pop-up a survey when it detects users’ smoking.

The fourth sensor is an electrochemical sensor called Smokerlyzer, shown in figure 5. It’s used to analyse the CO level in participants’ breath which is a biomarker of smoking status. Each examinee will use a disposable mouthpiece for infection control. The test needs 5 min to finish. Though the level of CO in the breath is correlated to both the number of cigarettes smoked during the past 24 hours and the time since the last cigarette, it usually exceeds 10 ppm among regular smokers. By contrast, normally the CO level among non-smokers is less than 5 ppm.

Apart from the annotations, the app will also be able to collect contextual information. This includes sound features (eg, volume), acceleration (activity), SMS/call log, system status, users’ schedule and GPS. The audio data won’t record anything about raw conversation contents, but only gather features, such as sound volume, to infer ambient circumstances. The GPS will be transformed and anonymised before being used to correlate locations with smoking episodes. Each of the above features will ask users for permission in the app, otherwise, the app is only used for annotation. The app is developed to reduce battery consumption as much as possible and can be used for at least one full day without charging.

Primary outcome
The primary and direct outcome of this study is a collection of diverse and representative data set about daily smoking habits of the general smoking population of...
office workers. Through analysis on these quantified data with reference of self-assessment report, it will lead us towards a more comprehensive and deeper understanding of smoking behaviour in the free-living environment, which, in turn, will facilitate the design of truly tailored and context-aware smoking cessation applications.

Data collection and management
Overall, data are collected through three channels, including website surveys, wearable sensors and smartphones. In addition, two platforms are used to store the data. Specifically, one Orient DB database is used to temporarily host streaming data from smartphones (ie, annotation EMAs and contextual data) and website (ie, questionnaires) information. Data in this platform are anonymised and will be ported into the local database once data collection is finished. The second platform is a local database used as the main storage containing both contextual data from the first database and physiological data retrieved from those two wearable sensors.

Besides, all data collected will be encrypted before being pushed to the cloud-based database platform. Access to both remote and local databases will be password-protected so entry is permitted to authorised users. Only the researchers directly involved in the study have access to the database, other external researchers can apply for access only to the preprocessed data where all identifiable and privacy-sensitive information is hashed or removed.

Analysis plan
The primary goal of the study is to detect patterns and triggers of smoking events. This is going to be addressed in two aspects. First and foremost, a significant effort will be put into physiological signal analyses to reveal quantitative biomarkers relevant to smoking. For example, several features can be extracted from ECG, such as the heart rate and HRV, which signifies stress level. In addition, GSR and temperature signal collected from the wristband sensor is believed to register affective arousal level, one important feature in Circumplex’s mood model.47 In fact, both stress and negative moods are commonly reported as trigger factors for smoking.

Second, contextual information will also be exploited to make sense the underlying circumstances which play an import role in triggering smoking. Having meals and drinking caffeinated beverages or alcohol are the most frequent events that accompany smoking in parallel or sequence. Generally, these two aspects are closely correlated and could act as a reference for each other. Furthermore, smoking timestamps collected by the lighters can be used as ground truth to pinpoint smoking episodes, which can be validated by hand-to-mouth movement using acceleration data from the wrist-worn sensor. Other context information, such as anticipating a meeting, may also impose an effect on the intrinsic feeling of stress in smokers, thus leading to smoking.

Apart from the analysis mentioned above, to answer our research questions, various modern data analysis methods, such as support vector machine, Bayesian networks, decision trees and eventually deep learning, will be exploited to fuse information together and draw inferences collectively. As the investigation needs to go beyond retrospective pattern analysis to a result that is more prospective and anticipatory, a regression analysis will be performed to examine the relationship between variables. Specifically, a multivariate multilevel analysis will explore the influence of smokers’ characteristics (sex, age, education, socioeconomic status, nicotine dependence, attitudes towards smoking, quit attempts, etc), contextual factors, and physical and emotional statuses on their smoking habits. A linear mixed model will be carried out to estimate whether any factor or pattern has a consistent correlation (p value<0.05) with smoking.

Patient and public involvement
Participant feedback received in the feasibility test on a previous version of the app was used to guide the design of the app used in the current study. Participants were not involved in the design of the study, research questions or recruitment. Overall burden and satisfaction of the study from the participants’ perspective will be assessed through the feedback survey. One week after the end of the study, participants will receive a plain English summary of the study outcomes showing basic information, such as the tracked number of cigarettes, number of filled-out surveys and so on.

DISCUSSION
As smoking cessation interventions are complex, its outcomes depend as much on the intervention itself as on participant characteristics and the context of intervention delivery.48 This could account for the efficacy variability shown in different smoking cessation applications, where users often have heterogeneous backgrounds and distinct habits. These differential factors are either not taken into consideration at all or significantly simplified which means any applications based on this assumption will often lead to varying results.

Given that smoking behaviour is rooted in the combination of contextual, sociological and psychophysiological factors, an efficacious intervention system should learn the habits of each participant and leverage the varying circumstances to deliver tailored feedback at the opportunistic moment. For this to happen, the collection of long-term smoking habits information with ubiquitous coverage in context and psychophysiology is a necessity and will be of great value to future work.

In this respect, hopefully, data and results from this study can both contribute to a better demonstration of the different smoking patterns among individual and general smokers, and empower a better design of prevention and intervention strategies through a comprehensive
understanding of the role of social circumstances and psychological factors in smoking.

Ethics
As data are collected from multimodal sources in the study, special attention will be paid to the appropriate storage and use of privacy-sensitive data. Data confidentiality will be guaranteed and deidentification will be implemented to make sure that no personal information will be disclosed.

Before the start of the study, participants are asked for their formal consent to the collection of privacy-sensitive data by signing a consent form. Additional permission is asked in the app for the collection of contextual information from their smartphones. If it is declined, the smartphone will only be used for EMA annotation. Participants can still stop any time during the study in case their mind changes.

Dissemination
In the future, any results based on this study will be published in international conferences and peer-reviewed journals.

Author affiliations
1 Connected Health Solution Group, IMEC, Leuven, Belgium
2 Department of Electrical Engineering (ESAT), KU Leuven, Leuven, Belgium
3 Connected Health Solution Group, Helst Centre, Eindhoven, The Netherlands
4 Health Psychology, Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium
5 Department of Neurosciences, Psychiatry Research Group, KU Leuven, Leuven, Belgium
6 Department of Electrical Engineering (ESAT), KU Leuven, Leuven, Belgium
7 Department of Neurosciences, Psychiatry Research Group, KU Leuven, Leuven, Belgium

Acknowledgements
The authors wish to thank Pauline De Wolf, Jozefien Speeckaert and Eva Steenberghs (IMEC) for their contribution to user recruitment and management in the study. The authors would also want to thank Matthias Stevens, Matthias De Geyter, David Vermeir and Robby Wauters (IMEC) for their work on the app and website used in our study.

Contributors
DZ: conception and design of the study, drafting of the protocol, design of the data collection and management system, development of the analysis plan and drafting of the manuscript. IVD and EV: revision of the study questionnaire and protocol. GS: conception of the analysis plan and revision of the manuscript. WD and CHW: conception and design of the study, and revision of the protocol. All the authors gave final approval of the version to be published.

Funding
The study is sponsored by an investigator-initiated grant from imec, a not-for-profit research institute. Imec has provided free Chillband and chest patch devices for study purposes. DZ is supported by a doctoral scholarship from KU Leuven of Belgium (000000008857).

Disclaimer
The funding sources had no role in study design; collection, analysis or interpretation of data; writing of the report; or the decision to submit the article for publication. Researchers were independent of the funders and sponsors of the study.

Competing interests
None declared.

Patient consent for publication
Obtained.

Ethics approval
Approval was obtained from the hospital ethical review board of UZ Leuven, Leuven, Belgium.

Provenance and peer review
Not commissioned; externally peer reviewed.

Open access
This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES
15. Sharmin M. Why do mobile phone-based smoking cessation interventions struggle and how can we make them more effective? University of Illinois at Urbana-Champaign, 2013.
44. Saadi DB. ePatch® - a clinical overview 2014:1–11.