Tracking Physical Activity: Problems Related to Running Longitudinal Studies with Commercial Devices

Daniel Harrison
UCL Interaction Centre
University College London
Gower Street, London
WC1E 6BT, UK
daniel.harrison@ucl.ac.uk

Paul Marshall
UCL Interaction Centre
University College London
Gower Street, London
WC1E 6BT, UK
paul.marshall@ucl.ac.uk

Nadia Berthouze
UCL Interaction Centre
University College London
Gower Street, London
WC1E 6BT, UK
nadia.berthouze@ucl.ac.uk

Jon Bird
City University London
Northampton Square
London EC1V 0HB, UK
jon.bird@city.ac.uk

UbiComp '14, Sep 13-17 2014, Seattle, WA, USA
ACM 978-1-4503-3047-3/14/09.
http://dx.doi.org/10.1145/2638728.2641320

Abstract
The problems with inactive and sedentary lifestyles are widely recognised. People believe that activity tracking systems, such as the *Fitbit*, may aid them in meeting recommended levels of physical activity. Similar systems have been the subject of previous research, but many of these studies were conducted over a short-term and some results may be attributable to reactivity or novelty effects. We ran a longitudinal mixed-methods effectiveness study using the *Fitbit Zip* activity tracker with 50 participants. In this paper we present two main challenges experienced during this study: the unreliability of the device and a lack of engagement by some of the participants. The issues we experienced can help inform the design of future studies.

Author Keywords
Quantified self; pedometer, activity tracker, personal device, fitness, health, Fitbit, behaviour change, physical activity, in the wild.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

Introduction
According to the World Heath Organization (WHO) approximately 3.2 million, or 6% of all, global deaths

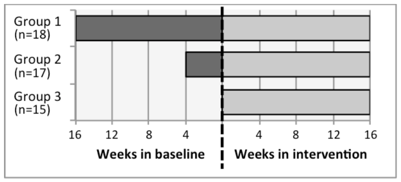

Figure 1: Study conditions. Groups 1 and 2 recorded baselines of 4 and 16-weeks, where they were unable to see how many steps they were taking. Participants in Group 3 did not record a baseline. All conditions were able to use the Fitbit normally during the 16-week "intervention" stage.

Figure 2: The Fitbit Zip activity tracker, showing a removed tamper-evident sticker. The sticker was applied to all devices used in groups 1 and 2, immediately after set-up. Each sticker was printed with the participant number to prevent participants replacing the sticker.

each year are caused by physical inactivity [7]. In addition obesity, which is strongly linked to physical inactivity, is a strong risk factor for many chronic diseases [5]. Physical activity policies and guidelines exist in many countries and are often accompanied with educational advertising campaigns aimed at increasing activity. These campaigns may be successful in increasing awareness, but many people struggle to meet recommendations [3].

Activity tracking devices, such as the *Nike Fuelband* or *Fitbit*, are often seen as systems which may aid in meeting activity guidelines or other goals [6]. However, studies of these, and other, behaviour change systems in HCI are often conducted over a short-term and as such their results may be attributable to novelty effects cf.[1]. In addition, after quantifying activity reactivity effects may play a role in short-term increases to activity levels [2]. One reason behaviour change studies in HCI have generally been quite short is that the devices utilised in research are often prototypes which may need high levels of maintenance. One potential solution is to use commercially available activity tracking systems in academic research. However, use of off-the shelf technology in longitudinal in the wild studies brings its own challenges.

Commercial devices

A wide-range of relatively inexpensive commercial activity tracking devices and smartphone applications are readily available. New and updated systems are frequently released with new features and functionality. The majority of these systems use accelerometers to infer and quantify, mostly ambulatory, activity. A large number of these systems offer application program interfaces (APIs) and software development kits (SDKs). APIs can allow researchers to collect (relatively) raw data from devices. Both APIs and SDKs give developers an opportunity to create software interventions using these devices, preventing the need to develop tracking hardware or use more costly research-oriented devices for in the wild research.

In addition to the research opportunities afforded though access to APIs and SDKs, commercially available activity trackers also embody a range of different implementations of behaviour change techniques (BCTs). Implementations of these BCTs have previously been studied within HCI and elsewhere in academia. These systems provide an opportunity to run longitudinal studies without the need for software or hardware development.

Fitbit Study

We conducted a longitudinal ethnographic study with 50 participants, each using a *Fitbit Zip* activity tracker. Our approach was to blend quantitative and qualitative methods, to identify behaviour changes and issues related to long-term use. Participants attended regular interviews (over 200 were conducted) and completed a diary throughout. In addition to these qualitative measures we used the *Fitbit* API to record step-data captured by the device. To compare reactivity and novelty effects we divided participants into three groups. Thirty-five participants in two groups recorded a baseline, where the device's screen was obscured (Figure 2) and participants were unable to see how many steps they were taking. In addition, participants could also not access their Fitbit account or synchronise with the mobile application. An overview of the study conditions is shown in Figure 1.

Problems experienced

Throughout the study we experienced complications, related both to the study design and technical issues with the activity tracking system. The Fitbit was chosen as it is a popular and well-reviewed device, which was reliable during a short pilot study. Four pilot participants undertook a shortened version of the study, which included baseline measurement, interviews and a diary. Longer-term use with a larger number of participants highlighted technical problems which were not apparent over a shorter term.

Study adherence

Participants in Group 1, who undertook a 16-week baseline record, experienced low levels of engagement. This manifested with participants not completing their diary or responding to requests for interviews, missing arranged interviews, regularly forgetting to wear the device and even losing the device altogether. To lessen these problems we moved affected participants from Group 1 (16-week baseline) to Group 2 (4-week baseline) where possible: removing the sticker from the screen earlier then planned, altering the study design and preventing reliable comparison between the different conditions.

Over the course of the study 14 participants lost their Fitbit. Nine of these participants chose to replace the device with their own money and the remaining five left the study early. The most common cause of a lost device was it falling off clothing when outside of the home or office. Problems with study engagement and adherence are not issues specifically related to personal informatics studies.

Hardware and technical issues

Almost one third (n=16) of our participants experienced technical issues with the device which prevented them from recording a complete set of activity.

The device is specified to store 30 days of data without synchronisation. Our study was designed to take advantage of this: we planned to meet participants every 28 days during the baseline period, to interview them and synchronise data with the Fitbit. However, in practice meeting participants every 28 days was not always possible, generally because participants did not respond to interview requests. This meant that interviews were sometimes not conducted until after 31, or more, days had passed. This unspecified use of the device often resulted in lost data: when the device was synchronized we would retrieve large numbers of days where 0-steps were recorded but the participant had been wearing the device and recording steps.

Data were frequently lost when synchronization was undertaken at intervals of less than 30 days. For example, the second interview with participant 10 was undertaken after 27 days of use, but data was only obtained for 4 days of use and all other data were lost. This problem did not affect all participants, but did result in large amounts of lost data at the beginning of the study. To counteract this participant were encouraged to regularly synchronise using the included USB dongle: they were still not able to access recorded data, but this lessened the effects of lost data.

Obscuring the screen also meant other technical issues went unnoticed. If the device was not working correctly neither the researchers, nor the participant, would easily be able to tell. Many participants experienced a problem where the device would repeatedly turn off and on throughout the day, drastically lowering the number of steps recorded and often preventing synchronisation. This issue was caused by a poor connection with the battery which was noticed by participants when they were using the device normally, but not during the baseline. Regular synchronisation allowed us to view the data and understand when issues were occurring. Fortunately this issue was more common later in the study, after the baseline record was complete.

Other participants experienced issues with the battery in the device depleting within days, rather than the specified 6-months. This also resulted in activity not being recorded until the battery was replaced. Because the tracking device was supplied by us, the majority of participants saw us as a first line of support for technical problems and also expected us to provide replacement batteries when required.

An additional problem we encountered was that we had no control over availability of the Fitbit system, which was sometimes unavailable during interviews. An additional problem with this lack of control is that updates to the system could potentially change BCTs and implementations, impacting participant usage and the study design. Fortunately no large changes were made during the study.

Discussion

The approach taken in this study allowed us to conduct a longitudinal in the wild study, but was not without limitations. One of the advantages foreseen of using a commercial device was that of time saved dealing with technical issues, but this was not the case. Researchers undertaking similar studies should be aware of the limitations laid out and realise that most commercial systems effectively act like a "black box", where it is not easy to troubleshoot or control the system.

Acknowledgements

Daniel Harrison is funded by an EPSRC DTG studentship. Work partially supported by: EPSRC EP/G043507/1 grant: Pain rehabilitation:E/Motion-based automated coaching.www.emo-pain.ac.uk

References

[1] Brynjarsdottir, H., Håkansson, M., Pierce, J., Baumer, E., DiSalvo, C. & Sengers, P. (2012). Sustainably unpersuaded: how persuasion narrows our vision of sustainability. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI 2012). ACM, New York, NY, USA, 947- 956.

[2] Clemes, S. A., & Deans, N. K. (2012). Presence and duration of reactivity to pedometers in adults. *Medicine and science in sports and exercise*, *44*(6), 1097-1101.

[3] Craig, R and Mindell, J. (2009) *Health Survey for England 2007*. London: The Information Centre.

[4] Harrison, D., Bird, J., Marshall, P. & Berthouze, N. (2013). Looking for bright spots: A bottom-up approach to encouraging urban exercise. Habits in HCI workshop, BCS HCI 2013.

[5] OECD (2013), *Health at a Glance 2013: OECD Indicators,* OECD Publishing. http://dx.doi.org/10.1787/health_glance-2013-en

[6] Rooksby, J., Rost, M., Morrison. A., and Chalmers, M. (2014) Personal tracking as lived informatics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI 2014).

[7] World Health Organization (2009).*Global health risks: mortality and burden of disease attributable to selected major risks*. Geneva.