Health Telescope: System Design for Longitudinal Data Collection Using Mobile Applications

Bas Willemse[1,2,3](✉), Maurits Kaptein[1,2,3], Nikolaos Batalas[3,4], and Fleur Hasaart[5]

[1] Jheronimus Academy of Data Science, 's-Hertogenbosch, The Netherlands
[2] Eindhoven University, Eindhoven, The Netherlands
[3] Tilburg University, Tilburg, The Netherlands
{b.j.p.c.willemse,m.c.kaptein}@tilburguniversity.edu, nikolaos.batalas@gmail.com
[4] RoboticBit, Eindhoven, The Netherlands
[5] CZ Zorgverzekeraars, Tilburg, The Netherlands
fleur.hasaart@cz.nl

Abstract. This paper describes the process of developing the technical infrastructure of the Health Telescope: an interventional panel study designed to measure the long term effects of eHealth usage. We describe the design and implementation of both the Health Telescope application—an Android application that allows us to interact with participants and obtain measurements—and the researcher authoring client—a web-based application that allows us to flexibly submit experience sampling tasks to participants. This paper serves as a blueprint for those wanting to study long-term behavioral change in the wild. The paper furthermore describes a pilot study that was conducted to evaluate the research software. We conclude with design guidelines aimed at those aiming to undertake a similar endeavor that are vital when developing similar software; this paper aims to highlight both the importance and challenges of measuring the effects of eHealth applications longitudinally.

Keywords: eHealth · mHealth · Longitudinal data collection · Interventional panel study · Wearable research · Technical guidelines

1 Introduction

Physical activity has decreased globally over the last 50 years [14]. In parallel, the last 50 years have brought forth a lot of research on health and physical activity. Given the developments in recent years, with new technological advances such as smartphones, wearable bands, smart scales, and many other examples entering the consumer market, we can now analyze health behaviors at an unprecedented level by generating high amounts of data on activity and physical health.

J. Ye et al. (Eds.): MobiHealth 2020, LNICST 362, pp. 145–165, 2021.
https://doi.org/10.1007/978-3-030-70569-5_10

However, research utilizing modern technology like cellphones and activity trackers still has some challenges: for one, long-term engagement with this technology has proven challenging [8,24]. The accuracy of devices is often inconsistent [6] and not validated [15]. Measuring the effects of usage may stretch beyond simply measuring step count for some period of time, and changes in health behaviors are made in different ways. These factors, among others, make it hard to conduct large-scale longitudinal research using activity trackers to measure participant physical activity. Furthermore, developing the technical infrastructure for such research can be challenging, as the technologies used are complex.

This paper describes the development of the system design and technical implementation of the Health Telescope, an interventional longitudinal study investigating the effects of eHealth app usage. With this paper, we aim to show the process through which the technical infrastructure for the Health Telescope project was developed. The paper covers the thoughts going into building the system, the details of the architecture, the evaluation process, done through a pilot study, followed by the improvements made after receiving feedback. These details on the development process may be interesting for readers trying to set up longitudinal research using mobile phones to generate data, or to communicate to participants. It can also be informative for researchers interested in using data generated in the project.

In the Health Telescope study, we measure the activity and mood of $N = 450$ participants for a minimum of four months. During this time, participants will be monitored and recommended to download existing applications for health behavior change. This study is significant for the field of eHealth for the following three reasons: (i) carrying out intensive longitudinal research contributes to helping the general understanding of the effects of eHealth application usage, as well as health behavior change; (ii) uncertainty exists regarding how to fully capture the effects of eHealth apps, in this study a broad range of information such as daily activity, sleep, mood, and phone usage is collected, allowing a unique, detailed look into the effects of usage of the apps distributed in the study; (iii) different approaches for health behavior change can help different users. In the study, we will experiment with personalization of eHealth app allocation: by prescribing different applications to different users, we aim to learn if an eHealth app is more effective for users with certain characteristics (i.e. "elderly women", "20-year olds that wake up at 6 am", "participants that make intense use of their smartphone"). We can then test these hypotheses by recommending the applications to users with similar characteristics. The study objectives are further detailed in Sect. 2. A detailed setup for the study can be found in [26].

The remainder of this paper is structured as follows: the paper starts by introducing the background, objectives and architecture of the Health Telescope Project (Sect. 2); then, the implementation of the system created for the study is described (Sect. 3). Next, we show how the system usage allows us to accomplish use cases (Sect. 4). We furthermore expand on a pilot that was used to evaluate and improve the system (Sect. 5) and close off by discussing improvements and related work, as well as describing a set of guidelines following our development process (Sect. 6).

2 The Health Telescope Software

The infrastructure described in the paper was created as part of the Health Telescope project. The project goals inspired us to build an infrastructure that can communicate with study participants' cellphones, and capture behavioral data like step count and phone usage to investigate the effects of using eHealth applications. In this section, we briefly discuss the Health Telescope project, listing its goals and motivations. We do so to show the rationale behind the development process and infrastructure that we discuss in the remainder of the paper. Additionally, we describe the requirements that implementations will need to satisfy. We first do this through sketching the data flow needed for the project, followed up by use cases that help make the details of the architecture concrete.

2.1 Study Objectives

The system is designed as part of the Health Telescope project, an interventional panel study measuring activity data using wearables. The objectives of this project are as follows:

- O_1. We aim to investigate the effect of using eHealth apps focused on increasing physical activity on activity and mood, for a long period,
- O_2. We want to see if there is a correlated effect between short-term measures and long-term measures after interacting with eHealth apps,
- O_3. We want to test different ways of personalizing eHealth offerings.

To effectively accomplish these objectives, we chose to not simply design or test a single app, but instead investigate how individuals respond to existing eHealth apps that focus on different, distinct persuasive elements. We will recommend study participants different, existing eHealth applications that all focus on improving physical activity, using different behavioral change techniques. For more information on the reasoning and design of the study, we refer to our study protocol [26].

2.2 System Architecture

To enable the goals of the Health Telescope project, we developed a system that allows a researcher to directly communicate to study participants, through a mobile app, as well as have that app transfer data generated by participants to the researcher. The architecture for this system has been largely inspired by a previous application for conducting event and signal contingent experience sampling studies, TEMPEST [11]. Like TEMPEST, the Health Telescope architecture consists of 3 software components:

– A mobile application for the Android OS, available to participants via Google play store, which collects data in the background on app usage and location from the phone, and steps, activity, and heart rate from the wristband. It also receives and renders content as produced by the researchers, such as questionnaires or cloud messages.
– An authoring web application, to be used by the researchers for creating content and managing its distribution to the participants in the study.
– A server that stores content and collected data in a database, and allows information to be shared between the authoring web application and the mobile application.

2.3 Considerations in Developing the System

We intend to create an application that serves as an observer app: that is, it gathers data relevant to the study from a user's phone, and transfers it to the database. The Health Telescope app is meant to a) transfer wearable and phone data to the database, and b) help evaluate existing apps. Importantly, the purpose of the app is not to function as a behavior change app in itself. Instead, we would like to use the app to observe how active users are before, during and after using eHealth apps that will be recommended to them during the study. To accomplish this, focus is put on making sure communication and data transmission happen correctly, not on experimenting with incentivizing activity or ways of engaging participants.

To ensure that the system can perform the tasks required to accomplish the study goals, we used use cases as a measure of internal testing during development [7]. These use cases resemble queries that need to be run during the study. The value of use cases shows in two different stages of development: a concrete, step-by-step plan on how the system should function can help during implementation by showing explicit restrictions that the system will need to uphold. During evaluation, each action in a use case can be tested to effectively isolate and communicate about existing flaws in the software. We expand further on use cases by giving two examples in Sect. 4 as a way of demonstrating the system's flexibility, as well providing detail on our evaluation process.

3 Implementation of the System

In this section, we discuss the framework that was created for the project. Below, we will detail the structure and interactions of the mobile app and authoring client.

3.1 Mobile Application

The goals of the Health Telescope application are:

1. To serve as a tool for users to self-report on their mood;

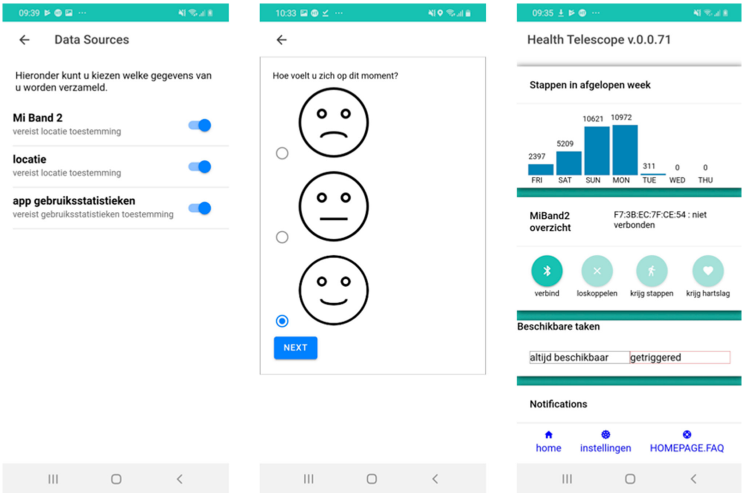

Fig. 1. Screenshots from the Health Telescope application. The first panel shows the privacy configuration for data collection, the second panel shows an example of a questionnaire loaded in the app, and the third panel shows the home screen, displaying the daily step counts, a wearable connection menu, and a notifications bar.

2. To gather GPS and phone usage data;
3. To collect sensor data from the wearable band, visually display this data, and transmit collected data to our servers;
4. To deliver interventions to participants; and
5. To monitor participant engagement.

The Health Telescope application is designed as a tool to observe participant behavior closely, without requiring active effort from participants. The data collection is completely passive, excluding the participant's self-reporting on mood, which is also designed to minimize the strain by making short multiple-choice questions. Figure 1 shows screenshots of different parts of the Health Telescope Application.

Data. The Health Telescope application collects a total of **six** different types of data: step count, heart rate, sleep, experience sampling, GPS location, and phone usage. Examples of these data can be seen in Table 1. We employ Experience Sampling as a tool for participants to self-assess how they are feeling throughout the week. To capture this, we investigated different constructs, such as well-being [17], happiness [21], and mood [28]. Out of these options, we chose mood and happiness. Mood is a construct chosen for its day-to-day variance, displaying direct effects of events happening in the life of participants. Different ways of evaluating mood exist, and one commonly used dichotomy fundamentally separates this mood into *valence* (the intrinsic attractiveness or aversiveness of the mood) and *activity/arousal* (expressing the calm or exciting nature of the mood) [5].

Table 1. Summary of data used in the Health Telescope

Data Type	Frequency	Example
Step Count	Every hour, the number of steps taken during that hour is measured	14.00–15.00 - 739 steps
Heart rate	Heart rate is measured on one hour intervals	14.00 - 73 bpm
Sleep Data	The start and end time of sleep are measured	February 23: 23:16–07:12
Experience Sampling	Up to once per day, we ask participants brief questions using push messages	February 23 14:00: I generally feel energetic: Yes I currently feel happy: Completely agree I feel good: Mostly agree
GPS Location	GPS location is saved every four hours	14:00: 38.8977N, 77.0365W
Phone Usage	We measure screen time and usage of applications on participants' mobile phones. Note: We only measure the duration of use, and do not in any way measure what happens within an application	Chrome: 14:03–14:04 Facebook: 14:04–14:17 Messages: 14:23–14:25 Mail: 14:25–14:37

Most existing surveys measuring mood in this way consist of 10–30 questions. Given our setup, it seems infeasible to regularly ask participants this many questions: participant burden is a large factor of dropout in longitudinal studies and questionnaires are lengthy. As a result, we directly ask for the underlying constructs of the questionnaires: one question asks participants about *activity*, one question asks participants about *valence*. A third question measures happiness using a three-point smiley face question, with a face for sad, neutral, and happy.

Technical Implementation. The mobile application is a hybrid of a web-based component and a native android component. The web-based component is an Ionic/Angular app, which renders the User Interface and also renders the questionnaires or other documents created by the researchers. The user interface consists of two major parts; one is a setup guide, for allow participants to log in and manage their system account, provide necessary permissions for data collection and pair their wristband with the phone; the other is the main screen which displays documents available for access, as well as an Inbox screen for notifications received so far. Notifications are Firebase cloud-messages that can consist of a single message body, but can also have documents attached. All documents in this environment can be any HTML/JavaScript content, which gives researchers the freedom to serve rich styling and interactions according to the needs of the study as it develops.

The native android component is tasked with downloading configuration and content created by the researchers, setting alarms for signal-contingent experience sampling, receiving Firebase cloud messages, collecting app usage statistics and location data, and connecting to the wristband via Bluetooth for sampling step count, activity data, and heart rate. Data is stored in the phone's SQLite3 database and is only uploaded to the server at opportune moments, as decided by the android OS's facilities for managing communications and power consumption. Additionally, Firebase cloud messages are not only notifications to be presented to the users but also commands to the application itself. This allows researchers to remotely control functions of the application, such as forcing the downloading of a new configuration, or the uploading of collected data.

3.2 Authoring Client

The authoring client is designed as a tool for researchers to carry out experiments. It accomplishes this goal by providing a simple interface for researchers to a) monitor enrolment and participant engagement; b) conduct experiments by setting up scheduled surveys and alerts; c) communicate to participants. The authoring client groups allow researchers to set up the following:

Participants. The authoring client allows researchers to monitor participants in their activity and engagement with the panel. While participant data can be directly accessed from the database, the authoring client provides a visual interface tracking data transfer and daily activity. The authoring client provides control of participant creation and deletion in the database.

Documents. The authoring client allows researchers to design any kind of content using HTML and JavaScript to create and serve rich styling and interactions as the study progresses. Specifically for questionnaires though, the application allows their creation using custom web components, which can be read in the semantic terms of the questionnaire's content and computational logic rather than the syntactic features of their HTML structure. The motivation and method of the approach for this particular feature can be found in [11]. It provides a live preview of the documents to prepare (see Fig. 2).

Allocation and Rules. The authoring client allows researchers to assign documents to (groups of) participants, using *rules* and *allocations*. Rules specify the conditions under which a document should be available for participants, and allows for researchers to set moments for the phone to trigger. Rules can be "always available", or "time-triggered". We can use the 'always available' setting for documents that users will need to have the ability to fill in whenever they want, as often as they want (tracking water consumption is an application for this rule). Time triggered rules, on the other hand, can be used when users should respond up to once per day. After completion, a document shown with a time-triggered rule will be hidden until the next day.

Allocations are the glue that holds everything together: an allocation is a triplet connecting *rules* specifying a timeframe, to a *document* that will be available for a (group of) *participants*. After an allocation has been created, participants will be able to see a document in their task list to be filled in, as well as receiving a system notification that they have a new task available (see Fig. 4). After completing a task, it disappears from the home screen of the app.

Communication. Furthermore, the authoring app allows researchers to send cloud messages to participants. Cloud messages are flexible: they can be used to "silently" contact the phone or can make notifications appear in the participants' Android notification screen. Cloud messages can contain the following: (i) a simple message, containing a head and body; (ii) a prompt for the participant phone to upload its allocation, forcing a new allocation onto the phone; (iii) a prompt for the phone to sample sensor data (as described in Table 1) to the server; and (iv) a notification that has a document attached to it. Cloud messages that contain notifications are stored and available to view for participants in the Messages tab of the app. In our setup, we use cloud messages to deliver interventions.

Technical Implementation. The authoring web-application was built using the Ionic/Angular framework. The interface components that are available in the authoring client belong to two main categories: lists and editors. Lists correspond to database tables and display the rows for a particular entity, such as participants, documents, rules, etc. Editors allow the researcher to create new such entities or edit existing ones, and they are generally custom UI components that are tailored to the properties of each. Both lists and editors offer options for accessing other related entities, e.g. a group editor can be opened from a participant list because a listed participant might belong in that group.

4 Testing the Software: Use Cases

In this section, we will demonstrate the flexibility and ease of use of the described system, by going over two distinct use cases and displaying step-by-step how these use cases can be implemented in the system. The use cases are as follows:

- **Use case 1 - Questionnaire**: Set up a questionnaire that asks every participant in the panel three questions daily for the next three months between 11:00 and 20:00.
- **Use case 2 - Intervention**: Send out a cloud message containing advice to download an interventional application to participants that walked an average of less than 7000 steps per day for the last 30 days.

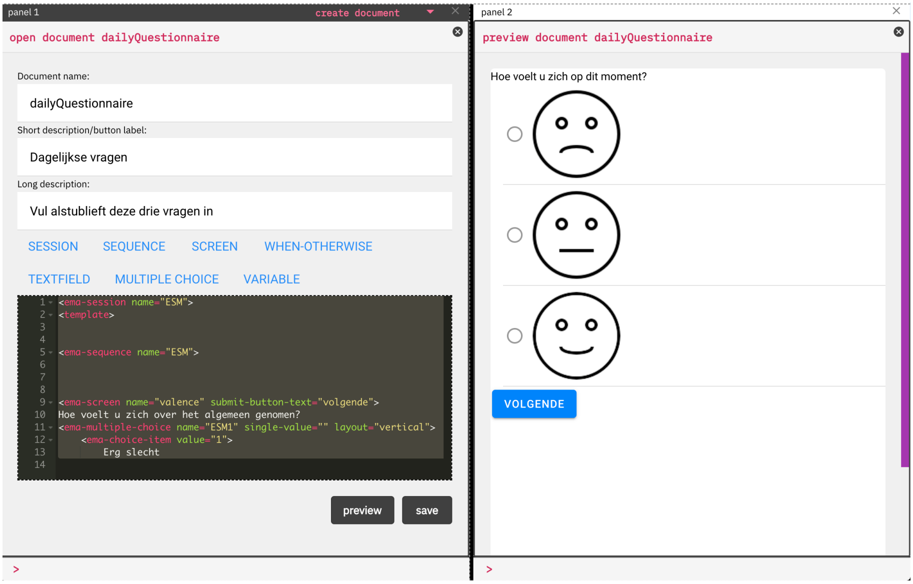

Fig. 2. Document creation interface. There are several pre-made buttons that create blocks of HTML code to create multiple-choice questions, create screen sequences, and more options. We are not limited to these pre-made buttons and could choose to manually insert other HTML. There is a live preview button, that renders the document in a separate pane.

4.1 Use Case 1 - Questionnaire

Here we describe how our system can be used to achieve the following steps:

1. Create a *document* in the authoring client to distribute to participants,
2. Select a *group* containing *participant* that will see the document,
3. Create a *rule* that makes this *document* appear within a time frame, and
4. Tie everything together by creating an *allocation*.

Figure 2 shows how a document can be created in the authoring client's *document creation* interface. On account creation, every new participant is added to the *'allParticipants'* group, containing the full set of participants, as part of the information sent from Google Scripts to the database.

Next, we create a rule that determines when the document is visible for participants. For this use case, we choose the latter, setting up a time-triggered rule. Here, we set the interval for the document to appear to 11:00–20:00, and enable the document for every day in the next three months.

After this, we have created the document we want to distribute, have chosen the time frame during which participants should see it, and have a group of participants that we would like to show the document to. To tie these three parts together, we set up the allocation by navigating to the *allocation creation*

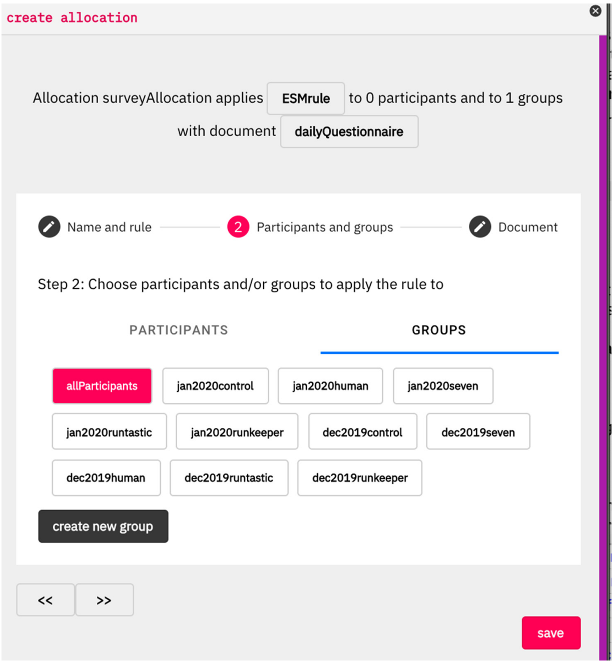

Fig. 3. Allocation creation interface in the authoring client. Researchers can use this to select documents to display for (groups of) users given some rule for when to show the content.

interface in the authoring client. The allocation interface (shown in Fig. 3) allows us to select the rule, participant, and document, which finalizes the use case. After taking these steps, every user will receive a notification, every day at 11:00, asking them to fill out a questionnaire.

4.2 Use Case 2 - Intervention

To complete this use case, we need to:

1. create a *group* of participants that will receive the intervention,
2. create a *document* containing the intervention, and
3. add the document to a *cloud message* that will be sent to the group of selected participants.

Use case 2 requires some work outside of the interface: we want to get to a point where we've identified the group of users that walk an average of fewer than 7000 steps per day, to feed this to the database. After this, we can create a *cloud message* with the advice, to send to this group. The first part of this can be done in any pre-made script, that connects to the database, selects the activity data for all participants for the last 30 days, then averages this out. From the

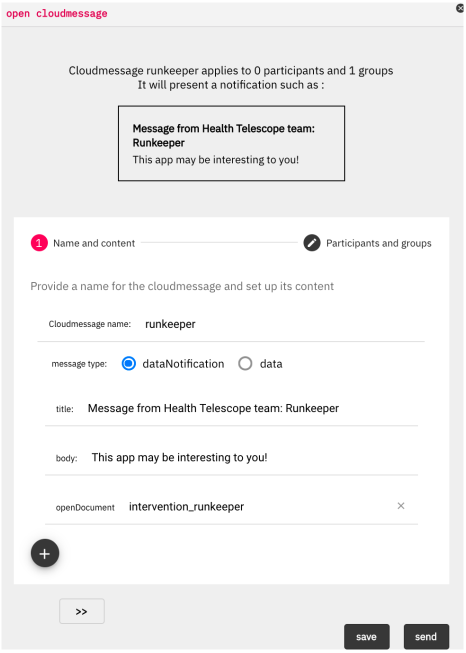

Fig. 4. Cloud message creation. Here, researchers can send notifications to participant cellphones, and specify the content for these notifications. There is a live preview of the notification at the top.

resulting list, select only the participants whose average activity is lower than 7.000 steps. Next, we add the IDs of these users that were insufficiently active to a group and use the database connection to create a new group in the database, containing this list of participants.

After these steps, the database contains the group of participants that will receive the intervention. We use the authoring client to create the cloud message. Cloud messages are messages appearing in the app's message box, that can be read back by users at any time. For this use case, we will attach a document to the notification, containing the intervention. We create this document in the same way we created the questionnaire in use case 1. Figure 4 shows how the message is given a title and body, that will appear to users as a notification, then add the document by selecting the key-value type openDocument, and selecting the earlier created message as our document. Then, we select the group that was created using the pre-made script as the receivers, and finally, we send the message. With this, the second use case has been completed.

5 Evaluating the System

Following recommendations for eHealth design in [20], part of the evaluation of the system before starting the study consisted of several pilot studies, testing elements of the study. Pilot studies can be used as an effective tool for testing experimental setups before starting a study or trial [1]. [30] argues that "There is a need for more discussion amongst researchers of both the process and outcomes of pilot studies". In this section, we aim to contribute to this discussion, by expanding on a short pilot that was conducted as part of the development of the app. Throughout development, three additional pilots have been held: one to test the technical functioning on a diverse set on smartphones; one to test the improved (after feedback from previous pilots) app, focusing mainly on user experience; and one pilot enrolling medical professionals that will be involved in recruitment, and were willing to learn more about the study by conducting a pilot. We aim to show how a small-scale pilot can help detect a wide range of necessary changes and improvements for a study.

A two-week pilot study was conducted, where participants were subjected to an accelerated version of the panel experience containing every element that participants will go through during the study. The goals of the pilot are to investigate the user experience of partaking in the panel, as well as testing the technical infrastructure as described in this paper. By testing the setup through a pilot, the developed software could be used on different phone configurations, and communication channels and data transfer can be validated. Below, we detail the design of the pilot, as well as providing a summary of the findings obtained in interviews that were held with pilot participants.

5.1 Pilot Design

The design for the pilot is as follows:

- **Intro survey.** Participants complete the introductory survey
- **Installation.** Participants download and install the application
- **Set-up.** Participants log into the app, set their data preferences, and set up the connection to the wearable
- **Daily survey.** Every morning, the participants receive a notification to fill in the three-question survey
- **Intervention I.** After three days, the participants receive a cloud message containing an app recommendation for an eHealth app
- **Intervention II.** After ten days, participants that have not installed the application recommended in Intervention I will receive another cloud message, containing a recommendation for a different health app.

The pilot concludes with individual, semi-structured interviews with each participant. During this interview, we have specific points on which we would like to gather information. Section 5.2 expands on these points.

Technical Functioning. One of the main objectives of the pilot was to test the technical functioning of the system. The pilot was held briefly after the first version of the software was developed, and while the system was tested on multiple phones during development, the use of Bluetooth, connecting phone and wearable, can function slightly differently for different phones. It is difficult to comprehensively test Bluetooth functionality during the development of an app, as Android Studio's virtual machine can not utilize Bluetooth. As such, connectivity and stability testing during development were limited - increasing the value of a pilot.

There are several specific points that we were interested in seeing during and after the pilot. The general stability of the software on different phone and OS configurations; the quality of the data transfer from phone to database; the functioning of our cloud messages, that should send alerts to users; the content of the data generated being in the desired format; the connectivity and general use of the wearable.

Users install the app through a platform called TestFairy. This platform provides valuable information for testing and developing software, by e.g. providing crash reports that help isolate technical issues.

User Experience. The other main objective of the pilot was to improve the user experience, by having a set of participants undertake the same actions that participants of the study will. While the decreased duration of such an accelerated version of the study may influence the perceived strain of participation, we believe such an accelerated path can be used to effectively find possible issues in user experience.

The interviews conducted after the pilot contained questions regarding the ease of enrolment, the general use of the mobile application, experiences in using the wearable, perceived strain of participation, privacy, and quality of the information provided to participants.

5.2 Pilot Results

This section describes the results of the pilot. We begin by describing the participant set, and split our findings up in two parts: first the technical functioning of the software, which had just finished the first cycle of development; secondly the user experience of participating in the study. We summarize the findings from the interviews and include relevant participant quotes.

Respondents. Recruitment for the pilot happened through one of the largest health-insurance companies in the Netherlands, respondents were all employees of this company. Seven respondents expressed interest in participating in the pilot. Two participants were unable to join due to not having an Android phone. One respondent changed his/her mind and did not participate in the pilot. The final pilot group consisted of four participants, aged 34 on average

(SD = 6.76). All participants were male. The pilot took place in Tilburg, Noord-Brabant, the Netherlands. Participants were given the information letter and informed consent form before participation, as well as a letter explaining the process of the pilot. These documents can be found at https://health-telescope.com/documents/. Participants all used Android phones and were given a Xiaomi MiBand 2.

Technical Functioning. The findings from the pilot and interviews are summarized below:

- **Stability of software.** On the first day of the pilot, two participants reported being unable to open the ESM questionnaires in the app. This issue got resolved within the next day, and participants were able to open the app, connect to the database, and respond to daily questionnaires. One participant noted that they were unable to open a questionnaire from their phone's home screen, but manually opening the app would make the questionnaire show up in his task list. This issue did not exist for the other participants.
- **Data transfer.** All of the data came into the database as expected: for each participant, activity, phone usage, and GPS data are periodically sent from the phone to the database.
- **Cloud messages and documents.** The cloud message functionality of the app showed clear issues: Intervention I was received by three out of four participants. Out of these three participants, two were unable to open the message. Intervention II was not received by any participant. One participant reported a measured step count of over 20.000 steps per day on average. This is likely a result of a malfunctioning wearable.
- **Wearable.** Participants reported that the wearables mostly functioned properly: there was one participant whose wearable reported abnormally high step counts, possibly caused by some malfunction in the band. Aside from this issue, the bands stayed connected and reported activity hourly.

User Experience. Participants noted that the introduction questionnaire that is asked on the website might be confusing, as it was written in English (as opposed to the native language of participants, Dutch). Participants noted that the questionnaire contains personal questions, but understood the significance of the questions asked. Excluding the technical issues that users experienced on the first day of the pilot, participants found the app easy in use. Three out of four participants found the wording used within the app clear and understandable. Participants thought the app UI was functional, but not visually attractive:

> *"I normally use a different wearable and am fairly spoiled with the app's dashboards. Because of that, I don't think this app is particularly attractive".*

Participants noted the visualization of the heart rate in the app's home screen is nice, but they hoped to see more: distance metrics, a visualization of their questionnaire responses, sleep analysis, and heart rate graphs were all suggested.

Participants found the effort of answering a survey every day to not be a problem, which they stated was because the questionnaires were very short and multiple choice. One participant pointed out that two years of answering the same daily questions could bore participants, and suggested changing the questions we asked periodically. Participants found the setup process to be well explained, but reported not knowing what to do after completing the setup, and trying to find actions that they were supposed to perform in the app:

> *"After the setup, I wanted to get more active, but I couldn't find any tips or activities in the app."*

When asked if it is realistic to expect panel members to participate for two years, the pilot participants responded positively.

6 Discussion

This section of the paper discusses the changes made based on the pilot, as well as listing related work that inspired us. Furthermore, we provide a list of guidelines that we believe are important to consider during development of longitudinal studies using electronic devices.

6.1 Lessons Taken from Pilot

The feedback received from the pilot allowed us to make the following changes to the system:

Technical Functioning. The pilot showed us important technical issues that needed to be worked out before starting the study. In the weeks following the pilot, various technical improvements were made that ultimately led to stability on, as far as we know, every Android phone version 7.0 and up. The pilot showed that the configuration of data gathering and the transfer was correctly set up. The issues concerning the allocation of cloud messages to different phones were traced down and resolved after the pilot. Lastly, one of the take-aways from the pilot is that while generally the wearable functions just fine, some wearables that we distribute may malfunction. It is imperative to detect this and replace the wearable. To allow for this, we now include a specific point on malfunctioning wearables and replacement in the participant briefing of the study. Participants can contact us if the wearable reports incorrect step counts, breaks, or gets lost, to have the wearable replaced.

User Experience. Participants indicated that the amount of effort that participation took was reasonable, and could be done on a longer-term. One participant described it as following:

"The effort needed to participate is not too high. I think it's reasonable to expect people to participate for longer periods. Also, if I were to enroll in a university study, there'd be some motivating aspect of contributing to science that helps me stay engaged".

After the pilot, we made significant changes to allow for easier understanding of the setup process: texts were clarified and made simpler; and more emphasis was put on the voluntary nature of the introduction survey, to ensure participants are aware they do not need to fill in every question. The pilot also allowed us to proceed with the planned rewards for participation: participants can earn their wearable over time, and we will host guest lectures relevant to the general understanding of eHealth that participants will be able to follow. Feedback from the pilot showed that participants did not fully comprehend the intended purpose for the app: despite information material outlining the interventional apps that are supplied during the study, some participants were under the impression that the Health Telescope app would provide them health advice. After hearing this, we have made significant changes to the briefing, putting more emphasis on the functions of the mobile application and communicating that the eHealth apps recommended during the study are what is being tested. Additionally, the introductory texts that guide participants through the setup now include at multiple points that there are no immediate actions for users after finalizing the setup.

6.2 Related Work

Here, we would like to highlight related research that guided us in this work. As mentioned before, the TEMPEST architecture, which contributed to the Health Telescope system, is described in [12], and the Health Telescope protocol is found at [26].

There exists a large amount of work on tracking physical activity. Before mobile phones became widely spread, these studies often had participants report on their physical activity [16,29]. These studies however often utilize self-reporting and surveys to measure activity/BMI. A possibly more reliable alternative used in longitudinal studies is to actively test their participants' physical condition periodically through exercise tests [18].

The rise of wearable technology may be able to provide researchers with a very direct way to measure activity. Usability and acceptability studies have been done on a wide range of participants and features [13,23,27], which can inform researchers when a target audience or outcome variable is chosen. Research analyzing the attrition rate of wearable use [2,8] should be considered in the design process. Additionally, research investigating the accuracy and reliability of data created by wearable devices [6,10] can help choose an appropriate device for a study, given the important trade-off in features, comfort, and accuracy for different features.

Wearables can be utilized in studies investigating physical activity in several ways: the reported activity can be used as an outcome variable [3,19]; notifications received on a wearable can be used to motivate participants [32], and there is a wide range of studies investigating individual wearables and their effect [25,31,33]. Regarding study design using wearables for physical activity, we advise researchers to use the informative design recommendations in [20]. Additionally, researchers should look into how wearables and their accompanying apps incorporate behavior change [22].

In designing the database and data sharing policies, we drew inspiration from various longitudinal Dutch studies, that we note here for their data sharing policies: the Lifelines cohort study [4] is a large-scale longitudinal study following the physical activity and weight in non-obese people. The Longitudinal Internet Study for the Social Sciences (LISS) panel [9] is a longitudinal study sampling the Dutch population and conducting periodical surveys and experiments. These studies may be interesting for researchers looking for longitudinal datasets.

6.3 Design Guidelines

As part of our findings, we would like to present a set of design guidelines, based on the experiences of developing this technical infrastructure:

Design and Implementation of Infrastructure

- When designing the technical infrastructure of a project, researchers should choose to measure data that is appropriate for the study objectives. As an example, for research regarding activity, step count is generally more accurate and reliable than heart rate, especially at higher intensity activity.
- It is important to consider the details on how data should be collected, transmitted and saved. We recommend regularly gathering the desired data, and periodically, encryptedly transmitting this data to a secure database.
- We recommend careful investigation of factors both in user experience such as comfort and battery life, as well as research done on the accuracy and reliability of the data generated by the instruments (e.g. wearable) considered.
- We strongly encourage researchers to consider privacy and participant rights while designing the study, to ensure the infrastructure is set up in a GDPR-proof manner. As an example, it is necessary for the system to allow for individual participant data excerpts, and full deletion of data belonging to individuals. These rights should be considered in the design.
- Design the setup process of a study or app in a way that informs participants of the purpose of the app, and actions they should take. As an example, we discovered through feedback that users were confused on their next actions after the initial setup, and considerably changed the information provided to users during this setup to inform them of the possible actions within the app.
- When implementing software of this sort, it is important, besides data collection, to collect data on how the client application functions, e.g., when a cloudmessage was received, when something was downloaded from the server, etc. This information can be helpful for isolating technical issues (e.g., if data is missing, is it because the application is not capturing that particular source

of data, or is the device perhaps turned off?). There are ways in which participants use their phones which can account for what data gets collected, which are not necessarily obvious from looking at the data of the experiment itself. The Health Telescope app does record such events into a part of the database that is separate from the data itself.
- Minimizing strain is important in long-term studies. Attrition rate is traditionally high in these studies, and efforts should be made to engage users, make participation simple, and clearly communicate expectations to users.

Testing

- We believe it is beneficial to set measurable, concrete goals for the system functionalities. We worked with specific, detailed use cases, that proved of great help in streamlining our repeated testing process by isolating technical issues.
- We recommend to test the software on multiple phones of different models and Android distributions during development. The stability of software often varies for different phones, in factors such as accessing specific data types, setting alarms, Bluetooth functionalities, etc.
- Bluetooth is commonly used in wearable studies. This further emphasizes the need to test the software on multiple phones, as it is not possible to test Bluetooth connection for different Android distributions through Android Studio's virtual machines.

Evaluating

- We emphasize the importance of feedback: try to involve feedback from as many people as possible in the development process. This can be colleagues, or potential participants, depending on the goal of the interaction. Pilot studies are an effective way of involving different individuals to gather feedback, and test the software on multiple devices.
- When designing a pilot study, think carefully of the things you hope to take away from it, and use this information in your design: Should the technical stability be tested? Include as many phones as possible, and set up the pilot to include these technical interactions, or stresstest the system. Should the pilot test the burden of participating? Focus on ways to capturing the user experience, by closely monitoring their actions and giving them opportunities to provide feedback.
- We advise to use systems such as TestFairy that track bugs and technical issues during testing. These systems can be used to quickly identify the specific issue when a technical problem occurs.

6.4 Conclusion

In this paper, we presented the development process of the system designed for data collection during the Health Telescope study. By detailing the steps in development, we aim to contribute to longitudinal research using smartphones and wearables. We gave a detailed look at the system architecture, and showed how it can be used to accomplish the study interventions. We displayed the role that pilot studies can play in the development of a system and showed the changes made following this pilot.

We concluded the paper by providing a list of guidelines for researchers, that we hope aid researchers interested in setting up longitudinal studies using electronic tools in developing their study designs and technical infrastructure. In doing so, we aim to show the value of sharing important lessons from developing work like ours. We encourage colleagues to share their work, to contribute to lowering the complexities in setting up longitudinal studies.

References

1. Arain, M., Campbell, M.J., Cooper, C.L., Lancaster, G.A.: What is a pilot or feasibility study? A review of current practice and editorial policy. BMC Med. Res. Methodol. **10**(1), 67 (2010)
2. Attig, C., Franke, T.: Abandonment of personal quantification: a review and empirical study investigating reasons for wearable activity tracking attrition. Comput. Hum. Behav. **102**, 223–237 (2020)
3. Benedetti, M.G., et al.: Physical activity monitoring in obese people in the real life environment. J. Neuroeng. Rehabil. **6**(1), 47 (2009)
4. Byambasukh, O., Vinke, P., Kromhout, D., Navis, G., Corpeleijn, E.: Physical activity and 4-year changes in body weight in 52,498 non-obese people: The lifelines cohort (2020)
5. Carson, T.P., Adams, H.E.: Activity valence as a function of mood change. J. Abnorm. Psychol. **89**(3), 368 (1980)
6. Case, M.A., Burwick, H.A., Volpp, K.G., Patel, M.S.: Accuracy of smartphone applications and wearable devices for tracking physical activity data. Jama **313**(6), 625–626 (2015)
7. Collins-Cope, M.: RSI-a structured approach to use cases and HCI design. Manuscrito no publicado (comunicación personal), Ratio Group Ltd. (1999)
8. Coorevits, L., Coenen, T.: The rise and fall of wearable fitness trackers. In: Academy of Management (2016)
9. Das, M., Knoef, M.: Experimental and longitudinal data for scientific and policy research: open access to data collected in the longitudinal Internet Studies for the Social sciences (LISS) panel. In: Crato, N., Paruolo, P. (eds.) Data-Driven Policy Impact Evaluation, pp. 131–146. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-78461-8_9
10. El-Amrawy, F., Nounou, M.I: Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? Healthc. Inform. Res. **21**(4), 315–320 (2015)
11. Batalas, N., aan het Rot, M., Khan, V.J., Markopoulos, P.: Using tempest: end-user programming of web-based ecological momentary assessment protocols. Proc. ACM Hum.-Comput. Interact. **2**D(EICS), 3:1–3:24 (2018). Redacted for anonymization

12. Batalas, N., Khan, V.-J., Franzen, M., Markopoulos, P., aan het Rot, M.: Formal representation of ambulatory assessment protocols in HTML5 for human readability and computer execution. Behav. Res. Methods **51**(6), 2761–2776 (2018). https://doi.org/10.3758/s13428-018-1148-y. Redacted for anonymization
13. Gimhae, G.-N.: Six human factors to acceptability of wearable computers. Int. J. Multimed. Ubiquit. Eng. **8**(3), 103–114 (2013)
14. Hallal, P.C., et al.: Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet **380**(9838), 247–257 (2012)
15. Henriksen, A., et al.: Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J. Med. Internet Res. **20**(3), e110 (2018)
16. Herman, K.M., Craig, C.L., Gauvin, L., Katzmarzyk, P.T.: Tracking of obesity and physical activity from childhood to adulthood: the physical activity longitudinal study. Int. J. Pediatr. Obes. **4**(4), 281–288 (2009)
17. Hills, P., Argyle, M.: The oxford happiness questionnaire: a compact scale for the measurement of psychological well-being. Personality Individ. Differ. **33**(7), 1073–1082 (2002)
18. Janz, K.F., Dawson, J.D., Mahoney, L.T.: Tracking physical fitness and physical activity from childhood to adolescence: the Muscatine study. Med. Sci. Sports Exerc. **32**(7), 1250–1257 (2000)
19. Kurti, A.N., Dallery, J.: Internet-based contingency management increases walking in sedentary adults. J. Appl. Behav. Anal. **46**(3), 568–581 (2013)
20. L'Hommedieu, M., et al.: Lessons learned: recommendations for implementing a longitudinal study using wearable and environmental sensors in a health care organization. JMIR mHealth uHealth **7**(12), e13305 (2019)
21. McGreal, R., Joseph, S.: The depression-happiness scale. Psychol. Rep. **73**(3_suppl), 1279–1282 (1993)
22. Mercer, K., Li, M., Giangregorio, L., Burns, C., Grindrod, K.: Behavior change techniques present in wearable activity trackers: a critical analysis. JMIR mHealth uHealth **4**(2), e40 (2016)
23. Profita, H.P.: Designing wearable computing technology for acceptability and accessibility. ACM SIGACCESS Access. Comput. **114**, 44–48 (2016)
24. Quiñonez, S.G., Walthouwer, M.J.L., Schulz, D.N., de Vries, H.: mhealth or ehealth? Efficacy, use, and appreciation of a web-based computer-tailored physical activity intervention for Dutch adults: a randomized controlled trial. J. Med. Internet Res. **18**(11), e278 (2016)
25. Randriambelonoro, M., Chen, Y., Geissbuhler, A., Pu, P.: Exploring physical activity monitoring devices for diabetic and obese patients. In: Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers, pp. 1003–1008 (2015)
26. Willemse, B.J.P.C., Kaptein, M.C., Hasaart, F.: Developing effective methods for ehealth personalization: protocol for the health telescope, a prospective interventional study. JMIR Res. Protocols **9**, e16471 (2020). Author redacted for anonymization
27. Ridgers, N.D., et al.: Wearable activity tracker use among Australian adolescents: usability and acceptability study. JMIR mHealth uHealth **6**(4), e86 (2018)
28. Ryman, D.H., Biersner, R.J., La Rocco, J.M.: Reliabilities and validities of the mood questionnaire. Psychol. Rep. **35**(1), 479–484 (1974)

29. Telama, R., Yang, X., Viikari, J., Välimäki, I., Wanne, O., Raitakari, O.: Physical activity from childhood to adulthood: a 21-year tracking study. Am. J. Prev. Med. **28**(3), 267–273 (2005)
30. Van Teijlingen, E.R., Hundley, V.: The importance of pilot studies (2001)
31. Veerabhadrappa, P., Moran, M.D., Renninger, M.D., Rhudy, M.B., Dreisbach, S.B., Gift, K.M.: Tracking steps on apple watch at different walking speeds. J. Gen. Intern. Med. **33**(6), 795–796 (2018)
32. Wang, J.B., et al.: Wearable sensor/device (Fitbit One) and SMS text-messaging prompts to increase physical activity in overweight and obese adults: a randomized controlled trial. Telemed. e-Health **21**(10), 782–792 (2015)
33. Yavelberg, L., Zaharieva, D., Cinar, A., Riddell, M.C., Jamnik, V.: A pilot study validating select research-grade and consumer-based wearables throughout a range of dynamic exercise intensities in persons with and without type 1 diabetes: a novel approach. J. Diab. Sci. Technol. **12**(3), 569–576 (2018)