Making sense in the long run: long-term health monitoring in real lives

Jochen Meyer, Elke Beck, Merlin Wasmann, Susanne Boll
OFFIS Institute for Information Technology
Oldenburg, Germany
{firstname.lastname}@offis.de

***Abstract*—**Long term self monitoring with connected personal health devices offers tremendous opportunities for wellbeing, health, and prevention. However, to date it is not fully understood how users perceive monitoring in the long term and how they implement it in their daily lives. We observed 7 participants that used a comprehensive set of connected personal health devices for 9 months and inquired their opinions and experiences. Users varied broadly in how they used the devices and how they engaged with the collected data. Implementation and use of long-monitoring evolved over time, leading to subtle but distinct differences to short term use. We found five relevant use cases: behavior support, improved self understanding, identification of trends and relations, decision making, and data collection for future use.

Keywords—* *Health tracking; self tracking; wellbeing; long term interventions, explorative field study

I. Introduction

Self observation and monitoring of parameters of health such as weight, blood pressure, or physical activity is ever since one of the basic tools for wellbeing, health, and prevention. Since a few years personal health devices are available that are connected to the internet and send their data to an online service where it is stored, analyzed, presented, and related with other values. This opens the door to a comprehensive self monitoring of health and wellbeing that (a) allows continuous monitoring of parameters of health integrated into the daily life, (b) covers a multitude of facets of health using combinations of different devices simultaneously, and (c) can be conducted over not just weeks or months, but over years and decades.

The new opportunities have turned self-monitoring from a tool for disease management to a matter of lifestyle and personal health and wellbeing. The users are not just patients or persons at risk, fitness enthusiasts, or number freaks, but also average healthy persons seeking general support in their daily lives. While a lot of research has been conducted on many aspects of self-monitoring, it is not yet fully understood how people realize and perceive self-monitoring in the long term, e.g.: How do people choose to monitor themselves in the way they do? How do they integrate the different devices in their lives? And what are differences to short term uses?

We therefore conducted the longitudinal "Lotus" (LOng Term Use of Smart health devices) study, observing the users' self-monitoring behaviors with a comprehensive set of devices over an extended period of time, and inquiring their experiences, perceptions, and opinions. Our findings provide insights into opportunities and challenges for long term self monitoring in real life.

II. Related Work

Since the 2000s technology is investigated and used to support a healthy lifestyle as. Early works such as UbitFit Garden [1] usually focused on inducing a behavior change towards a healthier lifestyle, implementing short term interventions with durations of often up to three months. Later, other opportunities to support wellbeing beyond behavior change have been investigated.

A. *Long term self tracking*

There is no clear definition of when short term use of trackers ends and long term use starts. Harrison et al. called a 32 week study using Fitbit a "longitudinal" study [2], whereas Merilathi et al. assumed a use duration of 43-99 (average 83) days in a feasibility study to be "long-term health monitoring" already. In an in-the-wild study Karapanos et al. [3] differentiate between initial use and prolonged use, without quantifying the borderline between the two. The current agreement seems that durations well beyond 6 months can indeed be considered long term.

Based on interviews with 30 users of activity trackers Fritz et al. [4] suggest design implications for long term support, including motivation to maintain gains already made, as well as to continue to change, and the need to support changes in activity and metrics. Further Karapanos [3] found that, in the long term, the user's thoughts about the behavior become less important as the incentive to continue; rather, the decision to continue is driven more by improved physical health and by social relationships.

Li et al. have put self-tracking into the broader context of "Personal Informatics" [5] aiming to "help people collect and reflect on personal information", not just related to health, but also to e.g. finance. The model has later been expanded and clarified, e.g. omitting the action phase for persons focussing on self-understanding [6], or identifying two more reasons for self-tracking beyond an action, namely instrumenting an activity, and curiosity [7]. Li et al. later found [8] that users ask six types of questions about their personal information: status, history, goals, discrepancies, context, and factors.

B. *Use and abandonment of activity trackers*

The long term potential of technology is lost by early abandonment of tracking tools. Activity trackers are particularly well investigated, and they allow us to get clues for other tracking tools. A market survey found that one-third of all activity trackers are abandoned within 6 months [9]. Various research studies report similar or even higher abandonment rates, such as an abandonment of 70-80% of all devices within two to four weeks in study settings (e.g. [10]).

There is growing consent among researchers on reasons for abandonment. Lazar et al. [10] found three categories: a misfit between devices and participants' self-conceptions, the collected data not being useful, and too much effort for use. Epstein et al. [11] found six themes among the reasons: cost of collecting, cost of ownership, discomfort with information, data quality concerns, learned enough, and change in life circumstances. And Clawson et al. [12] found various detailed reasons, including expectation mismatch, technical complexity, and goal met. Lazar also found reasons for ongoing use: perceived usefulness, novelty and curiosity, hope for potential future use, and developed routine of use.

C. Monitoring multiple health facets

While monitoring of many, albeit not all facets of wellbeing with smart devices is well supported, this imposes further challenges: The user is required to mix and switch between different systems [13], usually from more than one manufacturer. This means different user interfaces which implies major usability challenges [14]. This also results in distributed data, which requires integration and interpretation of different data sources [15]. Some, although not really much, research has focused on this topic.

Lane et al. have investigated multi-faceted wellbeing in the BeWell+ system [16]. However they focused on extensive manual logging and on feedback mechanisms, not on the sensing itself, for which they used a smart phone as a proof-of-concept. This technology has been shown not to be sufficient for long term self monitoring [14].

In the Health Mashups system [15], Bentley et al. investigate users' information needs when presented with multiple streams of health data. For sensing, a combination of an activity tracker and a scale was used in combination with manual logging. The system allowed some simple analytics to understand one's data and seek relations. The study showed that the users could integrate the system into their lives and were able to increase their awareness for health.

III. APPROACH

Research has provided numerous detailed and valuable insights into use and opportunities for connected health devices. However, most research has addressed relatively short use periods of a few months, and many have addressed the use of just one device, often an activity tracker. Long term monitoring is seen as a challenge and opportunity but has rarely been directly addressed, even less when taking into account the effects of using multiple devices and the interferences and dependencies appearing between them.

We focus on this spot by addressing a long term self monitoring of health conducted over an extended period of time in average persons' daily lives with multiple monitoring devices simultaneously. Our goal is to understand the users' perception of health monitoring in the long term as the fundamental basis for adoption and compliance. We are interested in two main questions: What are meaningful usage scenarios and the reasons and considerations why users ultimately would undertake long term self monitoring? And what affects and impedes long term adoption and makes people stop monitoring themselves?

This work inevitably replicates many of the existing work on health self monitoring, but puts it into the new context of long term monitoring with multiple devices. We will therefore focus on the characteristics of long-term monitoring and on the similarities and differences to short-term use.

A. General Approach

Inspired by the "n=1" approach [17] for understanding personal data traces, we conducted a qualitative study with quantitative backing, closely observing a small group of users in their self monitoring behaviors. The users were chosen to represent what might be considered "normal" end users, i.e. persons in normal working age that are neither technical nor medical experts and are without extensive self monitoring experiences. We gave them a comprehensive set of three up to date self monitoring devices but left it to them which devices they would ultimately choose and use. The study duration was chosen to be long enough to overcome the novelty effect and get into a long term routine of use: the results presented here have been collected after 9 months.

B. Study set-up

1) Participants

To ensure stable qualitative results we aimed for at least 6 participants, taking into account possible drop-outs. Inclusion criteria were: age 25-75, basic technical knowledge, fulfilling technical prerequisites for using self-monitoring tools (e.g. owns a recent smart phone), general interest in health, willingness to try out new technologies. Exclusion criteria were chronic diseases, limiting physical activities, and far-above-average experience in health or technology. Participants were recruited by a third-party recruiter. They received a small financial compensation (approx. $50) for the efforts involved.

2) Technology

We equipped the participants with three health devices for tracking various aspects of healthy living: A Fitbit One or Fitbit Flex for physical activity and sleep, a Withings body scale for weight, and a Beddit for sleep monitoring (see Fig. 1). To avoid overloading the participants with too much technology at once, the participants initially used the Fitbit for sleep monitoring (if they wanted), and we handed them the Beddit after three months only. The data collected was integrated into a study portal using the APIs provided by the producers, where the participants could access and we could analyze it.

Fig. 1: Trackers used in the study: Fitbit Flex, Fitbit One, Withings body scale, Beddit sleep monitor (photos: by the manufacturers)

3) *Study procedure*

The study was conducted in Germany, and all questionnaires and interviews were in German. We addressed the participants seven times (T0 to T6 meetings), either face to face (T0, T1, T3, T6), or online only (T2, T4, T5), in intervals of about 6 weeks, resulting in an overall study duration of 9 months. In every meeting the participants also filled out a questionnaire whose results are not reported here.

In the initial T0 meeting we gave the participants the activity tracker and the scale and guided them through the set-up procedure. We introduced the only task that they were expected to fulfill: to use the devices in their daily lives as they saw fit; they were not requested to pursue a specific lifestyle or achieve a goal. The participants filled out two baseline questionnaires: The TA-EG [18] is a 12 point questionnaire measuring affinity to use technology. And the International Physical Activity Questionnaire [19] assesses the level of physical activity; following the evaluation guide we classified the users as low, medium or very active. Moreover we conducted a short interview.

The T1 interview was mainly meant to solve problems. We also asked questions about usability and general impressions, but didn't expect much new due to the novelty effect.

In T3, after 3-4 months of regular use, we asked in more detail about their experiences with the use of devices, and started to address the impact from the use devices such as insights into personal health gained from using the devices, general usefulness of self-monitoring, or possible changes in health behavior and motivation. The questions were chosen based on our previous experiences with self-monitoring. The interviews were semi-structured; the participants were invited to speak freely about their thoughts and ideas.

We conducted an interim qualitative analysis of the results from the T1 and T3 interviews to identify an initial set of topics for subsequent consolidation with the participants. Four experts (computer scientists with experiences in HCI for health) identified clusters of relevant topics. Also taking into account the experts' knowledge about self tracking, we identified the following 9 topics relating to devices and their use, to the data collected, and to applications of data for health and wellbeing: How were the devices used in daily lives? How could the design of the devices be improved? Which data and which timescale did the participants find relevant? How much did the participants trust the correctness of the data? How did the participants engage with the data? How would they want to utilize the data for medical, wellness and long term purposes? What where the perceived effects of self-monitoring on oneself? What was the impact on and by third persons? And what is the significance of sharing and privacy? We developed guiding questions for each of the nine topics to stimulate discussions.

The T6 interview was the main tool to gain actual insights into long term self-monitoring. It was conducted as semi-structured 1:1 interview where we discussed the 9 topics with the participants. In preparation of each interview we also examined the participant's individual use of the devices and discussed specific observations with the participants. Each interview lasted between 60 and 90 minutes and was conducted by the same two researchers, one talking with the participant, and one taking extensive verbose notes. Additionally the interview was recorded. The recordings were subsequently partially transcribed to verify and complete the verbose notes and separated into a total of about 700 individual statements, between 70 and 130 per interview.

C. *Data analysis*

The verbose notes from the interviews were qualitatively analyzed, following an analysis model for identifying and applying categories in textual content by Mayring [20]. Two computer scientists with long-standing experiences in HCI and health related research, one communication scientist with long-standing experiences in HCI and qualitative research, and one undergraduate computer scientist with experiences in health research, did the analysis. We used an inductive approach to identify categories (major themes) from the individual statements.

Initial candidate themes were suggested by the two interviewers by walking though the textual content. Each of the four researchers then independently categorized all statements for two of the interviews (i.e. two experts per interview) and either assigned it to one of the candidate themes or marked it as "doesn't fit". In a first consolidation meeting all assignments were jointly reviewed and clustered, and a next iteration of candidate themes was suggested. The process was repeated twice, with a new allocation between researchers and interviews. At the third consolidation meeting saturation was reached, i.e. all researchers agreed that the participants' statements could be distinctly assigned to one of the major themes. In a next step, each researcher paraphrased and generalized all sentences for two interviews. In a final consolidation meeting, the four researchers jointly identified sub categories for the major themes based on the paraphrases and generalizations.

IV. RESULTS

A. *Demographics and data overview*

The study involved 4 female, 3 male persons. All participants were German. Age was 30-56 (mean 42.6, stdev 11.1). Affinity to use technology was above average, but not excessively high (30-40 of max 48 points, mean 36.7, stdev 3.7). The participants had varying levels of physical activity: 2 low, 2 medium, 3 high. Four participants had experiences with pedometers or tracking apps. Reasons to participate were, amongst others, curiosity about the devices (6 of 7 persons) and interest to learn about one's health (4 of 7), with just one person indicating the intention to change towards a healthier living as a reason.

Fig. 2 shows number of days that the participants used the different trackers. Most persons used the activity tracker most often and the sleep monitor least; however, there are individual differences. User P4 delivered practically no data, although he/she reported extensive use in the interviews. We suspect undetected technical issues.

B. *Qualitative findings*

Our analyses of the T6 interviews revealed five major themes and their topics (see Fig. 3). These themes partially confirmed our preliminary findings from the interim T3 interviews, in they relate to devices, data and information, and applications. They also considerably refine those findings, they distinguish between primary and secondary use of collected tracking data, and they introduce a new theme, general objections to self-monitoring.

Most of the topics are in principle known from prior work, but, in the context of long-term monitoring with multiple devices, they reveal new insights that we present below.

1) Usability of devices and services

This theme refers to the act of collecting data with a device and the interaction with that device and the related internet service for that purpose.

In line with findings from previous research (e.g. [14], [21]) **user experience** issues such as preferences for design and appearance, wearability in daily lives, and robustness were prevalent for activity tracking. Sleep monitoring with the Fitbit activity tracker as a body-worn device was not well accepted. The Beddit sleep monitor that was installed in the users' bed was valued for integrating well into the bed room and being invisible, and the figures show that some participants used it regularly. On the other hand participants also noted that users may have various sleeping places during the week (P4, P5), hence the stationary approach of the Beddit was not best suited for them. Nevertheless these participants didn't use body-worn sleep monitoring, either.

While the individual findings are known, combined they confirm our assumption that a comprehensive long term self monitoring requires multiple devices for multiple health facets – here: sleep –, and sometimes even multiple devices for the same facet in multiple situations – here: sleep at different places.

Most of our study participants reported different **technical problems**. These related to installing the software (P2), to ongoing synchronization (P1, P2, P5, P6), and to collecting data (P6). Our participants were able to cope with the issues and while the glitches impeded self-monitoring, they did not completely prevent it. However it is unclear what would have happened in an unobserved setting.

Our study participants quickly developed individual routines of **different intensities of using the devices** that they pursued quite consistently. The activity tracker was described as a daily companion (P6), being part of basic equipment (P1)

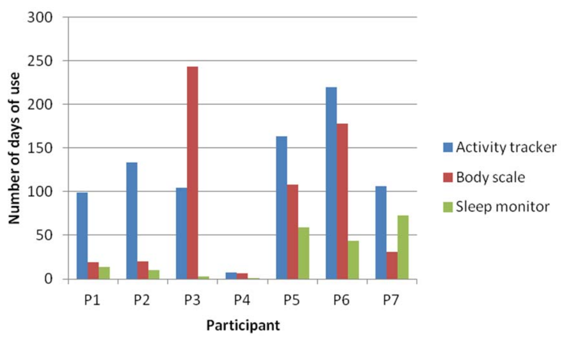

Fig. 3: Use of trackers per participant

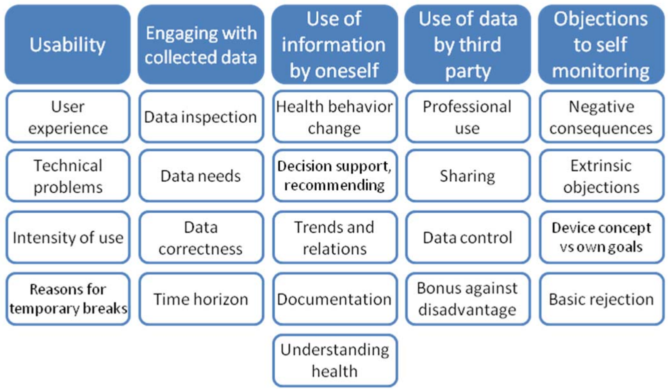

Fig. 2: Themes and topics

and being worn even during the night (P2) (but interestingly not used for sleep monitoring). Participants were creative in finding their routine of use; e.g. P4 wore the activity tracker in the wallet. Use of the body scale ranged from daily use (P3), weekly use (P7), quitting after a phase of regular use (P2, P6), to non-use since 6 months (P1). The sleep monitoring device was least used among the study participants. Some participants used the device occasionally (P1), while others refused or weren't able to use this kind of monitoring (P2, P3).

Again the results per device are not particularly surprising. However in combination they reveal an interesting finding: As can be seen in Fig. 2 there is no recurring "standard combination" of intensive and less-intensive used devices. Also we didn't observe a "super-user" using every device in maximum intensity. Rather every user chooses what suits her or him best. Long term monitoring therefore results in a heterogeneity of devices and in huge varieties of use intensities.

All participants took **temporary breaks from collecting data**, not just as part of their routine, but also as additional and irregular lapses. Reasons for breaks – e.g. forgetting to wear a device, not using to protect the device from loss or damage, suspending due to unwanted information – relate to known reasons. Breaks also happen due to non-mobile devices such as the fixated Beddit sleep monitor or the scale that could not be used while on travel.

Regarding continuity and long-term use it is interesting that we did not find changes over time, neither in the users' reports, nor in the data: Users do not "fade out" of the use of devices, reducing compliance over time. Rather they use it consistently also over a long period of time.

In our context taking breaks relates to single devices only; usually other devices were continued to be use. Reasons for breaks therefore seem to relate to individual data or devices, not to self-monitoring or health in general.

Finally, we found that users would resume use even after breaks of considerable length, e.g. many weeks. So while abandonment is often discussed one may ask whether in the long term there is really an ultimate abandonment of self-tracking, or whether any abandonment should be considered an extended break only, which would open up new opportunities to encourage users to re-use trackers.

2) Engaging with and understanding the collected data

This theme describes how the study participants dealt with their collected data and how they turned these data into knowledge. Now the collection of data using the devices has finished and the focus is on the data only. In this sense the data is independent of the device that was used to collect the data.

Our participants deployed different **types and frequencies of data inspection**. Confirming previous findings [15], [22] most interviewed users preferred the apps on the smart phone (P1, P2, P3, P6, P7) and partly the sensor device (P1, P4, P7) to access and view their collected data over the website or the study portal, pretty much in line with the concepts of glances, reviews, and engagement described by Gouveia et al. [22] for use durations of a few weeks.

Our findings add a long term perspective: In initial use the glances on the device were frequent e.g. to get to know the amount of steps for a specific activity (P1). For some participants, such as P5, this habit shifted over time to an occasional data inspection or even a disinterest in the data leading to hardly any inspection (P4): "*I don't need to have a look at the fitness tracker. That's like an investment portfolio: buy it, leave it, keeping it in mind.*"[1]

Our study participants also had **varying data needs**; they differed in their interests for specific types of data. Primary data that the devices offer are activity, sleep, and weight. The data was available at different levels of granularity, ranging from an overall score (e.g., sleep score) over individual features (such as sleep length, sleeping times, step counts, kilometers, calories) to single measurements (e.g. heart rate, breathing rate, snoring). Most participants (except P5) stated a prime interest in activity data. However while the other data was in general not interesting for all users, on the other hand virtually all data was interesting for at least some users.

Viewing the collected plain data without any further explanations of correlations (causal relations) might not be sufficient for some participants. P1, for instance, missed the "because" in his/her data visualization, telling him/her "*You've slept better, because you've moved enough. That's the right direction; please go on in that way*." Some participants would like to have tailored interpretations of their data and be provided even with a diagnosis: "*I wasn't able to draw conclusions from my raw data. It said: My heart rate is 5 percent or 10 points too high. Well, what does that mean? The interpretation is missing. The diagnosis is missing!*" (P4)

Similar findings have been documented e.g. by Bentley et al [15]. However two findings are particularly relevant in our context: First, several users stressed the importance of covering all dimensions of health relevant for their current health interests, without agreeing on one set of relevant features, though. E.g. for weight control P2 missed primarily food intake, whereas P4 said that "*my heart rate, my blood pressure, all that needs to be gathered to get the best results*." And P6 also was interested in contextual data, such as weather data "*in order to identify relations later on, like 'I've been running less on that day, because it rained'*." Given the heterogeneity of the answers, a comprehensive health monitoring will therefore always require multiple but heterogeneous devices.

Second, the measurements would be used as an "early warning system". In this case the most interesting data related to exceptional events and to outliers. P5 and P7, for instance, showed an interest in monitoring their respiration. Knowing about occasions of respiratory interruptions and snoring was important for P5. Similarly, P7 explained that his/her respiratory values always range between two levels, but when the value is above he/she knows that something is wrong.

The **correctness of data** was also an issue for the study participants. In general, they strived for a completeness of data. They wore the device even if they knew that little activity will take place. Or, they manually added data to the platform if data was lost during synchronization (e.g., P3). On the other hand while they regretted that some of their activities hadn't been tracked (such as bike rides mentioned by P3), they showed some tolerance with regard to missing or incorrect data, as long as the following conditions are given: First, the reasons for inaccuracy must be clear to the users. Second, as long as tendencies in the data were visible, our participants could cope with "inaccurate" measurements. Finally, incorrect data was less of an issue if participants (e.g., P6) were able to infer a health status from other correct data. Tolerance with incorrect data therefore is related to being able to interpret the data. This point is known, e.g. Consolvo et al. point out the importance to "give users proper credit for their activities" [21]. However, it should be noted that the ability to interpret is limited e.g. by time (external circumstances are forgotten after a few months) or, as we will see below, by use case (e.g. giving data to a third person that doesn't know the context). Somewhat paradox, correctness might therefore be more important in the long term than in the short term.

In the interview, we queried the participants about **interesting time horizons** when viewing the collected data. They had differing preferences ranging from high interest in past long term data (P1) to no (further) interest (P3, P7). P7 did not inspect data from the past because he/she considered the data of such a high importance, and thus knowing the values by heart, even over months. Those participants who were interested in accessing long term data viewed data from particular periods of time (quarters and beyond (P5), one or several weeks (P1, P6), from two days ago). For P6, additionally the start of collecting data was interesting. P5 explained: „*For me, the most exciting period was the first quarter, 3 to 5 months, because then one can discover how things are*." (P5). We will discuss the consequences of these findings below.

3) Use of the information by oneself

This theme relates to the effects that the self-monitoring had on the participants and how they used the gained information for the management of their health. The topics that we found are partially overlapping with previous findings. E.g. they refine Epstein et al's [7] three motivations for self-

[1] The participants' quotes in this section have been translated to English by the authors.

tracking (behavior change, instrumentation, and curiosity) and can be related to, but are more application related than Rooksby et al's [13] five styles of personal tracking (directive, documentary, diagnostic, collecting, and fetishised). However in their entirety they provide a new view on potentials of long-term self-tracking.

Many study participants used the tracking devices and their collected data to achieve personal health goals, such as a **health behavior change**, although initially only one participant had reported an intention to do so. Self-monitoring can act as an extrinsic trigger for obtaining (temporally) beneficial health-related behavior. It can support behavior retention, in as much as it can increase the motivation to get physically active, provide mechanisms to set health goals and evaluate their attainment. For instance, P3 stated that "*the tracking had a purpose: it nudged me to do more physical activity*." For P2, who wore the activity tracking device every day, it "*was super exciting. It was very motivating to move more*."

If health goals are met, the provided data seem to support the participant's perceived self-efficacy. E.g. P3 "*thought at the beginning, it would be very exhausting to walk that much. But then I discovered: it isn't! And so I've decided to go for a walk with the dog for at least half an hour each morning*." Some interviewed users also reported that self-monitoring helped them to reach an intrinsic, permanent change of behavior. With this, supporting a behavior *change* becomes the support of a *permanent* behavior, which makes sense particularly in the long term.

Understanding oneself and improving one's health knowledge is another effect of self-monitoring. Self-monitoring helped participants to reach an improved health self-understanding. P1 and P2 pointed out that they were not aware of their low daily amount of physical activity, unless the activity tracker enforced an awareness of the problem.

Many participants learned about the (positive) effects of physical activity and their own health-related behavior. It encouraged them, for instance, to reflect about phases of well-being during phases of poor health and to interpret their health-related situation on basis of the collected data. Furthermore, participants showed an interest in self-monitoring because it enabled them to compare their lived experiences with objective data. For instance, P1 was able to get an explanation for her/his degree of well-being by consulting the collected activity data, "*it's very valuable to establish a link between what I do and how I feel*."

Moreover, there is an interest in **identifying trends and relations** in data among our study participants. Continuous self-monitoring helped them to recognize changes in health-related data over time. P1 liked to use the sleep monitoring device, "*because one could get aware of the quality of sleep by reading the messages. At some point in time, the messages changed, the sleep evolved*." This is again a topic that is particularly relevant for long term applications.

Aside from learning about changes in health-related behavior, **recognizing connections between different kinds of collected data or events** is important for participants (e.g., P1, P2) too For instance, they established links between well-being and physical activity, sleep and weight, or life situations and health-related behavior. For instance, P2 could relate his/her life situation to increases of his/her weight, and P1 said: „*Especially when there was insufficient physical activity, it was important for me to relate that with events. For instance, I was on a conference and did not move much. That needs to be explicable for me, in order to get it balanced again*."(P1).

Some participants were overwhelmed by further interpreting the collected data, such as establishing connections between data. P4 said that he/she did not do any further investigations because it would be too complex: "*health experts are able to analyze such data much easier and faster than me*." P4 further criticized that the tracking devices and services did not help with analyzing and understanding the collected data and would prefer to get an (automated) diagnosis / analysis.

The interviews have shown that users would like to have **decision support and recommending**. They would prefer to be further supported in making health-related decisions and plans. They would favor device features such as the provision of daily feedback in form of recommendations and precise instructions for actions. Moreover, being reminded by the device to get physically active was also perceived as positive. P4 thinks that the tracking devices are "one-dimensional" because the user provides his/her data, but doesn't get any feedback (in form of recommendations for actions). Data gained in long-term monitoring could enable more fine-grained, personalized recommendations, in as much as it can adjust the recommendations to intra-individual changes over time.

At some point of time in device use, some users discovered that they pursued a **documentation of their life** but quit consulting their data. Participants like P5 were solely collecting the data for life and health logging purposes at a later stage of use, clearly a use case for long term monitoring.

4) Use of the collected data by third party

Given that users of self-monitoring devices have collected data – for whatever reason – for some time in the past, this theme relates to possible uses of the data by third party now or in the future and beyond the immediate effect on oneself. In the interviews, our study participants discussed passing their collected data to other people and how this would enable possible uses of their data.

Most participants can imagine a **professional use** of their collected data, i.e. giving the data to a health professional for an improved medical diagnosis and personalized advice. Providing data to the physician serves as a basis for analysis by the physician and should ease the physician-patient talk.

This might seem similar to what is frequently called "patient-generated data". However there is an important nuance: At the time of collecting the user is not a patient, but – in our scenario – a healthy person. The data is therefore not collected in a patient role. Using the data for medical purpose is a secondary use of data only, with restrictions on the applicability of the data for that context. As P1 notes: „*There* *needs to be ensured in advance: Are the data relevant for the particular case? And is the medical person, despite classical medical training, able to draw conclusions? The data would be suitable for lifestyle and wellbeing issues, where precision and completeness are less important. For medical issues, more precise analyses would be needed*." (P1)

Consequently professional use in this context relates more to prevention: E.g. P2 and P5 think of the collected data as being useful for preventive counseling, because it supports tracing the health status over a longer time and "*one could immediately see the origin of an illness when consulting the data*" (P2). Furthermore, users (like P6) see benefits in self-monitoring for an individualization of health programs.

Some interviewed users can imagine involving others **by means of sharing**. For P2, being a role model for others and setting a good example can be more plausible when sharing data about changed health-related behavior. While most users were rather skeptical about sharing personal health data in public, sharing with like-minded people was attractive to some users (e.g., within sports groups). These findings are somewhat in contrast to Karapanos et al. [3], who found social relatedness as a driver for long term engagement. One reason may be that Karapanos et al. investigated activity tracking, which is more open to competition and comparison, whereas our focus is broader, covering general health which may be considered more personal and intimate by the users.

Keeping control over their data was a big issue for the participants. For instance, P1 states that "*I have the feeling that I don't have any control over the data. Confidentiality is important to me, these are my data*!" Some participants were nevertheless concerned about directly providing their collected data to third parties, such as doctors or public health insurance companies, because these groups might "*use such data against me some time*" (P2). P6 also thought so, and therefore would prefer to give her/his data to a trustee if the compliance is ensured.

Some participants would agree to provide their **data to public health insurance companies** in order to get a bonus or better scale of charges. For P1, this is a tricky situation, because „*usage of the data for bonus options is great, but further usage to adjust dues is critical*."

5) Objections to self-monitoring

While the themes 3 and 4 focus on reasons *for* self-monitoring, this theme addresses general objections *against*, i.e. reasons that hinder a self-monitoring. The topics relate to known reasons for abandonment. However abandonment implies previous use, whereas these objections also point to reasons for not even starting. It should be noted that three of the four topics are independent of the device or data, but address objections against any kind of self-monitoring.

One type of objections that we found comprises **perceived negative consequences by tracking**. Some participants perceived risks caused by using tracking devices in their daily life. For example, P4 stated that there would be an information overload if the devices would provide her/him masses of health related data. Some interviewed users also noted general negative effects due to device use, such as the device acting as an authority which determines whether a behavior is problematic or not: "*It may well be that I sleep fitfully, but I don't really want to know that. Because I would say I don't have any sleep problems. And thus I don't want any device say 'on the contrary, you sleep very fitful'.*" (P3)

Another type of objection to device use is related to external sources, which we frame as **extrinsic objections**. A device-refusing immediate environment, such as skeptical family members or relatives, e.g., who have privacy concerns, may hamper the use of the devices. For instance, P1 installed the sleep monitoring device in the bed room, but P1's life mate found the device a bit suspicious because of the active microphone for monitoring snoring; "*my life mate is vehemently against storage of data, a definite no!*"

Or, some users have concerns about adequate data protection, either done by themselves (not being able to secure their data) or by third parties (e.g., risk of data theft if their data is given to public insurance companies). P7 explained that when he/she measures his/her blood pressure traditionally, no one can witness it, but when using a device which is connected to the Internet, "*someone can realize it*".

Furthermore, the **mismatch between device concept and own ideas/goals** is an objection to self-monitoring. If the required activities for using the device hindered the participant to pursue one's goals, the participant tended to reject the device. For instance, P3 stated that if he or she wants to reduce weight and tries hard not to think about food, he or she will reject a device which requires to pay attention to calorie intake, such as food diaries: "*If I had to note everything about my nutrition, I'd think about eating all the time.*"

Another objection to device use is a **basic rejection of quantification**, of measuring and quantifying one's health-related behavior. Despite her/his initial interest in trying tracking devices, P5 concludes after half a year that s/he doesn't „ *have to document anymore and wouldn't miss it too*" and "*either [physical exercise] has an effect on my well-being or it does not. I wouldn't need all the documentations and measurements to know that in the end*." Users like P5 see no need to document their habits, but prefer to rely on their own experiences, such as feeling well after physical exercise.

V. Discussion

Our findings provide highly interesting insights into long term self monitoring and reveal differences to short term monitoring. We subsequently describe the changes of use over time, find the need for short term rewards also in long term use, identify four main use cases for long term monitoring, and discuss the boundary conditions.

A. Three phases of monitoring

The way how devices are used is changing from short to long term use. We observe three main stages.

In the first weeks the users were very emotional about the devices and reported intense engagement in the T1 interviews. They were enthusiastic about the devices and data, but also very critical about points they didn't like. And they struggled with the challenges and problems evolving from technology or trying to integrate the devices into their daily lives. Our study setting influenced and encouraged them and gave guidance how to go through that phase, otherwise, we suspect, at least some users would have abandoned use of the devices or have not even started.

The second stage, which, given the schedule of our interviews, we observed in the T3 interviews after 4-5 months, was characterized as a phase of learning and exploring. In this time many users were on a quest trying to make sense of data. They aimed to understand and explain what's going on and were seeking relations between the different parameters. While they sometimes were disappointed by the lack of easy relations and overwhelmed by the complexity of the topic, they in general took their lessons learnt and continued to work with their findings. Making sense of the data remained an important topic also in later use, although the urgency reduced.

In the third stage the way of using devices changed in a subtle but perceivable way. Users had developed their routines of using the devices and had learnt to interpret the devices' data. They still regularly checked their devices, but often said they only did so only to compare their subjective perceptions with an objective measurement.

While the first and second stage can vaguely be related to Li's stage-based model of personal informatics [5], the third phase has characteristics of phases identified later by Li et al [8], namely maintenance and discovery: The users went into a routine use that often involves collection only, but usually not resulting in immediate actions. The discussion on use cases for long term monitoring below identifies five potential actions for immediate as well as delayed actions.

Correctness of data, which is one key issue in usability of devices [14] and was also reported in the early phases by our participants, became less important, and users were more tolerant to faults. They understood that gaps and errors in the data happen, and they developed strategies to cope with these faults. Users had learnt to interpret the data and found their ways to cope with imprecise data, such as the mis-estimation of cycling by activity trackers, although it still bothered them.

The demand for feedback by the devices changed, but continued to be important. Quickly checking the measured values on the device or glancing into the apps became the most frequently used way of interacting with the devices, often to compare subjective perceptions with the objective monitoring. Portals were less frequently checked, as were the more in-depth analyses provided by the apps; these were often considered too cumbersome to use and of little added value. However in times of exceptions and changes, looking more careful into the data became important again.

Major technical and usability problems were either solved by the third stage, or the respective devices had been abandoned. Users had found their ways and developed their routines of dealing with the specifics. However, a surprising number of glitches and issues continued to persist, such as updates to the apps and software that annoyed the users, or technical failures that remained unnoticed in the normal routines of use.

Our findings complement the aforementioned research on long term tracking [3], [4]. They confirm that long-term monitoring differs from short-term monitoring, although e.g. social relatedness identified by Karapanos et al. [3] or similarly appropriate social networks by Fritz et al. [4] does not play a role in our setting. The common ground is that the way of interacting with devices and data changes after some time. Long term interventions must be aware of the differences and make sure to support them appropriately to increase long term compliance.

B. Short term rewards versus long term advantages

We observe a somewhat paradox relation between the wish for short term rewards and the long term reasons for use. Users primarily observed the recent data of the past few days and rarely looked at historical data. And they requested short term findings explaining e.g. their current health status based on last week's behavior. Asked directly for advantages of long term self-monitoring, the users seemed a bit clueless and saw not many reasons. Some even suggested deleting collected data older than a few months.

On the other hand, users also reported interest in analyses that are only available in the long term, such as identification of changes. They took motivation from past achievements and demanded recommendations based on past behaviors.

To take advantage of long term monitoring it is therefore necessary to keep the user engaged in the short term too. Systems must provide short term rewards as a prerequisite to exploit the advantages evolving from long term use.

This complements the suggestion by Fritz et al. to "[support] changes in activity and metrics" for long-term applications. While we agree with that, it is also necessary to support activities and metrics that are relevant in the present to ensure engagement in the future. Some examples are implemented already, such as the Withings scale providing weather information to encourage users to regularly weigh themselves.

C. Use cases for long term monitoring

Ultimately long term monitoring must provide a benefit. For short term monitoring this is often a health behavior change. For long term monitoring the answer to the question "why would you do this" is less obvious. Our results clearly showed five main use cases, relating to behavior support, to health literacy, and to future use.

(1) *Supporting health behavior:* While this may seem to be the most obvious use case, the long term use differs from short term use in that a behavior *change* is usually short term, whereas long term focuses more *sustaining* an existing behavior. Considering this to be the maintenance phase of a behavior change, as framed by Fritz ("Motivating maintenance as well as change") might be too narrow: Sustaining may also happen without a previous change, and most of our participants left most of their behaviors unchanged, but nevertheless found self-monitoring valuable. Feedback from the devices is particularly useful to calibrate the user's perception (e.g. about activity) with an objective monitoring, particularly in exceptional, non-daily situations. Designs for such interaction would need further investigation.

(2) *Improved health understanding:* Self-monitoring may improve the understanding of health topics, both with respect to general aspects of healthy living, and to the individual and highly personal view of one's own health. The possibility to personalize the learning process about health with the learner's individual and health situation provides tremendous opportunities to improve the understanding of patients and healthy persons. Given the high relevance of health literacy for healthy populations this is an extremely powerful, although long-lasting approach to increase a population's health.

(3) *Trends and relations:* Users were eager to understand what's going on, how the different health parameters relate to each other, and how they change over time. Bentley et al. discussed challenges of analyzing and understanding complex health data already [15]. Persons would also use the data to identify slow and otherwise unnoticed slow changes, using self tracking as an "early warning system" for their health. This early warning system may be one of the predominant reasons for long-term self monitoring: Life changes as people age, and people change with that. It requires a high awareness and good knowledge to identify slow changes, and it's just too easy to neglect the negative impacts of e.g. gaining just 300 grams in weight per year. While identification of slow changes is possible with long term self monitoring, more research would be needed to design a system that reasonably presents this information to the user.

(4) *Making informed decisions:* users need to make decisions about their health, and they see that their data can help them with that. Here decision support can be passive – the users review their data and decide themselves – or active – a system provides analyses and makes recommendations for the users. Decision making may relate to medical decisions, i.e. using the collected long term data as one part of a health assessment. In this case quality and reliability of data are undoubtedly issues. Decisions can also be personal decisions, e.g. when looking for ways to cope with changes in life and using personal experiences and success stories from the past as encouragement and basis.

(5) *Storing data for some yet-unknown future use:* many users feel that the data that they collect now may have some still unknown use in the future, possibly also in a medical context. They therefore suggest storing data for this future use. This may relate to use case (3) and (4) above, but may – by definition – also be any other use. It's worth pointing out that this use case raises the fundamental question of data quality: While the users were tolerant regarding errors in the data when they were able to interpret it, this interpretation is not possible in the long term. The raw data that is valuable today may therefore lose its value in the future, and measures to ensure data quality are needed for a reasonable health log.

These five use cases should not be understood as being mutually exclusive. For each of our users there was not the single one reason, but it was a mixture of these five with different emphases.

While ultimately collecting data is not a purpose on its own and must provide a benefit, within the health context the benefit may well be considerably delayed until some possibly unknown time in the future. Not only is self-tracking possible in the long term, but it also provides some of its main advantages in the long term only.

D. Boundary conditions

While health obviously is a highly personal topic, it is still a bit surprising that users saw only very few opportunities for using the data beyond the individual use. Users had doubts that medical experts would be interested in the collected data. And they were very concerned about privacy and confidentiality of data. From the users' point of view self-monitoring, although clearly related to health, has only few overlaps with the classical medical sector.

Users reported objections against self-tracking that are very fundamental and probably difficult to address. There may be external reasons such as an environment that hinders self-monitoring. Or persons may have general reasons such as a basic rejection of quantification.

VI. Limitations

With its seven participants our study is fairly small. And while we addressed people of average age with average technical knowledge, today's health technology still requires above-average technical affinity and at least a basic interest in health to be used for an extended duration in daily lives; hence our group is probably biased. Nevertheless the saturation that we achieved when analyzing the interviews makes us optimistic that our themes are reasonable.

Within the limited setting of a relatively small study it is impossible to perform really long term observations covering years or even decades. This is a general challenge that researchers in technologies for wellbeing and life-long health are facing. We observed indications for an initial saturation effect between the interviews conducted after 4-5 months and those after 9 months: while opinions and perceptions for the latter provided more details and insights, the general trends per user remained stable. We therefore assume that major changes in the next months are not likely. Nevertheless we must be aware that after a considerably longer use time such as two years the users' behavior and perception might still change.

VII. Conclusion

Long term monitoring provides exciting new opportunities for the management of personal health. What so far was a playground for the data and technology enthusiast can now more and more be used by average persons who are interested in the results. Already now we can gain insights into our health that were not possible until a few years ago. However despite the technical opportunities, long term monitoring is still a major challenge. Usability in daily lives, data analysis, presentation and storage, applications in the overlap between personal wellness outside the tradition healthcare market and regulated medical interventions, design of long term interventions, ethics, privacy, data security and others are important topics for a successful deployment of long term monitoring in the real lives. Relevant use cases are not just in behavior support, but also in improving health knowledge, identification of changes, decision making, and data collection as a provision for the future. While technical advancements may make some of these topics obsolete there are still numerous open questions. Our work provides insights into the user's needs and requirements and contributes to answering these questions.

Acknowledgment

This work has been funded by the Hewlett-Packard Innovation Research Program.

References

[1] S. Consolvo, D. W. McDonald, T. Toscos, M. Y. Chen, J. Froehlich, B. Harrison, P. Klasnja, A. LaMarca, L. LeGrand, R. Libby, I. Smith, and J. a. Landay, "Activity sensing in the wild: a field trial of UbiFit garden," in *CHI '08 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2008, pp. 1797–1806.

[2] D. Harrison, N. Berthouze, P. Marshall, and J. Bird, "Tracking physical activity: problems related to running longitudinal studies with commercial devices," in *UbiComp '14 Adjunct: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, 2014, pp. 699–702.

[3] E. Karapanos, R. Gouveia, M. Hassenzahl, and J. Forlizzi, "Wellbeing in the making: Peoples' experiences with wearable activity trackers," *Psychol. Well. Being.*, vol. 6, no. 4, 2016.

[4] T. Fritz, E. M. Huang, G. C. Murphy, and T. Zimmermann, "Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness," *CHI '14 Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 487–496, 2014.

[5] I. Li, A. Dey, and J. Forlizzi, "A stage-based model of personal informatics systems," in *CHI '10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2010, p. 557.

[6] M. Whooley, B. Ploderer, and K. Gray, "On the Integration of Self-tracking Data Amongst Quantified Self Members," in *Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014 - Sand, Sea and Sky - Holiday HCI*, 2014, pp. 151–160.

[7] D. A. Epstein, A. Ping, J. Fogarty, S. A. Munson, C. Science, and H. C. Design, "A Lived Informatics Model of Personal Informatics," *UbiComp '15 Proc. 2015 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, pp. 731–742, 2015.

[8] I. Li, A. K. Dey, and J. Forlizzi, "Understanding my data, myself," *UbiComp '11 Proc. 13th Int. Conf. Ubiquitous Comput.*, p. 405, 2011.

[9] Endeavour Partners LLC, "Inside Wearables - How the Science of Behavior Change Offers the Secret to Long-Term Engagement," 2014.

[10] A. Lazar, C. Koehler, J. Tanenbaum, and D. H. Nguyen, "Why we use and abandon smart devices," *UbiComp '15 Proc. 2015 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, pp. 635–646, Sep. 2015.

[11] D. A. Epstein, M. Caraway, C. Johnston, A. Ping, J. Fogarty, and S. A. Munson, "Beyond Abandonment to Next Steps: Understanding and Designing for Life after Personal Informatics Tool Use," *CHI '16 Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2016.

[12] J. Clawson, J. A. Pater, A. D. Miller, E. D. Mynatt, and L. Mamykina, "No LongerWearing: Investigating the Abandonment of Personal Health-Tracking Technologies on Craigslist," in *UbiComp '15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, pp. 647–658.

[13] J. Rooksby, M. Rost, A. Morrison, and M. C. Chalmers, "Personal tracking as lived informatics," in *CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 1163–1172.

[14] J. Meyer, J. Fortmann, M. Wasmann, and W. Heuten, "Making Lifelogging Usable: Design Guidelines for Activity Trackers," in *Multimedia Modeling 2015*, vol. 8936, no. Lecture Notes in Computer Science, Sydney, NSW, Autralia, 2015, pp. 323–334.

[15] F. Bentley, K. Tollmar, and P. Stephenson, "Health Mashups: Presenting Statistical Patterns between Wellbeing Data and Context in Natural Language to Promote Behavior Change," *Tochi*, vol. 20, no. 5, 2013.

[16] N. D. Lane, M. Lin, M. Mohammod, X. Yang, H. Lu, G. Cardone, S. Ali, A. Doryab, E. Berke, A. T. Campbell, and T. Choudhury, "BeWell: Sensing Sleep, Physical Activities and Social Interactions to Promote Wellbeing," *Mob. Networks Appl.*, vol. 19, no. 3, pp. 345–359, 2014.

[17] D. Estrin, "Small data, where n = me," *Commun. ACM*, vol. 57, no. 4, pp. 32–34, 2014.

[18] K. Karrer, C. Glaser, C. Clemens, and C. Bruder, "Technikaffinität erfassen – der Fragebogen TA-EG (Assessing affinity to technology - the TA-EG questionnaire)," in *Der Mensch im Mittelpunkt technischer Systeme*, 2009, vol. 8, pp. 196–201.

[19] "Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire (IPAQ) – Short and Long Forms," 2005.

[20] P. Mayring, "Qualitative Inhaltsanalyse," in *Handbuch Qualitative Forschung in der Psychologie*, G. Mey and K. Mruck, Eds. Wiesbaden: VS Verlag f{ü}r Sozialwissenschaften, 2010, pp. 601–613.

[21] S. Consolvo, K. Everitt, I. Smith, and J. a. Landay, "Design requirements for technologies that encourage physical activity," *CHI '06 Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 457–466, 2006.

[22] R. Gouveia, E. Karapanos, and M. Hassenzahl, "How do we engage with activity trackers?: a longitudinal study of Habito," in *UbiComp '15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, pp. 1305–1316.