Identification and Classification of Usage Patterns in Long-Term Activity Tracking

Jochen Meyer
Merlin Wasmann, Wilko Heuten
OFFIS Institute for Informatics
Oldenburg, Germany
{firstname.lastname}@offis.de

Abdallah El Ali
Susanne C.J. Boll
Carl von Ossietzky University
Oldenburg, Germany
{firstname.lastname@uol.de}

ABSTRACT
Activity trackers are frequently used in health and well-being, but their application in effective interventions is challenging. While research for reasons of use and non-use is ongoing, little is known about the way activity trackers are used in everyday life and over longer periods. We analyzed data of 104 individuals over 14,413 use days, and in total over 2.5 years. We describe general tracker use, periodic changes and overall changes over time, and identify characteristic patterns. While the use of trackers shows large individual heterogeneity, from our findings we could identify and classify general patterns for activity tracker use such as try-and-drop, slow-starter, experimenter, hop-on hop-off, intermittent and power user. Our findings contribute to the body of knowledge towards the successful design of effective health technologies, health interventions, and long-term health applications.

Author Keywords
Activity tracker; usage patterns; quantitative analysis; longitudinal use; activity monitoring

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

INTRODUCTION
Physical activity is undoubtedly one of the most important behaviors in the prevention [33] or treatment [28] of numerous diseases such as obesity, diabetes, or cardio-vascular diseases. At the same time wellbeing and fitness are deemed socially desirable and being fit and active is part of a modern lifestyle. Within this context, activity trackers such as those from Fitbit, Garmin, Nike, Jawbone and other companies have become extremely popular.

CHI 2017, May 06 - 11, 2017, Denver, CO, USA

ACM 978-1-4503-4655-9/17/05…$15.00
DOI: http://dx.doi.org/10.1145/3025453.3025690

HCI research has investigated activity tracking for many years. Early systems such as UbiFit [6] or Fish-N-Steps[21] still used mobile phones as a sensing platform but already showed that interventions can successfully induce a change. While medical research has initially criticized activity trackers for their low precision in comparison to high-end accelerometers [30], they are now more and more understood as a new class of tools with distinct characteristics and acceptable precision [13] that is useful in e.g. epidemiological studies [3]. Activity trackers are used in numerous contexts such as personal wellbeing and fitness, life and health logging, but also health behavior change interventions and research studies. Despite their popularity, however, the use of activity trackers in health interventions is still challenging, and their success is not undisputed. While existing research has studied the effectiveness of activity tracker–based intervention, the reasons why people use activity trackers, and when and why they stop, the researchers have primarily looked at these subjects qualitatively [9] and only for short time periods.

To fully understand how activity trackers are used over longer periods of time, we studied long-term usage patterns across four different user cohorts in different studies. In this paper we present how activity trackers are used in different settings, in everyday life, and over prolonged periods of time. We provide an in-depth analysis of data from more than 100 users giving insights into when and how trackers are used. We identify usage patterns for short-term and long-term activity tracker users. From our observations we discuss implications for the design and utilization of activity tracker–based health interventions.

STATE OF THE ART
Activity trackers are a billion dollar market, with 22.5 million devices shipped in Q2/2016 alone [17]. The devices are easily used in daily life [23], including by elderly persons [12] and, with proper preparation, even by very difficult target groups such as dementia patients [32]. Nevertheless, their use still imposes technical and practical problems [16] and it is not yet clear if they will be used and contribute to personal wellbeing over longer periods of time.

Effectiveness of Activity Tracker–Based Interventions

Studies of the effectiveness of activity tracker–based interventions show mixed results. A meta-study conducted by Bravata et al. [2] showed significant increase of physical activity for pedometer users. And Connelly et al. [5] found that of 15 studies aiming to increase physical activity in type 2 diabetes patients using technology all reported an increase.

On the other hand, an analysis by Bort-Roig et al. [1] found that of 17 studies evaluating a mobile phone–based intervention to increase physical activity only 4 reported an increase in physical activity over up to six months. And in a pedometer-based intervention for elderly women conducted by McMurdo et al. [22] there was an increase after three months, but no significant difference beyond six months. Etkin [11] even found negative effects of short-term interventions: although step counting increased the participants' physical activity, it also reduced the joy of walking.

Long-Term Use, Abandonment, and Compliance

While behavior change interventions are the most obvious use for activity trackers, the potential beyond behavior change arising from long-term use of activity trackers is now receiving more attention. Karapanos et al. [18] 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 social relationships; rather user experience is driven by physical thriving and social relatedness. Based on interviews with 30 users of activity trackers [14] Fritz et al. 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.

The long-term potential of activity trackers, however, is lost by early abandonment. A market survey from 2014 found that one-third of all activity trackers are abandoned within 6 months [8]. Various research studies report similar or worse experiences: In a study by Lazar et al. [20] 80% of devices were abandoned after 2 weeks. And in a study by Shih et al. [31] 75% of 26 undergraduate students with activity trackers stopped using the tracker within four weeks.

Initial work has recently been presented discussing different types of user compliance in using activity trackers: In a long-term study, Purta et al. [29] found that users are either very compliant or not at all. Epstein et al. [10] suggest three basic types of use patterns: short use, long and consistent use, and intermittent use. While a detailed classification of use patterns has been done, e.g. for smartphone app use [35], we are not aware of similar work for activity tracker use.

Reasons for Use and Non-Use

There is growing consent among researchers on reasons for abandonment. Shih et al. 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. [9] 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. [4] found various detailed reasons, including expectation mismatch, technical complexity, and goal met.

Shih also found reasons for ongoing use: perceived usefulness, novelty and curiosity, hope for potential future use, and developed routine of use. Finally Gouveia et al. [15] found that activity trackers are "'deficit technologies' that 'scaffold' behavior during particular problematic moments in time, and 'transformational' technologies that instill and routinize new practices to the point that the technology is no more necessary."

APPROACH

We contribute to the ongoing research about use and abandonment of activity trackers by analyzing actual use of activity trackers in different settings, in the long term, and in real life. Complementing the existing research on *why* people use or do not use trackers, and going beyond initial work on compliance [10,29] we investigate *how* people use the trackers. Rather than looking at outcomes such as changes, we strictly focus only on use. Our findings contribute to understanding the users' needs and thus facilitate the design of better activity tracker–based interventions. To the best of our knowledge our approach has not been pursued before. Our work is therefore also meant to explore the opportunities from these analyses and to open the door for follow-up activities to confirm and extend our findings.

General Setup

Our analyses are based on data from more than 100 users collected in two intervention studies, one observational study, and in-the-wild data. These sources enable us to observe differences of use in different settings. In all four cases, the users tracked their physical activity in their daily lives using an off-the-shelf activity tracker. One intervention study is a short-term study of 12 weeks; the other three studies are long term, ongoing since at least 9 months and not limiting the duration of use of activity trackers, but leaving it to the user when to stop. The two intervention studies have been approved by ethical review committees and registered at public trial registers; all studies and the respective procedures have been approved by data protection officers. After receiving the users' consent we collected the data by accessing the tracker's corresponding internet service using the API provided by the manufacturers.

Depending on the study, different complementary data were available. We took into account a minimal core set that we deemed particularly relevant for activity tracker use. Not all data are available for all sources or all users, and for some data, measures are different between different sources. An

	IPP 1YearRehab	PROMOTE 12WeekIntervention	Lotus 9MonthObservation	VitaDock LongTermInTheWild
Study overview				
setting	Longitudinal intervention study	Short-term intervention study	Longitudinal observation study	Longitunal in-the-field use
duration of study	>= 1 year	12 weeks	>=9 months	arbitrary
influence of study on tracker use	low to medium	high	medium	low; opt-in
# users w/ activity data	43	17	5	39
Tools used				
tracker	Medisana ViFit	Fitbit Zip	Fitbit Ultra, One, Flex	Medisana ViFit
affinity to technology range	TA-EG 0-76 (highest)	n/a	BMTC 0-48 (highest)	self-assessed 1-5 (highest)
fitness, health, phys. activity; range	n/a	self-assessed health status; 1-5 (best)	IPAQ lo - med - hi	self-assessed fitness 1-5 (best)
Demographics				
# users with demogr.	22	6	5	38
age	34-74 ($\bar{x}$ 55.3, σ 11.3)	67-70 ($\bar{x}$ 68.0, σ 1.1)	33-56 ($\bar{x}$ 47.5, σ 9.0)	17-80 ($\bar{x}$ 53.4, σ 13.9)
sex	22 m	3m, 3f	2m, 3f	26m, 10f, 2 n/a
affinity to technology	5-68 ($\bar{x}$ 45.0, σ 15.7)	n/a	30-40 ($\bar{x}$ 36.0, σ 4.2)	3-5 ($\bar{x}$ 4.4, σ 0.7)
fitness, phys. activity	n/a	1-3 ($\bar{x}$ 2.3, σ 0.8)	2 hi, 2 med, 1 low	1-5 ($\bar{x}$ 3.1, σ 1.0)

Table 1. Overview of studies and characteristics.
Measures used: IPAQ [7], TA-EG [19]. BMTC - Brief Measure for Technology Commitment [26]

overview over the studies and the available data is given in Table 1. After the initial presentation we use meaningful codenames for the studies, also given in the table.

Studies

1YearRehab - IPP Intervention Study: IPP[1] is a multi-centric, comparative study involving patients aged 18-75 who suffered from a myocardial infarction in the last 30 days [34]. The primary goal of IPP is to improve rehabilitation in the 12 months after the infarct using an intensive prevention program (IPP) consisting of, among others, regular coaching, education, and clinical assessments. Patients in the intervention group were offered a Medisana[2] ViFit activity tracker and access to an online portal for regular observation of their compliance to heart friendly living, replacing a subjective self-assessment of physical activity with the objective monitoring by the ViFit tracker. 43 patients in the intervention group accepted and used a tracker; their data are used in our study. Of these, 22 filled out the demographic questionnaires at the beginning of the study. The patients were invited but not obligated to use the tracker. Moreover the tracker was not the focus but merely a side-measure of the intervention, so we assume that the study's influence on tracker usage is low to medium.

12WeekIntervention - PROMOTE Intervention Study: The PROMOTE study[3] investigates the effects of tailored physical activity interventions for elderly persons aged 65-75 [25]. The study is conducted successively in multiple local communities. Within two intervention groups and one control group, the participants' fitness is assessed pre- and post-study. The intervention groups conduct a 12-week training program tailored to their needs and abilities according to the initial assessment. Both intervention groups have access to a web portal for an online diary of their activities and as an exchange and information platform. One intervention group self-assesses their physical activity using the diary; the other intervention group additionally uses a Fitbit Zip for objective monitoring of physical activity. This group's data from two local communities are used in our study. For this group the use of the tracker was mandatory; we therefore assume that the study's influence on tracker usage is high.

9MonthObservation - Lotus Observation Study : The aim of the Lotus study is to observe the long-term use of

[1] Registered at ClinicalTrials.gov; study id NCT01896765

[2] A German company manufacturing consumer health devices and operating the "Vitadock" connected health service. see www.medisana.com and https://cloud.vitadock.com/?lang=en

[3] Registered at German Clinical Trials Register. https://drks-neu.uniklinik-freiburg.de/drks_web/; study id DRKS00010052.

networked health devices by average users under real-life circumstances. Seven persons are equipped with a comprehensive set of consumer devices for tracking physical activity, sleep, and weight for at least 9 months. The participants are requested to use the devices according to their own discretion and only as they see fit. Nevertheless we found that participants felt some urge to use the tracker "to deliver proper data". We therefore assume that the study's influence on tracker usage is medium. Due to technical access problems, two persons' data had to be excluded from analysis.

LongTermInTheWild - VitaDock Data Analysis Study: In a fourth study we acquired data from activity tracker users in the wild. In a newsletter sent to the customers of Medisana's connected health service "VitaDock" and in the VitaDock Facebook page we invited the recipients to provide us with their activity data and pointed them to an online system that we built for this purpose. After they granted our system access to their VitaDock account, their activity data were downloaded to our database. We then presented a short questionnaire with some demographic questions and self assessments which the users could, but were not required to, fill in. As this data is collected in the wild and ex-post there is no maximum duration in this setting, and there is no influence of the study on tracker use. However the opt-in procedure for providing the data may have induced a bias towards more engaged users.

Compared to general tracker users our population is slightly older and more male: A random sample of 1,463 ViFit users (independent of our population) showed a median of 47.8 years and 53.6% male; and US data from 2015 [27] showed 36% in age group 35-54 and 25% in age group 55+, and 46% male.

DATA PROCESSING

The activity data come as time series denoting for each user and each day the steps performed per minute (Fitbit) or per 15 minutes (Medisana ViFit). In rare cases we found some obviously wrong entries: 29 days (0.5% of the data) dated to the year 2000. And on 50 days we found 15-min intervals with more than 3,700 steps (long distance runners make less than 3,000 steps per 15 minutes). We ignored the respective days.

Measure	Description
use day	a day on which a user has recorded at least 500 steps
first / last day of use	a user's temporal first/last use day
duration of use	total time from first to last use day
density of period	relation of number of use days in a period to length of period, range [0..1]
total density	density for duration of use
use days per week	Alternative measure for density per week with range [0..7]
streak	uninterrupted series of use days
break	uninterrupted series of non-use days
phase	a series of streaks interrupted by short breaks, ending with a long break

Table 2. Terms and measures.

Measuring activity tracker usage is challenged by the fact that both non-use of the device and wearing the device while being inactive result in 0 steps for that time interval and cannot easily be distinguished. To the best of our knowledge this is true for all activity trackers today, and it is certainly true for the trackers used in the four studies. However, also in sufficiently long periods of inactivity some minimum number of steps is made, e.g. by moving around in the house or going to the bathroom. On the other hand we want to distinguish real use for monitoring physical activity (even if it relates only to short periods of a day) from unintended use where, e.g., the tracker is just taken from the bedside table and put into a drawer. After discussion with medical experts and after reviewing the data, we decided to use a simple threshold approach: Days with fewer than 500 steps are deemed non-use and are excluded from our analyses (8.3%, 1,320 of 15,812 days). An in-depth discussion of partial intra-day use of activity trackers with more insights is given below.

DATA ANALYSES

In the subsequent sections we first present general characteristics of activity tracker use such as duration and frequency of use, then describe longer-term characteristics such as weekly rhythms, and finally look in more detail into how users take breaks in the long term. We use a number of – mostly canonical – terms and measures shown in Table 2 to characterize the data. Where we use thresholds to categorize the data we define them quantitatively based on the observed data; we had also experimented with other e.g. heuristically defined thresholds and found that they don't considerably change the overall picture.

We analyzed the data from 104 users. The users collected data on 14,413 use days in the time from January 16, 2014, to August 1, 2016, and recorded 113,449,034 steps.

General Quantitative Observations

The key figures of our findings are summarized in Table 3.

Active users vs past users

Users frequently pause using the tracker and restart later. Therefore we cannot know for sure whether users are still using the tracker but are in a break, or whether they ultimately abandoned use. From our analyses of streaks and breaks – see below – we assume that a person is likely to have abandoned tracker use if that person hasn't used the tracker since 20% longer than their longest break so far, but at least for 50 days. We call these persons "past users".

With this, 66 of 104 users are active users, 38 are past users. Within the 1YearRehab study, abandonment is frequent. In contrast in the LongTermInTheWild study, most users are still active; this may be due to the opt-in

	1YearRehab	12Week Intervention	9Month Observation	LongTerm InTheWild
Active and past users				
Still active users	10	17	5	34
Past users	33	0	0	5
Typical duration (in days) and density				
Duration of active users	102-311 (196)	defined by study settings		270-566 (463)
Duration of past users	46-172 (125)	n/a	n/a	16-285 (163)
Total density	48-68% (59%)	90-100% (100%)	55-96% (91%)	20-93% (67%)
Typical streak and break lengths (in days)				
Streak length	1-2 (1)	3-17 (11)	3-28 (12)	2-13 (6)
Avg streak length per user	1.5-4 (2.2)	11-42 (15)	11.3-26.4 (18.8)	4.6-35.2 (12)
Max streak length	4-9 (7)	14-51 (15)	34-57 (38)	12-119 (36)
Break length	1-1 (1)	1-3 (2)	1-2 (1)	1-4 (1)
Average break length per user	1.2-3 (1.5)	1.3-3.9 (2.8)	1.4-9.3 (2)	3.4-35.4 (8.2)
Maximum break length	3-14 (6)	2-5 (4)	3-51 (4)	13-157 (53)
Use during day				
Average 3-a-day-ratio	73.6%	92.8%	88.2%	78.8%

Table 3. Quantitative observations. Ranges indicate Inter-Quartile Range, i.e. quartiles 2 and 3, figures in brackets are median

procedure for data acquisition that may primarily have attracted engaged users. With 12WeekIntervention being a short-term study with limited duration, past users are not relevant. Finally the 9MonthObservation study is still ongoing without dropouts.

Use Duration by Active Users

In the 1YearRehab study, active users are using the tracker for 3 to 10 months already, some up to 1½ years. In the LongTermInTheWild population, use durations are considerably longer up to 2.5 years. For 12WeekIntervention and 9MonthObservation, the maximum use duration is determined by the study settings.

Abandonment by Past Users

Abandonment in 1YearRehab happens often in the first 6 months. Of all 1YearRehab users, 11 (25.6%) definitely use or used the tracker for more than 6 months, 27 (62.8%) definitely abandoned within the first 6 months, and 5 (11.6%) are active users with a use duration of so far less than 6 months. The rate of abandonment is thus lower than found in the aforementioned studies [20,31], but higher than in the market survey [8].

In LongTermInTheWild, abandonment happened after two weeks to 13 months; again this may be influenced by the opt-in procedure for data acquisition and cannot be considered representative for the general population.

Density

The density, i.e. the percentage of use days in a period of time, is very different in the four settings. In the 12WeekIntervention study, it is up to 100%; this might be influenced by the high pressure induced by the study to use the tracker. In 9MonthObservation it is between 50% and close to 100%; due to the small population size this is strongly influenced by individual users' behaviors. 1YearRehab shows a fairly narrow density with typically 4.1 use days per week. In LongTermInTheWild densities are in general higher than in 1YearRehab but they are much more spread out. Some users have extremely low densities of less than 20%. We provide an explanation for the latter below.

Streaks and Breaks

Streaks, i.e. uninterrupted use of the tracker over successive days, in general last a few days only, with some considerable differences in the different populations and with outliers up to one year. In 1YearRehab most streaks are very short, with no more than 2 days, whereas in the more controlled 12WeekIntervention and 9MonthObservation settings streaks are much longer, often up to 3-4 weeks. The LongTermInTheWild setting is somewhere in between.

In all populations, breaks are on average shorter than streaks, and 94.8% of all breaks are one week or shorter. Outliers are more extreme, with maximum almost 2 years of break before resuming use. These long breaks are one explanation for the very low densities we observed particularly in the LongTermInTheWild population.

In each population a considerable number of users have a quite high maximum streak length. This is most visible in the LongTermInTheWild population, but can also be observed in other populations. In 1YearRehab this is least distinct, but there are outliers with longer streaks. Therefore many but not all users use the tracker at least once uninterruptedly for a longer period.

The longest break before using the tracker again is 3.5 to 53 days (median per study). With this we believe that the minimum 50-day period of non-use that we apply to estimate past users is a fair choice.

Usage During the Day

Investigating tracker use throughout the day is challenging, as both non-use of the tracker and inactivity result in 0 steps for the given interval. We therefore examine on which days the tracker has been used at least once in the morning from 3 a.m. to 11 a.m, once around noon from 11 a.m. to 3 p.m., and once in the afternoon after 3 p.m. This heuristics, subsequently called “3-a-day”, implies a continuity of use during the whole day. Applied to a long-term Fitbit user with known very high consistency, this measure shows 97.6%, reinforcing that the measure is a fair indicator for whole-day use: the higher the ratio of days fulfilling the 3-a-day measure in relation to all use days, the more likely it is that the user has used the tracker consistently the whole day.

With this measure we see that in all populations the tracker has been used in the morning, noon, and afternoon between 74% and 93% of all days. The 12WeekIntervention population, which in our previous measures, such as density, has already shown high engagement in wearing the tracker, scores best here, with 9MonthObservation performing only slightly worse. And the 1YearRehab population performs worst, with the LongTermInTheWild population being in between.

Use Rhythms per Weekday, Month, Over Time

Analyzing recurring use rhythms requires sufficiently long use durations that we only find in the 1YearRehab and the LongTermInTheWild populations (N=82). We therefore exclude 12WeekIntervention and 9MonthObservation from the subsequent analyses. For specific analyses we also need to further exclude participants; we mention this in the respective sections.

Use per Weekday

We observe how trackers are used on the days of the week. To mitigate the influence of the novelty effect we here take into account only users with at least 60 use days (N=62).

The distribution gives an unclear picture. In 1YearRehab (N=26) we observe a below-average use on Tuesdays and Wednesdays, and an above-average use on Saturdays and Sundays (see Fig. 1). For LongTermInTheWild users (N=27) the deviations are smaller, but the effect is somewhat reverse – less use on Saturdays and Sundays, and more during the week.

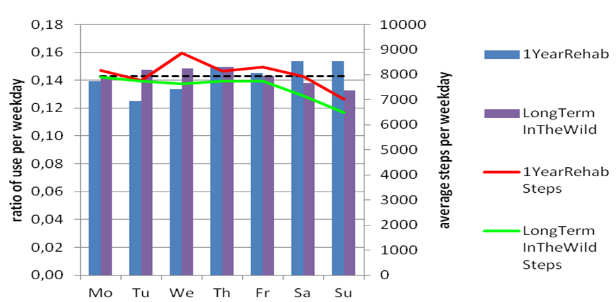

Fig. 1: Use and average steps per week day for 1YearRehab and LongTermInTheWild. Dashed line indicates equal distribution of tracker use

The relation between use of tracker and actual activity is subtle, as indicated by the average steps per weekday in Fig. 1. For LongTermInTheWild, use and average steps are about constant during the week, and lower on the weekend, confirmed by a strong positive correlation (Pearson’s r = 0.803, p = 0.0297) between use and average steps per day. Somewhat counter-intuitive, for 1YearRehab tracker use is higher during the weekend, but again average steps are lower; indeed there is no significant correlation between use and average steps (Pearson’s r = -0.399, p = 0.376). The users’ choice to wear the tracker is thus independent of the activity level.

Use per Calendar Month

To understand if there are periodical changes per year, we analyzed the density (measured in use days per week) per month of year, which only makes sense for users with at least one year of activity tracker use. And to avoid a bias by single users we only observe a time span with at least 10 users. This leaves us with Jan’15 to Jun’16 of the LongTermInTheWild population (N=24).

The analysis shows no obvious change of density throughout the year and no seasonal effects, e.g. during summer time or over Christmas. Wearing the tracker seems not to be influenced by, e.g., bad weather conditions or seasonally reduced activity.

Use over Time

To understand whether users change their use behavior over time, we analyze the density per week after first day of use. To avoid a bias by individual users, we only observe time spans with at least 10 users. We end up with 79 weeks with LongTermInTheWild data and 32 weeks with 1YearRehab data (N=81).

We observe (see Fig. 2) that aside from some minor change in the first weeks and some fluctuation, the use over time is on average practically constant. Users use the tracker consistently until the end, and then they stop. This is in contrast to Karapanos et al. [18] who found that intensity of use decreases over time; we will discuss this finding in the

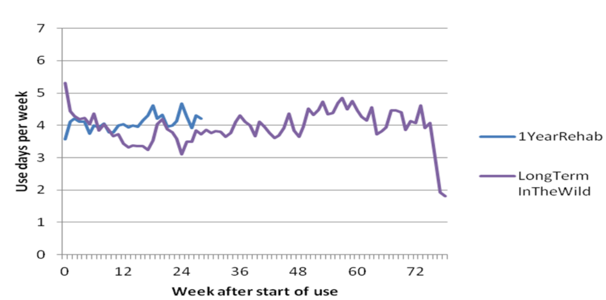

Fig. 2: Use after start of use

next section. The dip towards the end of LongTermInTheWild curve is an artifact caused by 5 long-term low-density users; we will explain this in the next section.

Taking Breaks from Use

Definition of Use Phases

To understand how users take breaks from using the tracker, we observe *phases*, i.e. series of streaks separated by short breaks and ending with a long break, as a more relaxed measure than the streak, allowing for a couple of days off. Our definition of a *long break* separating two phases is a break of more than one week: A week is a strong cultural rhythm that is perceived and lived as a basic time structure for differentiating "recent" from "past" also in the context of reflecting health behaviors [24]. Also we saw above that 94.8% of all breaks are one week and shorter, so anything longer is indeed an exception. Finally we also experimented with other durations up to 28 days and found that the overall picture drawn below remains consistent.

Number of Phases per User

Our data contain 218 phases in total, of which 177 (81.2%) are finalized, and 41 are still active. The number of phases per user is often low: 57 of the 104 users have exactly one finalized or active phase; and 94 users have finalized or are still active in four phases or less. Maximum number of phases is 9, achieved by two users. 41 users are currently active in a phase, 63 users have finalized their last phase, being either past users (N=38), or being in the break towards a possible next phase (N=25). Users in the LongTermInTheWild setting have more phases than users in the other three settings; this can partially be explained by the generally longer duration.

Since the characteristics of active phases may still change, we restrict our further analyses on finalized phases. To avoid a bias by individual users, we further restrict our per-study analyses to phases where at least 10 users finalized a phase. This again excludes 12WeekIntervention and 9MonthObservation and leaves the first two phases of 1YearRehab (N=41 / 12 in phase 1 / 2) and the first four phases of LongTermInTheWild (N=34 / 20 / 17 / 13).

Temporal Distribution of Phases

The length of the first phase varies widely from 1 to 529 days. Median in 1YearRehab is quite high, with 125 days. For LongTermInTheWild the first phase is much shorter, with median 29.5 days (see Fig. 3). The subsequent phases decrease considerably in length. In 1YearRehab, median of phase 2 drops drastically to 23 days with outliers from 1 to 156 days. For LongTermInTheWild, variations are smaller but still show a reduction to up to half the length of the first phase. Breaks between phases are usually shorter than the phase lengths. They have typical durations of 2-3 weeks up to 2-3 months, although they can also be longer than 1.5 years.

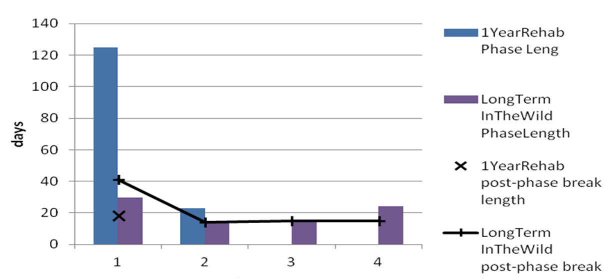

Fig. 3: Median length of phases (bars) and breaks after phases (marker, line)

Characteristics of Phases

We use four core measures to describe usage characteristics within the phases: density, average and maximum streak length, and ratio of 3-a-day use days to total use days (see Fig. 4). Median density in phase 1 is 65% to close to 100%. In 1YearRehab density is increasing from phase 1 to 2. In LongTermInTheWild we observe only a minor decrease in density from phase to phase; this is in line with our observations on use over time remaining mostly constant. We discussed above that in contrast to Karapanos et al. [18] we found on average no change of intensity over time; with this observation on use phases we now explain this contrast by the fact that per user in phases of use the density indeed stays mostly stable, but longer breaks and shorter use phases reduce the overall density. Average streak length is between 2 and 15 days in the first phase (with an impressive maximum of 179 days). In 1YearRehab there is an increase on a low level from phase 1 to 2, whereas in LongTermInTheWild there is a clear decrease from the first to the later phases. Maximum streak lengths per phase show little variation. 3-a-day ratio, i.e. intra-day use, decreases for 1YearRehab and fluctuates considerably for LongTermInTheWild.

Changes in Use between Phases

In general we see some tendency that the first phase has different characteristics from the subsequent ones. This might be explained in two ways: On the one hand, this might be inter-user effects caused by less than half the users entering into a second phase; this may result in a bias

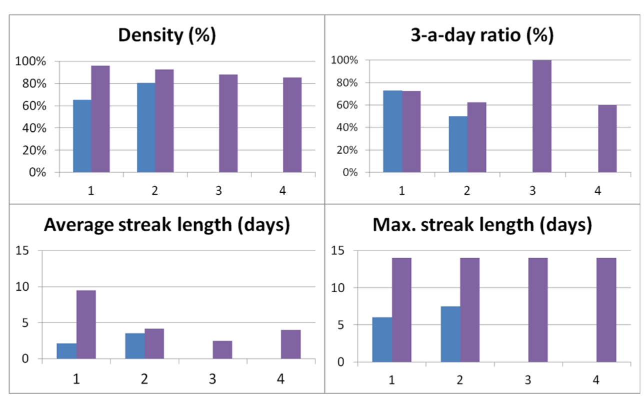

Fig. 4: Median of key characteristics of the first phases (horizontal) for 1YearRehab (left bar, blue) and LongTermInTheWild (right bar, purple).

towards more engaged users for later phases. On the other hand this may be intra-user changes, with the first phase possibly influenced by a novelty effect with more engagement and adherence, whereas a user's later phases may be more routine use. We discuss reasons for this in the next section.

PATTERNS OF USE

Approach

Based on the quantitative analysis of activity tracker use, we aim to identify patterns of use. We chose an expert driven, qualitative approach to explore our data and to identify potential patterns of use. Three HCI experts with experiences in activity tracker based health interventions, and one communication scientist with a focus on user experience and user analysis reviewed the data in a two-step process: In the first step they clustered phases based on common combinations of phases' attributes such as length and density. In the second step they identified the actual usage patterns by finding frequently occurring combinations of types of phases and other general use characteristics such as total use duration. For thresholds used to define phases and patterns, the experts discussed both, heuristic reasoning based on expert knowledge, and quantitative definitions based on observed data, and finally decided to go for quantitative definition which seemed less arbitrary and was still close to the experts' intuition. The experts individually reviewed the data and developed hypotheses for phases and patterns, taking into account their experiences from their previous work. The hypotheses were jointly discussed and the criteria applied to the data to check for occurrences of patterns, consistency with the quantitative observations such as the percentiles of the respective measures in the observed data, and ability to discriminate different usages. The process was repeated until a consensus was reached

Types of Phases

The experts identified the following, non-disjoint characteristic types of phases (in parentheses: the number of instances found):

- **minor use phase**: a short phase (in 1st quartile of phase length, 10 days) with either at the most 3 use days, or 3-a-day ratio below 50%. Here the user has tried but not really used the tracker. (N=45, 20.6%)
- **long streak phase**: a phase that contains a long streak (in 4th quartile maximum streak length per user, 41.25 days). Here the user has used the tracker at least once for a longer period of time without interruption. (N=45, 20.6%)
- **low density phase**: a phase of significant duration (>10 days) with density below median density (64%). Here the tracker is used regularly but sparsely. (N=37, 17.0%)
- **high intensity phase**: a phase of significant duration, density in 4th quartile (>92%), and 3-a-day ratio in 4th quartile (>89.6%). Here the tracker is used on most days and consistently. (N=24, 11.0%)
- **restarting phase**: a phase starting after a break with length in 4th quartile of maximum break length per user (32 days). Here the user is resuming use after having taken a long break (N=41, 18.8%)
- **very long phase:** a phase in 4th quartile of phase length (>103 days). Here the user has used the tracker over a very long time in a row (N=54, 24.8%)
- 76 phases (34.9%) have no specific characteristic.

Other types of phases were hypothesized, such as a low density phase that also includes a long streak, indicating a change of behavior within a phase. However no significant number of occurrences could be found for such phases in the data, hence they were rejected.

Use Patterns

The experts identified 12 non-disjoint use patterns. The number of occurrences of patterns, as well as the co-occurrences with other patterns, are reported in Table 4. Four of the patterns relate to the start of use (dotted frame top left in the table), five are relevant for longer-term use (dashed frame in the middle), two are additional attributes of other patterns, and one is without further specification.

Patterns for Start of Use

- **beginner**: *active user with short use duration (in 1st quartile; ≤ 62 days).* These users are still in their early days of use, so we cannot say much about them yet.
- **try-and-drop**: *past user with use duration ≤62 days and only minor use phases.* These users took the initial barrier of installing the tracker, but stopped using very fast before being able to assess the value of tracker use.
- **short-term** *past user with use duration ≤62 days*. These users tried the tracker for a short time. At some quite early point they must have made a – conscious or unconscious – evaluation of the benefits and found that they do not outweigh the perceived costs and disadvantages, in line with the reasons for abandonment already discussed [9,20,31].
- **slow-starter:** *first phase minor use, all other phases of significant duration (>10 days).* These users needed some time after setting up the tracker before going into actual use. This pattern is relatively rare; only 2 of the longer-term users are slow-starters, and 3 of the slow-starters are try-and-drop users with just the one minor use phase. So most users go into regular use right from the beginning,

Patterns for Long-Term Use

- **experimenter**: *≥ 2 minor use phases, and <90% of use days in non-minor use phases.* These users try out the tracker often for short periods of time. Experimenters can be found in both long-term settings, 1YearRehab and LongTermInTheWild. Some have very long use durations of over two years, and all have a low number of use days. These users seem to use the tracker more playfully, using it from time to time for a few days, possibly answering some very concrete question, and then stopping again.

- **hop-on hop-off**: *more than 3 phases, at least half of phases (excluding phase 1) and at least 2 phases are restarting phases*. These users often take long breaks from tracker use, but regularly resume use. We observe hop-on hop-off use in LongTermInTheWild only, but not in study and intervention-oriented settings. These users' first phase in general has a normal length (average 32 days), whereas in most cases (with two exceptions) the next phases are much shorter (average usually around 11 days). Users with this pattern have some reason to restart tracker use after some time. We hypothesize that the earlier phases help build a basic understanding, whereas later phases have a more concrete use such as re-evaluating a behavior.
- **intermittent user**: *>62 days of use and with only low density phases.* These users use the tracker consistently but sparsely. This pattern is particularly interesting. All intermittent users are in the 1YearRehab population. 3-a-day ratio is low with on average 74%, so intra-day use seems to be selective. On the other hand, use duration is relatively high; average of the intermittent users is 197 days. All but one users have only one phase; this makes most of them long-phase users. The engagement of intermittent users with trackers is therefore, although low, very consistent over a long time. It seems that intermittent users are not primarily interested in a complete logging and understanding of their activity, but have other reasons for use. This pattern seems to be related to external motivation induced by a study setting.
- **power user**: *>62 days of use and only high intensity phases.* These users use the tracker consistently and intensely. Power users appear in all but the 1YearRehab settings. All are still active; use duration is high in the open-ended LongTermInTheWild setting (average 370 days). 2 have two phases; the others have one phase only. 3 of those who are not long-phasers are in the 12WeekIntervention population, where the study is shorter than our definition of a very long phase. All have very long maximum streaks between 57 and 277 days, 3-a-day ratio is average 97%, so also intra-day use seems to be high. These users are compliant in using the tracker the whole day for a very long period of time, taking only exceptional breaks. They are found in study settings as well as in the wild, but their occurrence is relatively low.

	beginner	try-and-drop	short-term	slow-start	experimenter	hop-on hop-off	intermittent	power use	gen con	long-phaser	health logger	other
beginner	14			1							2	
try-and-drop		4	4	3	1							
short-term			13	4	1							
slow-start				7						1	1	
experimenter					8	1						
hop-on hop-off						7						
intermittent							14		14	14		
power use								9	9	5	9	
gen con									50	38	19	
long-phaser										43	17	
health logger											26	
other												8

Table 4. Occurrences and co-occurrences of patterns. Boxes mark patterns for start of use (left) and long-term use (right).

- **generally consistent user**: *>62 days of use, has only phases of significant duration.* This is a "Jane and John Doe" pattern, where the tracker is used for some time with moderate requirements to consistency of use. By definition all intermittent and all power users fall in this pattern. Leaving these out, 26 generally consistent users (25%) remain. This is the largest group that we observe. These users appear in all but the 12WeekIntervention settings; 15 are still active; average use duration is 11 months, density 73%, 3-a-day ratio 80%. This pattern is pursued by many users and shows a moderate but quite long-term engagement.

Attributive Patterns

The subsequent two patterns usually co-occur with other patterns. We therefore consider these to be primarily attributes of existing patterns rather than independent patterns themselves.

- **long-phase user:** *at least 75% of all use days in very long phases.* These users usually use the tracker without major interruptions for a longer period of time. This pattern overlaps particularly with generally consistent, intermittent and power users. The relatively low overlap with the latter is due to 12WeekIntervention users who cannot fulfill the long phases due to the study setting. There is no overlap with hop-on hop-off users or experimenters. Having a long phase thus seems to be a frequent behavior for users with a generally regular use.
- **health logger**: *at least one phase is a long-streak phase.* These users log complete activity data for a longer period of time. Health logging can be observed for all power users, for many generally consistent users, but only for one intermittent user and occasionally in some other patterns. Logging complete activity seems therefore to be related to generally higher engagement.

Miscellaneous

- **Other**: *a user fulfilling none of the patterns above.* 8 users fall into this pattern; 4 more fulfill only the long-phase and health logging attributes. The users in this pattern are diverse. They appear in all but the 12WeekIntervention settings. Some are long-term users with high consistency that just missed one of our strict criteria for another pattern, whereas others have unique combinations of phases such as mixtures of very long and short phases, or phases without specific characteristics. With 11.6% of such non-classified users our patterns have a good coverage.

DISCUSSION

At first glance our results may not look particularly surprising and merely give evidence to the seemingly obvious. However, looking closer our findings reveal important insights:

Most users fall in only one or two of the use patterns, and they don't change this pattern over long periods of time. Although these patterns are likely influenced by the study settings as indicated e.g. by the high density in 12WeekIntervention, they also appear across very different studies and thus also seem to be inherent to a user. We therefore conclude that a user has an individual and "natural" way of using an activity tracker.

While forcing users into other patterns may be possible, long-term interventions and studies using activity trackers are probably well advised to respect the user's natural patterns to increase adherence and reduce abandonment. Systems might even use these patterns to their advantage: If after a few weeks of use the user's pattern can be identified, interventions may adapt accordingly to the user's preferred way of tracker use. And deviations from a use pattern might indicate a change in use that makes changes in the intervention necessary.

Furthermore for the majority of users their natural way of tracker use involves regularly taking breaks. Temporary non-use is therefore not the exception but the rule and must be considered appropriately in interventions and studies.

Breaks can be as short as a couple of days or even just hours. Such breaks result in gaps in the data where the activity reported is less than the activity actually performed. This must be taken care of when analyzing the data. And in such short breaks users may be open to reminders and encouragement to use the tracker again soon.

Breaks may also be longer, covering many weeks or even months. This may be considered temporary abandonment and users may or may not resume use. Long-term interventions therefore need a great staying power when such breaks occur. Strategies to persuade the user to resume are likely different from those for short breaks. Accepting the break and waiting for the user to return on their own or using small nudges at the most may be better suited than continuously sending obtrusive reminders.

Nevertheless abandonment happens. While indeed many users stop after 3-6 months, there are individual differences ranging from a couple of days to more than a year. Reducing abandonment in early stages is particularly relevant for many interventions. Here two different challenges must be addressed: the one is the try-and-drop user; she or he must be convinced to give the tracker a fair chance. The other is the general short-term user who has tried the tracker but found no value in using it.

LIMITATIONS AND NEXT STEPS

Our work is to the best of our knowledge the first to use quantitative analyses of tracker use to understand user interaction. It is meant to provide initial reliable insights and open the door for follow-up activities which may confirm and broaden our results and can help to overcome the limitations of our work:

The number of users per setting is small to moderate, with 5 to 43 persons. The settings and the demographic data were heterogeneous, and the influence of the study settings on tracker use is not fully understood. Therefore some findings per setting and overall statistical relevance are limited. We based our identification of patterns on qualitative analyses. Future work could be based on quantitative analyses of larger populations, also including a greater number of different settings. We investigate observed use only. Follow-up work could also use a mixed-methods approach, identifying users' motivations and relating these to observed use patterns.

Our work is descriptive. Larger data might also allow identifying predictors for patterns, e.g. based on a user's demographic data and the study setting. This would allow tailoring an intervention to the individual user's needs from the very beginning, thus better matching the intervention to the user's actual behavior, increasing compliance and hopefully intervention success.

We discuss when users abandon use, but we don't examine the non-starting users who had been invited to tracker use but chose not to do so. The try-and-drop users are one step further than the non-starters but may have similar reasons for abandonment. Understanding their reasons might also help in building interventions that are not abandoned after a few weeks.

CONCLUSION

Our analyses provide detailed insights into activity tracker use in long-term and short-term studies. They add a quantitative perspective to the qualitative analyses of reasons for use and abandonment. The insights facilitate the design of activity tracking–based health technologies, interventions, and long-term applications better matching the users' preferences and needs.

ACKNOWLEDGMENTS

This work would not have been possible without the work of the teams from the PROMOTE and the IPP studies. Thanks also to Medisana for the support of the data acquisition.

This work has partially been funded by the German Federal Ministry of Education and Research (Project No. 01EL1422C), by the Hewlett Packard Innovation Research Programme, and by the Stiftung Bremer Wertpapier-Börse, (Bremen Stock Exchange Foundation).

REFERENCES

1. Judit Bort-Roig, Nicholas D Gilson, Anna Puig-Ribera, Ruth S Contreras, and Stewart G Trost. 2014. Measuring and influencing physical activity with smartphone technology: a systematic review. *Sports medicine (Auckland, N.Z.)* 44, 5: 671–86. http://doi.org/10.1007/s40279-014-0142-5

2. Dena M Bravata, Crystal Smith-Spangler, Vandana Sundaram, et al. 2007. Using pedometers to increase physical activity and improve health: a systematic review. *JAMA Journal of the American Medical Association* 298, 19: 2296–304. http://doi.org/10.1001/jama.298.19.2296

3. Steindorf K. Bruhmann B A, Schmidt M E. 2014. Assessment of physical activity in epidemiological studies : Are questionnaires obsolete in the era of accelerometry ? *GMS Med Inform Biom Epidemiol* 10, 1: 1–12. http://doi.org/10.3205/mibe000155

4. James Clawson, Jessica A. Pater, Andrew D. Miller, Elizabeth D. Mynatt, and Lena Mamykina. 2015. No longer wearing. *UbiComp '15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*: 647–658. http://doi.org/10.1145/2750858.2807554

5. J Connelly, A Kirk, J Masthoff, and S MacRury. 2013. The use of technology to promote physical activity in Type 2 diabetes management: a systematic review. *Diabetic medicine : a journal of the British Diabetic Association* 30, 12: 1420–32. http://doi.org/10.1111/dme.12289

6. Sunny Consolvo, David W. McDonald, Tammy Toscos, et al. 2008. 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*, 1797–1806. http://doi.org/10.1145/1357054.1357335

7. Cora L Craig, Alison L Marshall, Michael Sjöström, et al. 2003. International physical activity questionnaire: 12-country reliability and validity. *Medicine and science in sports and exercise* 35, 8: 1381–95. http://doi.org/10.1249/01.MSS.0000078924.61453.FB

8. Endeavour Partners LLC. 2014. *Inside Wearables - How the Science of Behavior Change Offers the Secret to Long-Term Engagement*.

9. Daniel A. Epstein, Monica Caraway, Chuck Johnston, An Ping, James Fogarty, and Sean A. Munson. 2016. Beyond Abandonment to Next Steps: Understanding and Designing for Life after Personal Informatics Tool Use. *CHI '16 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. http://doi.org/10.1145/2858036.2858045

10. Daniel A. Epstein, Jennifer Kang, Laura R. Pina, James Fogarty, and Sean A. Munson. 2016. Reconsidering the Device in the Drawer: Lapses as a Design Opportunity in Personal Informatics. In *UbiComp '16: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 829–840. http://doi.org/10.1145/2971648.2971656

11. Jordan Etkin. 2016. The Hidden Cost of Personal Quantification. *Journal of Consumer Research* 42, 6: 967–984.

12. Cara Bailey Fausset, Tracy L Mitzner, Chandler E Price, Brian D Jones, Brad W Fain, and Wendy A Rogers. 2013. Older Adults' Use of and Attitudes toward Activity Monitoring Technologies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 57, 1: 1683–1687. http://doi.org/10.1177/1541931213571374

13. Ty Ferguson, Alex V Rowlands, Tim Olds, and Carol Maher. 2015. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. *The International Journal of Behavioral Nutrition and Physical Activity* 12: 42.

14. Thomas Fritz, Elaine M Huang, Gail C Murphy, and Thomas Zimmermann. 2014. Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. *CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: 487–496. http://doi.org/10.1145/2556288.2557383

15. Rúben Gouveia, Evangelos Karapanos, and Marc Hassenzahl. 2015. 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*, 1305–1316. http://doi.org/10.1145/2750858.2804290

16. Daniel Harrison, Nadia Berthouze, Paul Marshall, and Jon Bird. 2014. 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*, 699–702. http://doi.org/10.1145/2638728.2641320

17. IDC. 2016. *Worldwide Quarterly Wearable Device Tracker*.

18. Evangelos Karapanos, Rúben Gouveia, Marc Hassenzahl, and Jodi Forlizzi. 2016. Wellbeing in the making: Peoples' experiences with wearable activity trackers. *Psychology of Well-Being* 2013. http://doi.org/10.1186/s13612-016-0042-6

19. Katja Karrer, Charlotte Glaser, Caroline Clemens, and Carmen Bruder. 2009. Technikaffinität erfassen – der Fragebogen TA-EG (Assessing affinity to technology - the TA-EG questionnaire). In *Der Mensch im Mittelpunkt technischer Systeme*, 196–201.

20. Amanda Lazar, Christian Koehler, Joshua Tanenbaum, and David H. Nguyen. 2015. Why we use and abandon smart devices. *UbiComp '15: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*: 635–646. http://doi.org/10.1145/2750858.2804288

21. James J Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B Strub. 2006. Fish ' n ' Steps : Encouraging Physical Activity with an Interactive Computer Game. In *UbiComp '06 Proceedings of the 8th international conference on Ubiquitous computing*, 261–278. http://doi.org/10.1007/11853565_16

22. Marion E T McMurdo, Jacqui Sugden, Ishbel Argo, et al. 2010. Do pedometers increase physical activity in sedentary older women? A randomized controlled trial. *Journal of the American Geriatrics Society* 58, 11: 2099–106. http://doi.org/10.1111/j.1532-5415.2010.03127.x

23. Jochen Meyer, Jutta Fortmann, Merlin Wasmann, and Wilko Heuten. 2015. Making Lifelogging Usable: Design Guidelines for Activity Trackers. In *Multimedia Modeling 2015*. Sydney, NSW, Autralia, 323–334. http://doi.org/10.1007/978-3-319-14442-9_39

24. Jochen Meyer, Anastasia Kazakova, Merlin Büsing, and Susanne Boll. 2016. Visualization of Complex Health Data on Mobile Devices. In *MMHealth'16: Multimedia for personal health and health care Proceedings*. http://doi.org/10.1145/2985766.2985774

25. Saskia Muellmann, Sarah Forberger, Tobias Möllers, et al. 2016. Effectiveness of eHealth interventions for the promotion of physical activity in older adults: a systematic review protocol. *Systematic Reviews* 5, 1: 47. http://doi.org/10.1186/s13643-016-0223-7

26. Franz J. Neyer, Juliane Felber, and Claudia Gebhardt. 2012. Entwicklung und Validierung einer Kurzskala zur Erfassung von Technikbereitschaft (Development and validation of a brief measure of technology commitment). *Diagnostica* 58, 2: 87–99. http://doi.org/10.1026/0012-1924/a000067

27. NPD. 2015. *Consumers and Wearables Report*. Las Vegas.

28. Massimo F Piepoli, Ugo Corrà, François Carré, et al. 2010. Secondary prevention through cardiac rehabilitation: physical activity counselling and exercise training. *European Heart Journal* 31, 16: 1967–76. http://doi.org/10.1093/eurheartj/ehq236

29. Rachael Purta, Stephen Mattingly, Lixing Song, et al. 2016. Experiences measuring sleep and physical activity patterns across a large college cohort with fitbits. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers - ISWC '16*, 28–35. http://doi.org/10.1145/2971763.2971767

30. Patrick L Schneider, Scott E Crouter, Olivera Lukajic, and David R Bassett. 2003. Accuracy and reliability of 10 pedometers for measuring steps over a 400-m walk. *Medicine and Science in Sports and Exercise* 35, 10: 1779–1784. http://doi.org/10.1249/01.MSS.0000089342.96098.C4

31. Patrick C. Shih, Kyungsik Han, Erika Shehan Poole, Mary Beth Rosson, and John M. Carroll. 2015. Use and Adoption Challenges of Wearable Activity Trackers. In *iConference 2015 Proceedings*.

32. Torben Wallbaum, Melina Frenken, Jochen Meyer, and Andreas Hein. 2014. Acceptance of sensor networks by dementia patients. In *7. Deutscher AAL-Kongress*.

33. WHO. 2011. *Global status report on noncommunicable diseases 2010*.

34. H. Wienbergen, T. Backhaus, J. Stehmeier, et al. 2016. Assessment and monitoring of physical activity with pedometers and and online documentation in secondary prevention after myocardiac infarct (Erfassung und Kontrolle körperlicher Aktivität mit Schrittzählern und Online-Dokumentation in der Sekundärprävent. In *Tagung der Deutschen Gesellschaft für Kardiologie*.

35. Sha Zhao, Julian Ramos, Jianrong Tao, et al. 2016. Discovering different kinds of smartphone users through their application usage behaviors. In *UbiComp '16: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 498–509. http://doi.org/10.1145/2971648.2971696