Proceedings of the 53rd Hawaii International Conference on System Sciences | 2020

Fitness Tracking Technologies: Data Privacy Doesn't Matter? The (Un)Concerns of Users, Former Users, and Non-Users

Kaja J. Fietkiewicz
Heinrich Heine University Düsseldorf
Kaja.Fietkiewicz@hhu.de

Aylin Ilhan
Heinrich Heine University Düsseldorf
Aylin.Ilhan@hhu.de

Abstract

To be concerned about data privacy in the fitness tracking world is apparently not the question of age or fitness level. It also does not necessarily influence the actual use of fitness tracking technologies. In this empirical study, 590 participants from the EU and USA, who are current users, former users or non-users of fitness tracking applications, were surveyed (online) on their sensitivity perception of several data pieces collected with fitness trackers as well as their data privacy concerns. Furthermore, subgroups of different fitness tracking users were detected based on their different privacy unconcerns.

1. Introduction

Today, ubiquitous technologies spread rapidly in different spheres of our lives. Even though the use of these technologies is not forced on anybody, the shift towards increased application of digital goods and new trends appears omnipresent and somewhat inevitable. The adoption of these new trends can be based on genuine interest or gained benefits, but also on social pressure or the need to belong. Depending on many factors, the usage of these technologies might feel safe and solely beneficial or it can be accompanied by uneasy feeling, e.g., of being dependent, surveilled or, in general, uncertain of the security of personal data collected or generated with this technology. One good example are the users of fitness tracking and similar wearable technologies, who apply them while often having many concerns about privacy risks. Still, in order to profit from the (fitness and health) benefits, they need to accept the challenges and threats. Data privacy and security became one of the prominent concerns in this area, especially since wearable technology encourages collection, storage and sharing of health-related data, which might be perceived as more sensitive than the usual name-gender-age information, nowadays rather willingly shared on many social networks.

Even though the fitness tracking tools give (health and fitness-related) benefits to the consumers, they also pose new and partially unpredictable challenging threats to data privacy and security. These threats exist due to the possibility of ubiquitous collection of large amounts of data in real time and creation of detailed user behavior patterns, e.g., when people eat, sleep (and how good or bad), exercise or go home from work [24]. The new tracking devices and applications are collecting both, personal information as well as health data, and create "a quantified self for their users," which becomes especially risky when the companies (being in custody of users' data) might violate their privacy and misuse it [22, 28:13].

Activity tracking technologies are collecting different kinds of data (e.g., steps, heart rate, sleep stages, geolocation), which might be considered to have different degrees of sensitivity. This led Lidynia, Brauner, and Ziefle [18] to investigate the users' perceived sensitivity of different data types. They online-surveyed 82 participants from Germany, where 46 participants were characterized as non-users of wearables and 36 participants as wearable users. Their results show that data types such as GPS, sleep analysis, and weight are perceived as (rather) sensitive in comparison to, for example, step count, hours spent standing, and the number of climbed stairs. Lehto and Lehto [17] investigated the user perception of privacy and sensitivity of health information collected with wearable devices as well as their willingness to share such information with other parties. The participants of their qualitative study "described the information collected by their devices as not sensitive, not secret, not confidential, and quite general" [17:247]. Even though the collected information was not perceived as sensitive, some interviewees expressed concerns when the data should be connected with individual's name and address.

Previous studies showed that people are mainly concerned about the collection of GPS data [16, 22, 23] as well as data about their mood or stress level [23, 25] and the detailed health information [23]. This topic attracts attention also outside the scientific community. For example, last year, The Guardian reported about the case "Fitness tracking app Strava gives away location of

URI: https://hdl.handle.net/10125/64163
978-0-9981331-3-3

secret US army bases" [15]. Even though this breach was not concerning data collected by daily users or runners, it again showed the sensitivity of information pieces obtained through different fitness applications and which potential risks might be lurking [15]. Although people seem to agree on sensitivity of location or detailed health data, the users of fitness trackers do not express one specific privacy concern about data collection on their device, as it appears to change depending on various factors [11, 16, 22, 30]. Lower concerns or even unconcerns can be explained by the lacking awareness of how users' privacy can be compromised due to collection of "granular data about users over a long time" [30:230].

In order to identify how do former and current user as well as non-users of fitness tracking applications perceive sensitivity of several data types collected by this technology, we formulate the first research question (**RQ1**): What is the perceived sensitivity of different data types by current users, former users, and non-users of the fitness tracking applications?

Not without reason, many users of tracking apps have concerns about privacy protection [20], third party access to data [7], as well as access to personal information by apps [4]. Still, even when users understand and care about potential data privacy risks, "they feel that once information is shared, it is ultimately out of their control. They attribute this to the opaque practices of institutions, the technological affordances of social media, and the concept of networked privacy, which acknowledges that individuals exist in social contexts where others can and do violate their privacy" [13:3738].

Fitness tracking technologies are popular not only among the consumers, but also researchers on human-computer interaction and health informatics. The number of studies on activity tracking technologies increased over the last years [27]. Recently, it focuses more on the ubiquitous data collection and privacy [1, 5, 6, 24, 25]. Due to the "mobile and networked nature of fitness trackers […] they automatically and persistently collect data, which companies share with or sell to third parties" [30:230]. Although seemingly anonymous, the collected user data can be more easily re-identified due to the increasing uniqueness of the datasets [12, 24].

There is scientific interest in users' behaviors when sharing the so-called personal fitness information and the privacy concerns coming from the collection, aggregation, and sharing of these information pieces [30:229]. How sensitive do people perceive their fitness information to be? And what data privacy concerns do they have? These questions are increasingly discussed in context of the so-called privacy paradox [1, 3, 5], meaning that even though users express some privacy concerns, they "behave in ways that appear to undermine their privacy" [30:230].

Based on the disagreement regarding what privacy concerns about fitness tracking technology do users and non-users indeed have, we formulated the second research question (**RQ2):** What are the general privacy concerns about fitness tracking by current users, former users, and non-users of the fitness tracking applications?

Finally, previous research indicates that some users apply fitness tracking applications to the fullest extent even though they have data privacy-related concerns (so-called privacy paradox). Also, there are users who do not voice any specific concerns about the fitness tracking technologies. Hence, there appear to exist different groups of fitness tracking users when considering the state of (perceived) data sensitivity and security. This leads us to the final research question (**RQ3):** What types of fitness tracking applications' users can we distinguish based on their data privacy concerns?

2. Methods

In order to collect suitable data for this study, an online survey was conducted. This way it was possible to reach as many participants from the European Union and the USA as possible. The origin of fitness tracking users can impact their attitude towards data privacy [2, 3, 21] and should be considered as an influencing factor during the interpretation of the results; especially considering the fundamentally different data protection history and regulations in the USA and the EU [10]. The survey targeted not only current users of fitness tracking technologies, but also former users and non-users, who also might have data privacy concerns.

The online survey started with questions about the use of activity tracking applications and wearables, as well as their usage frequency and duration. Inquired was also the general opinion on (online) data privacy ("I am not concerned about security on the Internet, e.g. people finding out what websites I visit or getting to know my real identity," and "I do not care what companies whose services or applications I use do with my (personal) data"), which could be valued on a 5-point Likert scale. These questions provided data to include further factors possibly influencing privacy-related concerns about fitness tracking applications as well as the perceived sensitivity of data pieces collected by fitness tracking technology.

Seven items were formulated to inquire participants' data privacy-related concerns about fitness tracking applications (e.g., misuse of data by the company). The questions could also be answered on a 5-point Likert scale. Five of the seven items were adopted from Lidynia, Schomakers, and Ziefle [19], who among others investigated the data privacy concerns of fitness tracking users and non-users in Germany. The other two items were added based on the research about involvement of heath insurances and possible third parties inferences [14, 17, 26]. All three types of participants (users, former users, and non-users of activity tracking applications) had to answer those questions.

In order to measure the perceived sensitivity of different data types, the following data pieces were adopted from the work by Lidynia, Brauner, and Ziefle [18:45]: "Step count," "Pulse,"* "GPS,"* "Calories,"* "Blood pressure," "Stairs,"* "Standing hours," "Sleep analysis,"* "BMI,"* "Blood sugar," and "Weight." Data pieces marked with "*" were labeled differently than in research by Lidynia, Brauner, and Ziefle [18] in order to clarify the meaning of the data pieces to the survey participants. Considering the functionalities of activity tracking technologies, further data pieces were added: menstrual cycle, completed workouts, fitness level/experience points, trophies, badges, lost and won challenges, real name, gender, birthday, e-mail, contacts/friends, and joined groups. All in all, the survey included 23 data pieces, which had to be assessed by all three groups of participants. The data pieces were grouped into the categories: personal data, health-related data, activity data and progress data. The rating scale for each data piece ranged from "1—not sensitive at all. I would make it public" to "5—Very sensitive. I don't want anyone to know it." Here, also the answer possibility "I don't know what it is" (especially for non-users) or "Not applicable" (e.g., information piece being "menstrual cycle" had to be answered by male participants) were included.

The survey was pretested by six participants, two non-users and four current or former users of activity tracking technologies. Three pretesters were males and three were females. After the pretest was concluded, the survey was slightly modified in regard to language (e.g., statements formulated more objectively), clarification of any ambiguities, adding of open questions for further remarks, and making the survey more user-friendly by different positioning and segmentation of the questions.

The online survey was non-probabilistically distributed from February 26, 2019, until May 28, 2019. It was spread through different social media channels, both private profiles and social media groups (e.g., Facebook, Reddit, Twitter, or Xing), scientific communities (ASIS&T), or portals for survey sharing (SurveyCircle, SurveyTandem).

The collected data was cleaned—incomplete answers and answers provided by pretesters were excluded, and the data was recoded into numerical values with the Syntax of IBM SPSS Statistics 25. The data collected from answers marked on the Likert scale was handled as ordinal.

In order to answer the first two research questions, the Kruskal-Wallis H Test and a subsequent post-hoc test were conducted to investigate the differences in perceived data sensitivity and data privacy concerns between three groups of participants (users, non-users, and former users of fitness tracking applications). Kruskal-Wallis H Test is a rank-based nonparametric test used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable [29]. It is adequate for our approach and collected data since the dependent variables (perceived data sensitivity and data privacy concerns) are measured on an ordinal scale. In order to determine which group(s) exactly are different from which other group(s), a post-hoc test—all pairwise comparisons using Dunn's [8] procedure with a Bonferroni adjustment, was conducted [29].

In order to determine the characteristics of possible subgroups of fitness tracking applications' users based on their perceived data sensitivity as well as data privacy concerns, the *K*-means clustering procedure was conducted. The *K*-means clustering algorithm was run for a range of *K* values in order to determine the most suitable one. Since the scale of the included ordinal variables ranges only from 1 to 5, the most distinctive group differences were given for *K*=3.

3. Results

Out of 777 online survey participants, 649 completed the survey (83.53%). Only participants who stated to be from the USA or the EU (N=590) were included in further analysis. The origin of fitness tracking users was considered as possibly influencing factor during the interpretation of the results.

The descriptive information about the sample is listed in Table 1. The distribution by gender is almost balanced (with 56% female participants). The survey addressed not only users of fitness tracking applications (55.9%), but also former users (9%) and non-users (35%). The age distribution is satisfactory, since both, elderly (over 60) and young adults (up to 23 years old), are represented within the sample. The participants of the survey had to indicate their year of birth. For further analysis a categorization into four generations, based on research on inter-generational differences in digital media usage [9, 10], was conducted. The four generations include: Silver Surfers (born before 1959, hence at least 60 years old), Gen X (or Digital Immigrants, born between 1960 and 1979, hence 40-59 years old), Gen Y (also Digital Natives or Millennials, born between 1980 and 1995; between 24 and 39 years old), and finally, Gen Z (born after 1996, hence, up to 23 years old).

Table 1. Demographic information (N=590).

	Freq.	%
Origin		
EU	477	80.8%
USA	113	19.2%
Gender		
Female	331	56.1%
Male	253	42.9%
Other	6	1.0%
Fitness Tracking Application		
Current Users	330	55.9%
Non-Users	207	35.1%
Former Users	53	9.0%
Generation		
Silver Surfers	55	9.3%
Gen X	115	19.5%
Gen Y	327	55.4%
Gen Z	93	15.8%

The inclusion of non-users in the survey gives us a third perspective on the perceived data sensitivity and data privacy concerns with regard to fitness tracking. However, their answers can be influenced by further factors like inexperience with fitness tracking or disinterest in fitness activity in general. The possible distortion of the results by non-users' lacking knowledge about different data pieces etc. was minimized by inclusion of the answer possibility "I don't know".

In order to verify, if the participating non-users are at all physically active, which might have an influence on their attitude towards fitness tracking in general, the Kruskal-Wallis H Test was conducted to determine any significant differences between users, non-users, and former users regarding their "daily activity level" (from "predominantly not active, e.g., due to an office job," to "highly active") and their fitness or exercise intensity. As for the daily activity level (ranging from 1 to 5), the median equals 3 ("moderately active") for all three groups, there is, however, a significant difference in the distribution between current users (Mean Rank = 317.10) and non-users (Mean Rank = 263.31), H(2) = 14.058, p = .001. As for the question about how often do the participants exercise (frequency ranging from 1 to 8), the differences in medians are rather small. For current users the median equals 7 ("exercising 3 or more times per week"), whereas for former users and non-users the median equals 6 ("exercising 1-2 times per week"). There are, however, significant differences in the distributions, H(2) = 36.268, p < .001, between current users (Mean Rank = 327.70) and non-users (Mean Rank = 240.20) (p < .001) as well as between former users (Mean Rank = 310.98) and non-users (p = .016). Even though there are significant differences in fitness or exercise activity, on average the non-users of fitness tracking technologies are still quite active (exercising 1-2 times per week), which indicates a general interest in fitness (just not fitness tracking).

3.1. Perceived data sensitivity (RQ1)

The first research question concerns the differences in perceived data sensitivity by users, non-users, and former users of the fitness tracking applications. The results of the Kruskal-Wallis H Test for perceived sensitivity of "personal data" (Table 2) indicates significant differences in distributions for only two data pieces— "gender" and "interest groups." A post-hoc test revealed statistically significant differences between users and non-users in both cases. When looking at the mean perceived sensitivity values for all groups (medians), there are barely any differences, except for "gender." The users and former users perceive those pieces of information as rather not sensitive, whereas non-users perceive them as neutral.

There are more significant differences in the distribution of the perceived sensitivity of health-related data (continued in Table 2). Except for the information about "menstrual cycle" (sensitive for all three groups), the perceived sensitivity of all remaining information pieces is different between users and non-users. Moreover, there is a significant difference between former users and non-users for the data pieces "heart rate" and "sleeping times." When considering the mean perceived sensitivity, the non-users valued "calories intake" and "heart rate" higher than former and current users, who perceive them as neutral. Interestingly, current users and non-users perceive "blood pressure" and "sleeping times" as rather sensitive, whereas the former users have a neutral attitude towards them.

Regarding the activity and progress data, all three groups agree on high sensitivity of GPS data (median of 5 for all groups, no significant differences in distribution). For the remaining data pieces, there are significant differences between users and non-users, and additionally between former users and non-users for the information about "step count." Except for GPS data, users and former users perceive all remaining activity and progress data as neutral (median of 3). Non-users also perceive most of the data pieces as neutral, except for the "step count" (interestingly seen as very sensitive, median equals 5), "fitness level or experience points" as well as "lost challenges" (rather sensitive, median equals 4). Interestingly, even though non-users perceive the information about "lost challenges" as rather sensitive, their perception of the information about "won challenges" is neutral (median equals 3).

Table 2. Differences in perceived sensitivity of different data pieces by mean ranks (MR) and medians (Mdn).

			Users (Y)	Non-Users (N)	Former Users (F)	Kruskal-Wallis H Test	Post-Hoc Test
Personal data	Real name	*MR*	277.53	312.17	287.42	H(2) = 5.714	-
		Mdn	4 (n=325)	4 (n=202)	4 (n=53)	*p* = .057	
	Gender	*MR*	274.20	322.87	277.81	H(2) = 11.621	Y- N *p* = .002
		Mdn	2 (n=327)	3 (n=203)	2.5 (n=52)	*p* = .003	
	Birthday	*MR*	281.14	310.42	283.34	H(2) = 4.272	-
		Mdn	4 (n=328)	4 (n=202)	4 (n=52)	*p* = .118	
	E-Mail	*MR*	296.76	280.21	302.07	H(2) = 1.660	-
		Mdn	4 (n=327)	4 (n=202)	4 (n=53)	*p* = .436	
	Contacts/ friends	*MR*	280.29	306.20	304.64	H(2) = 4.264	-
		Mdn	5 (n=327)	5 (n=202)	5 (n=53)	*p* = .119	
	Interest groups	*MR*	274.66	320.44	296.75	H(2) = 10.042	Y - N *p* = .005
		Mdn	4 (n=329)	4 (n=202)	4 (n=53)	*p* = .007	
Health-related data	Calories intake	*MR*	270.88	324.52	278.09	H(2) = 13.835	Y - N *p* = .001
		Mdn	3 (n=322)	4 (n=205)	3 (n=53)	*p* = .001	
	Burned calories	*MR*	269.87	331.45	287.00	H(2) = 17.662	Y - N *p* < .001
		Mdn	3 (n=327)	3 (n=205)	3 (n=53)	*p* < .001	
	Heart rate	*MR*	278.19	322.67	259.78	H(2) = 11.496	Y - N *p* = .028
		Mdn	3 (n=327)	4 (n=203)	3 (n=53)	*p* = .003	F - N *p* = .038
	Blood pressure	*MR*	275.28	320.23	270.50	H(2) = 10.384	Y - N *p* = .006
		Mdn	4 (n=325)	4 (n=202)	3 (n=53)	*p* = .006	
	Sleeping times	*MR*	280.13	321.51	256.34	H(2) = 11.084	Y - N *p* = .013
		Mdn	4 (n=326)	4 (n=205)	3 (n=53)	*p* = .004	F - N *p* = .027
	BMI	*MR*	274.76	320.40	299.55	H(2) = 9.966	Y - N *p* = .005
		Mdn	4 (n=327)	4 (n=205)	4 (n=53)	*p* = .007	
	Weight	*MR*	274.02	319.67	312.81	H(2) = 10.686	Y - N *p* = .005
		Mdn	4 (n=328)	4 (n=205)	4 (n=53)	*p* = .005	
	Menstrual cycle	*MR*	216.72	241.01	208.37	H(2) = 5.113	-
		Mdn	5 (n=244)	5 (n=159)	4 (n=45)	*p* = .078	
Activity & progress data	Step count	*MR*	209.15	404.51	254.40	H(2) = 175.95	Y - N *p* < .001
		Mdn	3 (n=329)	5 (n=159)	3 (n=53)	*p* < .001	F - N *p* < .001
	GPS	*MR*	281.70	305.09	300.15	H(2) = 3.584	-
		Mdn	5 (n=327)	5 (n=202)	5 (n=53)	*p* = .167	
	Climbed stairs	*MR*	265.09	335.07	302.46	H(2) = 23.264	Y - N *p* < .001
		Mdn	3 (n=327)	3 (n=205)	3 (n=53)	*p* < .001	
	Standing hours	*MR*	268.89	325.82	298.01	H(2) = 15.446	Y - N *p* < .001
		Mdn	3 (n=325)	3 (n=204)	3 (n=53)	*p* < .001	
	Completed workouts	*MR*	262.61	337.80	295.08	H(2) = 26.553	Y - N *p* < .001
		Mdn	3 (n=325)	3 (n=205)	3 (n=53)	*p* < .001	
	Fitness level, XPs	*MR*	260.09	325.45	285.78	H(2) = 20.278	Y - N *p* < .001
		Mdn	3 (n=320)	4 (n=196)	3 (n=53)	*p* < .001	
	Trophies, badges	*MR*	261.86	326.45	285.54	H(2) = 19.645	Y - N *p* < .001
		Mdn	3 (n=324)	3 (n=194)	3 (n=53)	*p* < .001	
	Lost challenges	*MR*	251.02	333.41	277.01	H(2) = 32.779	Y - N *p* < .001
		Mdn	3 (n=315)	4 (n=195)	3 (n=53)	*p* < .001	
	Won challenges	*MR*	256.40	324.08	279.34	H(2) = 22.092	Y - N *p* < .001
		Mdn	3 (n=315)	3 (n=195)	3 (n=53)	*p* < .001	

3.2. Data privacy-related concerns about fitness tracking applications (RQ2)

The second research question addresses differences in data privacy-related concerns (Table 3) about fitness tracking applications between current users, former users, and non-users of fitness tracking applications. The Kruskal-Wallis H Test revealed significant differences in distribution between some of the groups for all concerns, except for "health insurances will access my data and use it against me." For the remaining concerns there are significant differences in distributions for former users and non-users and additionally between current users and former users (for the concerns that "collected data is too sensitive" and "the app companies will forward my personal data to third parties"). Interestingly, the former users seem less concerned about the listed aspects and see most of them as neutral (median equals 3), except for the concern that "it will be possible to create an exact profile of my movements, habits or preferences," which they slightly agree with (median equals 4). The users and non-users on average agree with all the statements (median equals 4).

Table 3. Differences in data privacy concerns about fitness tracking by mean ranks (MR) and medians (Mdn).

Concerns about fitness tracking applications		Users (Y)	Non-Users (N)	Former Users (F)	Kruskal-Wallis H Test	Post-Hoc Test
Collected data is too sensitive.	*MR* *Mdn*	285.26 4 (n=323)	303.99 4 (n=194)	212.51 3 (n=52)	H(2) = 13.528 *p* = .001	Y - F *p* = .007 N - F *p* = .001
The app companies will forward my personal data to third parties.	*MR* *Mdn*	286.49 4 (n=320)	292.76 4 (n=196)	221.19 3 (n=49)	H(2) = 8.558 *p* = .014	Y - F *p* = .020 N - F *p* = .013
Health insurances will access my data and use it against me.	*MR* *Mdn*	280.26 4 (n=322)	292.55 4 (n=193)	251.21 3 (n=48)	H(2) = 2.713 *p* = .258	-
The app companies will misuse my data.	*MR* *Mdn*	273.73 4 (n=318)	300.64 4 (n=194)	243.96 3 (n=48)	H(2) = 6.428 *p* = .040	N - F *p* = .074
I have no control over what will happen to my data.	*MR* *Mdn*	280.19 4 (n=322)	306.82 4 (n=197)	230.01 3 (n=50)	H(2) = 10.022 *p* = .007	N - F *p* = .007
It will be possible to create an exact profile of my movements, habits or preferences.	*MR* *Mdn*	282.35 4 (n=322)	299.05 4 (n=195)	235.90 4 (n=50)	H(2) = 6.585 *p* = .037	N - F *p* = .032
There will be interference risks from hackers and other unauthorized parties.	*MR* *Mdn*	284.66 4 (n=320)	292.13 4 (n=194)	231.32 3 (n=50)	H(2) = 6.113 *p* = .047	N - F *p* = .044

3.3. Fitness tracking user types by privacy concerns (RQ3)

The final research question concerns identifying and characterizing subgroups of fitness tracking applications' users based on their perceived sensitivity of different data pieces and privacy concerns.

The *K*-means cluster analysis with K_1=3 revealed three very distinctive groups of users. For better identification of data privacy concerns, the medians for each cluster and data piece were aggregated into groups of perception as "sensitive" (for medians equaling 4 or 5), "neutral" (median equaling 3) and "not sensitive" (medians equaling 1 or 2), see Table 4.

The first cluster (CL1, with 64 users) includes users that can be described as rather cautious about data sensitivity, since except for "gender" (perceived as neutral), all remaining data pieces are regarded as sensitive. A more detailed differentiation between "sensitive" and "very sensitive" perception of data pieces can be gathered from Table 5. Here, we can see that for CL1, the most sensitive data pieces are "contacts /friends," most of the health-related data pieces, and the GPS location.

The second cluster (CL2, with 120 users) can be described as rather neutral or balanced in the valuation of the data pieces. Here, eleven of the data pieces (personal and health-related information) is perceived as sensitive (however, only "GPS" is valued as "very sensitive" (Table 5)). Most of the activity and progress data is perceived as neutral. The "not sensitive" information pieces are gender, step count, and climbed stairs.

Finally, the third cluster (CL3, with 43 users) can be described as rather indifferent or unconcerned about the different data pieces. The only sensitive data seem to be the "e-mail," "contacts/friends," and the "GPS" location (however, none of them are perceived as "very sensitive"). The data pieces "real name," "birthday," "interest groups," and "menstrual cycle" are perceived as neutral, whereas others are seen as "not sensitive."

In order to detect further differences between the three clusters that could influence the perceived data sensitivity, the cluster membership of each case was saved into a new variable and the Kruskal-Wallis H Test was conducted for these subgroups of fitness tracking applications' users. Several factors, e.g., fitness level or origin, were investigated.

Table 4. Results of *K*-means clustering procedure on perceived data sensitivity, grouped into perception as "not sensitive" (1-2), "neutral" (3) and "sensitive" (4-5). Abbreviations: Blood Pressure (BP), Heart Rate (HR).

	CL1 (n=64)	**CL2 (n=120)**	**CL3 (n=43)**
Sensitive	Real name, Birthday, E-Mail, Contacts/friends, Interest groups, Calories (burned/intake), HR, BP, Sleeping times, BMI, Weight, Menstrual cycle, Step count, GPS, Climbed stairs, Standing hours, Completed workouts, Fitness level or XPs, Trophies or badges, Lost challenges, Won challenges	Birthday, E-Mail, Contacts/friends, Interest groups, HR, BP, Sleeping times, BMI, Weight, Menstrual cycle, GPS	E-Mail, Contacts/friends, GPS
Neutral	Gender	Real name, Calories (burned/intake), Standing hours, Completed workouts, Fitness level or XPs, Trophies or badges, Lost challenges, Won challenges	Real name, Birthday, Interest groups, Menstrual cycle
Not sensitive		Gender, Step Count, Climbed stairs	Gender, Calories (burned/intake), HR, BP, Sleeping times, BMI, Weight, Step count, Climbed stairs, Standing hours, Completed workouts, Fitness level or XPs, Trophies or badges, Lost challenges, Won challenges

Table 5. Results of *K*-means clustering procedure on perceived data sensitivity (scale from 1 to 5).

Data Pieces		**CL1 N=64**	**CL2 N=120**	**CL3 N=43**
Personal data	Real name	4	3	3
	Gender	3	2	2
	Birthday	4	4	3
	E-Mail	4	4	4
	Contacts/Friends	5	4	4
	Interest groups	4	4	3
Health-related data	Calories intake	4	3	2
	Burned calories	4	3	2
	Heart rate	4	4	2
	Blood pressure	5	4	2
	Sleeping times	5	4	2
	BMI	5	4	2
	Weight	5	4	2
	Menstrual cycle	5	4	3
Activity & progress data	Step count	4	2	1
	GPS	5	5	4
	Climbed stairs	4	2	2
	Standing hours	4	3	2
	Workouts	4	3	2
	Fitness level, XPs	4	3	2
	Trophies, badges	4	3	2
	Lost challenges	4	3	2
	Won challenges	4	3	2

Indeed, the Kruskal-Wallis H Test revealed significant differences in distribution between the three clusters (CL1-CL3) for the fitness or exercise activity (ranging from 1 to 8), H(2) = 10.628, *p* = .005; CL1 (Mean Rank = 93.8; Median = 6), CL2 (Mean Rank = 118.20; Median = 6.5) and CL3 (Mean Rank = 132.33; Median = 7). According to the post-hoc test, the significant differences are given between CL1 and CL2 (*p* = .039) and between CL1 and CL3 (*p* = .006).

Further significant differences in distribution between the three clusters are given for the general attitude towards online privacy, namely "I am not concerned about security on the internet, e.g. people finding out what websites I visit or getting to know my real identity" (answered on a 5-point Likert scale), H(2) = 6.069, *p* = .048; CL1 (Mean Rank = 99.77; Median = 2), CL2 (Mean Rank = 115.92; Median = 2) and CL3 (Mean Rank = 129.81; Median = 3). There was only one significant difference between CL1 and CL3 (*p* = .047).

The last significant difference in distributions was given for the general opinion on online privacy: "I do not care what companies whose services or applications I use do with my (personal) data" H(2) = 19.326, *p* < .001; CL1 (Mean Rank = 89.20; Median = 1), CL2 (Mean Rank = 116.79; Median = 2), CL3 (Mean Rank = 141.12; Median = 2). The significant differences were given between CL1 and CL2 (*p* = .010) and between CL1 and CL3 (*p* < .001).

According to the Kruskal-Wallis H Test, there were no significant differences between the three clusters regarding the everyday activity level, the usage frequency as well as usage duration of the fitness tracking application, and the age of the participants. In order to detect possible cultural differences in cluster membership between participants from the EU and from the USA, the Pearson Chi^2 was calculated. However, there were no significant differences between participants from these two regions.

The first three clusters were estimated based on the users' perceived sensitivity of different data pieces. Another three clusters (CL4-CL6) were calculated based on the data privacy-related concerns regarding fitness tracking applications (Table 6). Here, the CL4 (n=104) includes users agreeing with the most concerns. Except for the one: "collected data is too sensitive," they highly agree with all the remaining statements (median equals 5). The next cluster, CL5 (n=63), includes rather unconcerned users. They do not agree with the most statements and are neutral (median equals 3) with concerns about the collected data being too sensitive as well as the statement "it will be possible to create an exact profile of my movements, habits or preferences." Finally, the last cluster, CL6 (n=137), consists of users having slight concerns. They somewhat agree with most of the statements, except for the two about the collected data being too sensitive and the one stating that "health insurances will access my data and use it against me," towards which they have a neutral attitude (median equals 3).

Similar to the first three clusters, the Kruskal-Wallis H Test was conducted for the Clusters CL4-CL6. The results show that there are significant differences in distribution between the clusters for general online privacy concerns, namely the statement "I am not concerned about security on the Internet": H(2) = 31.151, $p < .001$; CL4 (Mean Rank = 118.11; Median = 2), CL5 (Mean Rank = 189.06; Median = 3), and CL6 (Mean Rank = 161.80; Median = 2). The post-hoc test revealed significant differences between CL4 and CL5 ($p < .001$) and between CL4 and CL6 ($p < .001$).

There are also significant differences in the agreement with the statement "I do not care what companies whose services or applications I use do with my personal data," H(2) = 34.248, $p < .001$; CL4 (Mean Rank = 119.70; Median = 1), CL5 (Mean Rank = 195.15; Median = 2) and CL6 (Mean Rank = 157.78; Median = 2). According to the post-hoc test, the significant differences are given between all clusters: CL4 and CL6 ($p = .001$), CL4 and CL5 ($p < .001$), and CL6 and CL5 ($p = .008$).

The tests revealed no significant differences between the clusters for the everyday activity level, the fitness or exercise level, the usage frequency and usage duration of the fitness tracking application as well as the age of the user. Furthermore, according to Pearson Chi^2, there were no significant differences in cluster distributions between users from the EU and the USA.

Table 6. Results of *K*-means clustering procedure on data privacy-related concerns regarding fitness tracking applications (scale from 1 to 5).

Concerns	**CL4 (n=104)**	**CL5 (n=63)**	**CL6 (n=137)**
Collected data is too sensitive.	4	3	3
The app companies will forward my personal data to third parties.	5	2	4
Health insurances will access my data and use it against me.	5	2	3
The app companies will misuse my data.	5	2	4
I have no control over what will happen to my data.	5	2	4
It will be possible to create an exact profile of my movements, habits or preferences.	5	3	4
There will be interference risks from hackers and other unauthorized parties.	5	2	4

4. Discussion

How do different groups of participants perceive the sensitivity of various data pieces collected by fitness tracking technologies? And what specific privacy concerns do they have, when thinking about this technology? When comparing current users, former users, and non-users of fitness tracking applications, there are only two significant differences between users and non-users in perception of "personal information"—the sensitivity of "gender" (perceived as neutral or not sensitive) and "interest groups." All other personal data pieces were perceived as at least sensitive by all groups.

More significant differences were given for health-related data. All groups agreed on the sensitivity of information about "menstrual cycle." All remaining information pieces were perceived differently between users and non-users. In general, current users perceive calories ("burned" or "intake") and "heart rate" as neutral, and the remaining data pieces as sensitive. The non-users perceive only "burned calories" as neutral and rest as sensitive. Finally, the former users only perceive information about "BMI," "weight," and "menstrual cycle" as sensitive.

Regarding the activity and progress data, all three groups agree on high sensitivity of "GPS," which confirms the results by Lidynia et al. [19]. Except for "GPS," users and former users perceive all remaining activity and progress data as neutral. Non-users perceive most of the data pieces as neutral, except for "step count" (very sensitive), "fitness level or experience points," and "lost challenges" (rather sensitive). Even though they perceive "lost challenges" as rather sensitive information, their perception of the information about "won challenges" is neutral.

The second research question addressed the data privacy-related concerns about fitness tracking applications. There were no significant differences in distribution between the three groups for the statement "health insurances will access my data and use it against me." In general, the former users seem less concerned about the aspects and see most of them as neutral, except for the concern that "it will be possible to create an exact profile of my movements, habits or preferences," which they slightly agree with. The users and non-users on average agree with all the statements. Here, an interesting question arises, why the former users stopped using these applications or wearables and whether any privacy-related concerns played a role. Since users in this investigation still appear to have some reservations about data privacy, but continue using the fitness tracking technologies, it might not be a key aspect, when making a decision to stop or continue using the technology.

The third research question regarded potential subgroups of fitness tracking applications' users based on their (a) perceived data sensitivity and (b) data privacy-related concerns about fitness tracking applications. The first *K*-means clustering procedure (K_1=3) yield three distinctive subgroups of users: CL1 (*concerned users*, n=64), CL2 (*neutral users*, n=120), and CL3 (*unconcerned users*, n=43). The *concerned users* indeed perceive all data pieces as (very) sensitive, except for "gender" (neutral). The *neutral users* are more balanced in their perception, as only "GPS" was perceived by them as "very sensitive," whereas 11 data pieces (personal and health-related information) as "sensitive." They perceive most of the activity and progress data as neutral and information like "gender," "step count," and "climbed stairs" as "not sensitive." Finally, the *unconcerned users* do not perceive any of the information pieces as "very sensitive," and valued only three data pieces ("e-mail," "contacts/friends," and "GPS") as "sensitive" and four data pieces ("real name," "birthday," "interest groups," and "menstrual cycle") as "neutral." They perceive the remaining information as "not sensitive." The differences between these three clusters are not limited to the perceived data sensitivity.

Subsequent Kruskal-Wallis H Test revealed that the *unconcerned users* are on average the most active ones (regarding "fitness or exercise" activity), followed by *neutral users*. It could also mean that users of activity tracking technologies, who are very active, might not fear the "publicity" of the collected data that supports their healthy lifestyle. As one would probably expect, users who are generally doubtful about data privacy online, are also more concerned about the sensitivity of different data pieces. Their perceived sensitivity of data might be this high due to (perceived) lack of safe (data) environment, where personal data is protected from hackers and other misuse, and due to very limited (or non-existent) trust in the companies who have custody of the data. For example, the *concerned users* tend to disagree more with the statement "I am not concerned about security on the internet" than the *unconcerned users* (who are rather neutral towards it). Furthermore, the *concerned users* tend to strongly disagree with the statement "I do not care what companies whose services or applications I use do with my (personal) data," whereas *neutral users* and *unconcerned users* only somewhat disagree. Interestingly, there are no significant differences between the three user groups regarding age as well as the usage duration and usage frequency of the fitness tracking application. Finally, there was no significant association between the cluster membership and the origin of the users.

The second clustering procedure (K_2=3) involved users' data privacy-related concerns about fitness tracking applications. The identified subgroups include: *highly concerned users* (CL4, n=104, strongly agreeing with almost all statements), *unconcerned users* (CL5, n=63, not agreeing with most of the statement or being neutral), and *slightly concerned users* (CL6, n=137, somewhat agreeing with most of the statements). Further differences between these three subgroups regarded the general online privacy concerns, which were again higher for the cluster with *highly concerned users*. Interestingly, there were no significant differences between the clusters regarding the usage frequency and usage duration of the fitness tracking application, the age of the user as well as for the everyday activity and the fitness or exercise level. There were also no significant differences in distributions between users from the EU and the USA, indicating a rather similar distribution of data related unconcerns between users from these two regions.

6. References

[1] Ball, K., Di Domenico, M.L., and Nunan, D. Big Data Surveillance and the Body-subject. *Body and Society 22*, 2 (2016), 58–81.

[2] Bellman, S., Johnson, E.J., Kobrin, S.J., and Lohse, G.L. International differences in information privacy concerns: A global survey of consumers. *Information Society 20*, 5 (2004), 313–324.

[3] Brashear, T.G., Milne, G., and Kashyap, V. Internet Culture And Information Privacy Concerns In Developing Countries Autoria: Thomas G. Brashear, George Milne, Vishal Kashyap. *English*, (2001), 1–17.

[4] Chen, J., Bauman, A., and Allman-Farinelli, M. A Study to Determine the Most Popular Lifestyle Smartphone Applications and Willingness of the Public to Share Their Personal Data for Health Research. *Telemedicine and e-Health 22*, 8 (2016), 655–665.

[5] Christovich, M.M. Why Should We Care What Fitbit Shares?: A Proposed Statutory Solution to Protect Sensitive Personal Fitness Information. *Hastings Communications and Entertainment Law Journal 38*, 4 (2016), 91.

[6] Crawford, K., Lingel, J., and Karppi, T. Our metrics, ourselves: A hundred years of self-tracking from the weight scale to the wrist wearable device. *European Journal of Cultural Studies 18*, 4–5 (2015), 479–496.

[7] Dennison, L., Morrison, L., Conway, G., and Yardley, L. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. *Journal of medical Internet research 15*, 4 (2013), e86.

[8] Dunn, O.J. Multimple comparisons using rank sums. *Technometrics 6*, (1964), 241–252.

[9] Fietkiewicz, K.J. Jumping the digital divide: How do "silver surfers" and "digital immigrants" use social media? *Networking Knowledge 10*, 1 (2017), 5–26.

[10] Fietkiewicz, K.J., Lins, E., Baran, K.S., and Stock, W.G. Inter-Generational comparison of social media use: investigating the online behavior of different generational cohorts. *2016 49th Hawaii International Conference on System Sciences (HICSS)*, IEEE Computer Society (2016), 3829–3838.

[11] Gorm, N. and Shklovski, I. Sharing Steps in the Workplace: Changing Privacy Concerns Over Time. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM (2016), 4315–4319.

[12] Grundy, Q., Held, F.P., and Bero, L.A. Tracing the potential flow of consumer data: A network analysis of prominent health and fitness apps. *Journal of Medical Internet Research*, (2017).

[13] Hargittai, E. and Marwick, A. "What Can I Really Do?" Explaining the Privacy Paradox with Online Apathy. *International Journal of Communication (19328036) 10*, (2016), 3737–3757.

[14] Henkel, M., Heck, T., and Göretz, J. Rewarding Fitness Tracking---The Communication and Promotion of Health Insurers' Bonus Programs and the Use of Self-tracking Data. *Social Computing and Social Media. Technologies and Analytics*, Springer International Publishing (2018), 28–49.

[15] Hern, A. This article is more than 1 year old Fitness tracking app Strava gives away location of secret US army bases. *The Guardian*, 2018. https://www.theguardian.com/world/2018/jan/28/fitness-tracking-app-gives-away-location-of-secret-us-army-bases.

[16] Klasnja, P., Consolvo, S., Choudhury, T., Beckwith, R., and Hightower, J. Exploring Privacy Concerns about Personal Sensing. *Pervasive Computing*, Springer Berlin Heidelberg (2009), 176–183.

[17] Lehto, M. and Lehto, M. Health Information Privacy of Activity Trackers. *16th European Conference on Cyber Warfare and Security*, (2017), 243–251.

[18] Lidynia, C., Brauner, P., and Ziefle, M. A step in the right direction - Understanding privacy concerns and perceived sensitivity of fitness trackers. *Advances in Human Factors in Wearable Technologies and Game Design*, Springer International Publishing (2018), 42-53.

[19] Lidynia, C., Schomakers, E.M., and Ziefle, M. What are you waiting for? – perceived barriers to the adoption of fitness-applications and wearables. *Advances in Intelligent Systems and Computing 795*, (2019), 41–52.

[20] Lieffers, J.R.L., Vance, V.A., and Hanning, R.M. Use of Mobile Device Applications In Canadian Dietetic Practice. *Canadian Journal of Dietetic Practice and Research 75*, 1 (2014), 41–47.

[21] Miltgen, C.L. and Peyrat-Guillard, D. Cultural and generational influences on privacy concerns: a qualitative study in seven European countries. *European Journal of Information Systems 23*, 2 (2014), 103–125.

[22] Motti, V.G. and Caine, K. Users' privacy concerns about wearables: Impact of form factor, sensors and type of data collected. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 8976*, (2015), 231–244.

[23] Nissenbaum, H. A Contextual Approach to Privacy Online. *Daedalus 140*, 4 (2011), 32–48.

[24] Patterson, H. Contextual Expectations of Privacy in Self-Generated Health Information Flows. *SSRN Electronic Journal*, (2013), 1–48.

[25] Peppet, S.R. Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent - ABI/INFORM Complete - ProQuest. *Texas Law Review*, (2014), 85–1176.

[26] Pingo, Z. and Narayan, B. Users' responses to privacy issues with the connected information ecologies created by fitness trackers. *Lecture Notes in Computer Science 11279 LNCS*, (2018), 240–255.

[27] Shin, G., Jarrahi, M.H., Fei, Y., et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: A systematic literature review. *Journal of Biomedical Informatics 93*, March (2019), 103153.

[28] Stach, C. Big Brother is Smart Watching You - Privacy Concerns about Health and Fitness Applications. Icissp (2018), 13–23.

[29] Statistics, L. Kruskal-Wallis H test using SPSS Statistics. Statistical tutorials and software guides. 2015. https://statistics.laerd.com/.

[30] Vitak, J., Liao, Y., Kumar, P., Zimmer, M., and Kritikos, K. Privacy attitudes and data valuation among fitness tracker users. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 10766 LNCS*, (2018), 229–239.