

RESEARCH ARTICLE

Studying health-related internet and mobile device use using web logs and smartphone records

Ruben L. Bach¹*, Alexander Wenz^{1,2}

1 Collaborative Research Center 884 Political Economy of Reforms, University of Mannheim, Mannheim, Germany, **2** Institute for Social and Economic Research, University of Essex, Colchester, United Kingdom

* These authors contributed equally to this work.

* r.bach@uni-mannheim.de



Abstract

Many people use the internet to seek information that will help them understand their body and their health. Motivations for such behaviors are numerous. For example, users may wish to figure out a medical condition by searching for symptoms they experience. Similarly, they may seek more information on how to treat conditions they have been diagnosed with or seek resources on how to live a healthy life. With the ubiquitous availability of the internet, searching and finding relevant information is easier than ever before and a widespread phenomenon. To understand how people use the internet for health-related information, we use data from a sample of 1,959 internet users. A unique combination of data containing four months of users' browsing histories and mobile application use on computers and mobile devices allows us to study which health websites they visited, what information they searched for and which health applications they used. Survey data inform us about users' socio-demographic background, medical conditions and other health-related behaviors. Results show that women, young users, users with a university education and nonsmokers are most likely to use the internet and mobile applications for health-related purposes. On search engines, internet users most frequently search for pharmacies, symptoms of medical conditions and pain. Moreover, users seem most interested in information on how to live a healthy life, alternative medicine, mental health and women's health. With this study, we extend the field's understanding of who seeks and consumes health information online, what users look for as well as how individuals use mobile applications to monitor their health. Moreover, we contribute to methodological research by exploring new sources of data for understanding humans, their preferences and behaviors.

OPEN ACCESS

Citation: Bach RL, Wenz A (2020) Studying health-related internet and mobile device use using web logs and smartphone records. PLoS ONE 15(6): e0234663. <https://doi.org/10.1371/journal.pone.0234663>

Editor: Simone Borsci, Universiteit Twente, NETHERLANDS

Received: January 22, 2020

Accepted: May 30, 2020

Published: June 12, 2020

Copyright: © 2020 Bach, Wenz. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Minimal Data Sets to reproduce all tables and figures in the paper are available through the Open Science Framework at https://osf.io/6v9a5/?view_only=0aa5bd7a616f45acb7419e2542e4d77b With these datasets, all tables and figures can be replicated.

Funding: Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 139943784 – SFB 884. The publication of this article was funded by the Baden-Württemberg Ministry of Science, Research and the Arts and by the University of Mannheim.

Introduction

Many people use the internet to seek information that will help them understand their body and their health [1–3]. Motivations for such behaviors are numerous. For example, users may wish to figure out a medical condition by searching for symptoms they experience. Similarly,

Competing interests: The authors have declared that no competing interests exist.

they may seek more information on how to treat conditions they have been diagnosed with or seek resources on how to live a healthy life. With the ubiquitous availability of the internet, searching and finding relevant information is easier than ever before and a widespread phenomenon. In the U.S., up to two out of three adults regularly search for health information online [4, 5]. Likewise, two out of five internet users in Germany search the internet for health information before their doctor's appointment, and around half of internet users after their appointment [6].

In addition, the rise of health applications (apps) on mobile devices such as smartphones and tablets, as well as accompanying health and fitness trackers ("wearables"), make it possible for people to track their health and fitness without the help of medical professionals and at lower costs. Just like researching health information online, the use of health apps on mobile devices is spreading rapidly. For example, about one in five smartphone owners in the U.S. used a health app in 2012 [7]. Likewise, about one in four U.S. citizens regularly or occasionally use health apps for self-diagnosis [8] and up to 45% of U.S. citizens report using a mobile phone or tablet to manage their health [5]. The numbers for health app use in Germany are very similar: Around two out of three smartphone owners used a health app in 2019 [6]. Moreover, the number of health apps in Apple's App Store is estimated to be about 90,000 [9].

Understanding who uses health apps on mobile devices and searches and consumes what kind of health information on the internet is crucial for several reasons. Health apps and mobile devices are used to improve self-reflection [10], change behaviors [11] and track physical activity [12]. Likewise, health information collected from the internet influences users' decisions about their own health and decisions they make for others (e.g., their children) [13]. Similarly, it may affect users' decisions when to go see a physician or to change eating habits and physical activity [14–16]. Moreover, health-related internet use has the potential to reduce shortcomings in health knowledge for certain subgroups of the population, such as individuals with lower education [17]. Online health information may also influence how users treat conditions or symptoms they experience. However, (mis)information disseminated via the internet may also drive and facilitate the emergence of phenomena such as the spread of anti-vaccination sentiments [18]. This point is exacerbated by the fact that users often do not check the source and data of health information found online [19].

In this paper, we study health-related internet and app use by relying on a unique data set covering passively tracked browsing behavior of 1,959 German internet users over a period of four months in 2017. For some of these users, browsing behavior as well as app use was in addition monitored on smartphones and tablets during the same time period. The data were collected by a commercial vendor who keeps a pool of participants who are occasionally invited to answer short surveys for money.

Previous work mostly relied on self-reported measures of online behavior through questionnaires, which is often inaccurate due to recall error, and previous research often used samples small in size. Our approach of passively tracking browsing behavior addresses these limitations and allows more accurate and detailed insights into how people use the internet. In addition to the tracking data, participants also provided socio-demographic information, information on health issues and information on other lifestyle behaviors that may influence their health (e.g., exercising and smoking). This combination of web logs and survey data creates a unique data source for studying individuals' health activities in the online world, going far beyond previous research. Overall, we demonstrate what the field can learn from such records of individuals' online and app activities and point to avenues for future research.

Using these data, we examine who engages in health-related internet and app use and to what extent. That is, we study how online health information searches and app use are

associated with socio-demographic characteristics, health conditions and health-related lifestyle behaviors. Results indicate that women, young users and users who have a university degree are more likely to engage in health-related internet/app use. We find limited evidence that health-related internet/app use is related to health conditions, but that it is related with lifestyle behaviors such as smoking. Overall, however, frequent use of health-related apps is not a widespread phenomenon: Only 16% of all app users frequently use a health app. Regarding topics that users engage in the most, we classify participants' visits to health-related internet domains and apps into broader categories. Our results indicate that users are particularly interested in exercising and weight loss as well as nutrition and alternative medicine. Analysis of health-related search queries made to search engines reveals that users are most interested in finding pharmacies, symptoms of medical conditions, various forms of pain and remedies for health problems.

Related work

In this section, we review two streams of previous work that are relevant for our study: Correlates of health-related internet and app use and health information search through search engines.

Correlates of health-related internet and app use

Internet use for health information seeking is associated with a variety of socio-demographic characteristics. Consistent findings are reported regarding education (users with lower education are less likely to seek health information online) [4, 19–21, 21–24] and gender (women more likely) [4, 17, 19, 20, 21, 21, 22, 25]. Moreover, previous research agrees regarding the role of age in determining health-related internet use (younger users more likely) [4, 19, 21, 23, 24]. Some studies also report effects regarding income (users with higher income more likely) [4, 21, 22] and race/ethnicity (non-white users less likely) [4, 25].

Less is known regarding associations between users' health behavior and health-related internet use. Several studies report that users with "fair" or "poor" health are more likely to use the internet for health purposes [22, 26]. Users who already have a chronic health condition are more likely to seek information online [27], while those who are at risk of getting a medical condition (for example, cancer) rely more often on information obtained from health professionals [23]. Furthermore, there seems to be a positive association between obesity and health-related internet use [28].

Regarding determinants of health-related app use, socio-demographic relationships similar to those mentioned above are reported (e.g., age, education, gender, income) [7, 29–32]. In addition, having a history of chronic medical conditions [33], being obese [31] and engaging in physical activity are all positively associated with health-related app use [30, 33]. Overall, however, regular use of health apps and other digital health solutions is not widespread, even among users with a high digital affinity (see, for example, [34]).

To sum up, previous work identified several socio-demographic correlates of health-related internet and app use, such as age, gender and education. Moreover, a few studies find relationships between users' health and their online and app activities. One major drawback of all of the studies mentioned here, however, is that they rely on survey data. That is, users self-report whether they engage in health-related internet or app use. However, it is well known that survey reports are often inaccurate as users tend to forget or overestimate actual internet use (see, e.g., [35]). Analyzing web logs and records of mobile device use such as those used in our study offers a more complete and fine-grained picture of users' online and app activities.

Moreover, they allow us to study not only who engages in health-related internet and app use, but also how users obtain their information.

Health information search through search engines

Besides literature regarding the question of who uses the internet and apps for health-related purposes, previous work studying health-related search queries is relevant for our study. Most of the studies in this domain rely on the analysis of search query data.

Cartright et al. identify health-related search queries (about 20% out of all queries) made to three major search engines in the U.S. over a period of six months [36]. The resulting queries are classified into different foci (symptom, cause or remedy). In addition, the authors train a classifier that predicts what the next focus of a user in a single session will be. Using similar data, White et al. show that users who start a session searching for simple symptoms easily end up searching for serious diseases [13]. Thereby, the authors show that likely innocuous health searches can quickly lead users to seek information about serious, but rare disease with similar symptoms. Using Google search queries, Ginsberg et al. demonstrate that (for some time) search activity for influenza-like symptoms was an accurate predictor of actual influenza epidemics [37]. A few years later, however, the performance of their algorithms decreased rapidly due to changes in Google's search algorithms and in the ways people used the search engine [38].

Abebe et al. study health information needs related to HIV/AIDS, malaria and tuberculosis in 54 African countries [39]. Using Bing search data from those countries, the authors show that users are mostly interested in gathering information about symptoms, testing and treatment, but also stigma, discrimination and natural cures. Using 18 months and billions of search queries posted to Bing's web search engine, Fourney et al. show how concerns about pregnancy and childbirth change over the course of pregnancy [40].

Furthermore, Yahoo! search activity for cancer correlates with estimated cancer incidence, mortality and, especially, news coverage [41]. Similarly, search activity for information about cancer in the U.K. and the U.S. increased from 2008 to 2010, with almost half of all searches dealing with breast cancer, followed by lung and prostate cancer [42]. Most common topics include different treatment forms, diagnosis and screening.

Google Trends, a tool allowing the estimation of aggregated Google search activity for specific queries, is popular for studying health-related search activity [43–45]. For example, research found that queries about breast cancer screening made to Google correlate with changes in legislature and news coverage [44]. Likewise, Google Trends shows that consistent seasonal patterns in search activity exist and that breast, pancreatic and ovarian cancer are among the most searched for forms of cancer [43, 45].

Another approach to obtain information about users' online search behavior is used in [46]. 56 women were recruited from a commercial vendor in market research. Women answered an online survey and were then instructed to search information online about a hypothetical body change. Monitoring the online searching behavior of the participants through a browser plugin, the authors find that seeking information about unfamiliar symptoms online does not necessarily help women understand their condition.

To sum up, previous work mainly relied on search query data obtained from various search engines or tools built on top of them (Google Trends). One drawback of the latter is that it does not allow analyses on the user level because Google Trends only provides aggregate search activity information. In addition, even if user-level data are available (such as in [36] or [13]), the data does not contain detailed socio-demographic and/or additional health information about users. Moreover, query data is often difficult to obtain [40] or relies on small samples with specific foci [46].

Data and methodology

We analyze four months (July-October 2017) of web log and mobile device use data for a sample of 1,959 German internet users. Data were collected by a commercial vendor (respondi AG [47]) in market research for a third-party (SINUS Markt- und Sozialforschung GmbH [48]) not involved in this study for market research purposes. The vendor keeps an opt-in panel of participants for marketing and social research. Participants of the panel are occasionally invited to answer surveys in exchange for small cash incentives. In addition, participants can agree to the monitoring of their browsing behavior and mobile device use for additional incentives, after giving informed consent that their survey, web log and app use data will be used for academic and market research purposes. Details on the panel, the vendor's recruitment process as well as the privacy policy can be found at [47].

To ask participants questions, researchers provide them to the vendor who then implements them in its online survey platform. At no times is there contact between researchers and participants of the vendor's panel. Moreover, the vendor provides all data in pseudonymized and de-identified form. That is, all data accessible to us are stripped of users' names, addresses or birth dates and cannot be linked back.

Given these circumstances (pseudonymized data provided by a survey platform from users of the platform who gave informed consent in combination with no possibility for us to de-identify individuals), this study was exempted from an approval by the Ethics Committee of the University of Mannheim (Ethics Committee of the University of Mannheim, Decision "EK Mannheim 15/2020").

The vendor gathers web logs from participants' personal computers and mobile devices (smartphones and tablets) through a tool based on software provided by Wakoopa [49]. Users install a plugin in web browsers used on their personal computers (e.g., Safari, Firefox, Microsoft Edge, Chrome). In addition, they download an app on their mobile devices (smartphones and tablets, Android and iOS devices only). This app collects web logs from the native browsers (i.e., Safari on iOS devices and Chrome on Android devices) as well as information about the apps participants use.

Each time a participant navigates to a website, the complete URL of the website (e.g., <https://en.wikipedia.org/wiki/URL>), the domain ([wikipedia.org](https://en.wikipedia.org/wiki/URL)), the current date and time as well as the time spent on the website are recorded (both on personal computers and mobile devices). In addition, on mobile devices, information about the apps that participants use are recorded. Every time a participant opens an app on a device, the name of the app, the duration of use and information about the device are logged. Information on activities that individuals perform in an app are not recorded. At any time, participants can turn off data collection temporarily or stop data collection completely.

The vendor also provides background information about the users, including socio-demographics and information on various health issues, which were collected through a web survey. Age, gender, and education quotas were used to achieve a sample approximately representing the German adult population. [Table 1](#) shows characteristics of the participants in our study. Overall, about half of all participants are female (54.72%), and the mean age is about 42 years. 58.96% of participants work full- or half-time. More than 80% have some education beyond basic secondary school and most of the participants have a personal net monthly income between €1,000 and €2,000. Besides socio-demographic information, [Table 1](#) lists also the most common health issues reported by the participants. 32.06% of all participants indicated back problems and about 26.08% of having (any) allergies. 20.93% reported having high blood pressure, 18.07% problems with sleeplessness, 16.23% depression and 11.84% reported obesity. Regarding other lifestyle behaviors, 41.45% indicated smoking and 76.88% participating in any

Table 1. Descriptive statistics of sample.

<i>Variable</i>		<i>%</i>
Gender	Female	54.72
	Male	45.28
Age	(mean)	41.87
	(standard deviation)	14.50
Education	Basic secondary school	19.50
	Extensive secondary school	36.35
	High school	23.99
	University degree	20.16
Employment status	Work full-time or part-time	58.96
	Do not work	41.04
Personal income	≤ €999	32.01
	€1,000-€1,999	36.86
	≥ €2,000	31.14
Health issues	Back problems	32.06
	Allergies	26.08
	High blood pressure	20.93
	Sleeplessness	18.07
	Depression	16.23
	Obesity	11.84
Lifestyle	Smoking	41.45
	Physical Activity	76.88

N = 1,959

Physical activity denotes whether the participant engages with at least one of the following activities: aerobics, badminton, basketball, fitness, football, handball, hockey, jogging, judo, karate, Nordic walking, Pilates, horse riding, swimming, squash, dancing, diving, tennis, volleyball, yoga, cycling or mountain biking, golf, sailing, skiing, surfing.

<https://doi.org/10.1371/journal.pone.0234663.t001>

physical activity (such as running, swimming and playing football). Thus, we find that some participants in the sample do have to cope with several health problems and engage in health-related lifestyle behaviors.

We create three datasets from the data described above. The first one contains web logs from both personal computers and mobile devices (28,524,036 total logs from 1,959 participants). Overall, participants visited 194,389 unique domains. To identify web logs that refer to domains with health content, we used Webshrinker [50], an online service offering domain categorization. Each of the 194,389 unique domains found in the web logs dataset was categorized into one of the 26 categories of the Interactive Advertising Bureau's domain taxonomy [51]. If available, Webshrinker also included the appropriate subcategory (for example, chronic pain, dental care or alternative medicine). We then defined a binary indicator denoting whether a domain belongs to the category "health and fitness" or to a different category. Furthermore, we recorded the subcategories for all domains with the "health and fitness" category. Overall, 10,371 out of the 194,518 unique domains (that is, 5.33%) were categorized as being health-related. Table 2 shows the number of health-related domains per subcategory.

The second dataset contains records of app use (8,957,760 total records from 1,328 participants). The remaining 631 participants either used a personal computer only or did not use apps on their mobile devices. Therefore, the number of participants in this dataset is smaller than the total number of participants in our study. Overall, participants used 10,123 unique

Table 2. Number of unique domains, by subcategory.

<i>Subcategory</i>	<i>Number of domains</i>
Abortion	14
AIDS/HIV	30
Allergies	32
Alternative Medicine and Holistic Healing	1,031
Arthritis	12
Asthma	7
Attention Deficit Disorder	2
Autism	25
Bipolar Disorder	6
Brain Tumor	30
Cancer	102
Cholesterol	16
Crohn's Disease	64
Chronic Fatigue Syndrome	12
Chronic Pain	62
Cold and Flu	48
Deafness	70
Dental Care	423
Depression	25
Dermatology	371
Diabetes	95
Epilepsy	10
Exercise and Weight Loss	1,248
GERD/Acid Reflux	26
Headaches/Migraines	73
Health and Fitness (no subcategory)	3,071
Heart Disease	100
Incest/Abuse Support	113
Incontinence	39
Infertility	29
Men's Health	396
Nutrition	142
Orthopedics	326
Panic/Anxiety Disorders	30
Pediatrics	93
Physical Therapy	181
Psychology/Psychiatry	700
Senior Health	67
Sleep Disorders	95
Smoking Cessation	53
Substance Abuse	74
Thyroid Disease	50
Vitamins and Food Supplements	576
Women's Health	402
Total	10,371

Health-related domains only. Web logs dataset.

<https://doi.org/10.1371/journal.pone.0234663.t002>

apps. Health-related apps were identified based on the classification of apps used in both Android's Playstore and Apple's App Store. All apps classified as "health and fitness" or "medicine" were labelled as health-related. Overall, 476 out of the 10,123 unique apps (that is, 4.70%) were categorized as being health-related. To get a more detailed insight into these apps, we manually coded all apps labelled as health-related into one of the subcategories shown in [Table 3](#).

Closer inspection of the app use records reveals, however, that only about one third of health apps are actually frequently used (column three of [Table 3](#)). Limiting app use records to apps that were used for at least thirty minutes by at least one user (during the four months of data collection), the number of apps decreases to 157. Thus, it seems that many health apps are

Table 3. Number of apps, by subcategory.

<i>Subcategory</i>	<i>Number of apps</i>	<i>Frequently used apps</i>
Allergies	2	0
Alternative Medicine	3	0
Baby Care	12	4
Beauty Care	1	0
Blood Pressure	2	2
Children	3	0
Dental Care	4	2
Diabetes	9	2
Donate Blood	2	0
First Aid	5	2
General Health Information	31	6
Health Diary Keeping	9	4
Health Insurance	19	4
Health Tracking	69	38
Heart	5	1
Hydration	19	7
Information for Disabled People	1	0
Meditation	27	6
Mental Health	3	0
Migraine	2	1
Neck and Back Problems	2	0
Nutrition	11	2
Palliative Care	1	0
Pregnancy	21	7
Reminder	3	3
Sexual Health	5	2
Sleep	27	5
Smoking Cessation	8	1
Tinnitus	1	0
Unclear	1	0
Veins	1	0
Weight Loss	47	16
Women's Health	22	13
Workout and Exercise	98	29
<i>Total</i>	<i>476</i>	<i>157</i>

Health-related apps only. Apps dataset. Frequently used apps: Used for at least thirty minutes by at least one user.

<https://doi.org/10.1371/journal.pone.0234663.t003>

hardly ever used. Regarding apps that users actually use, we find that tracking one's health and fitness, as well as exercising and weight loss and women's health are the popular categories.

The third dataset is a subset of the first one. It contains search queries sent to the 19 most common search engines in the web logs dataset (for example, Google, Bing, Yahoo and DuckDuckGo). From the first dataset, 1,197,421 (4.20%) URLs point to search engines. We extract the search queries from these URLs (for example, the search queries extracted from the URL <https://www.google.com/search?q=high+blood+pressure> are "high blood pressure"). 1,656 (84.53%) participants used one of the search engines at least once. In order to identify health-related search queries in this dataset, we scraped health-related terms from ten German websites. We chose websites that listed diseases and organs and other parts of the human body, technical medical terms, medical disciplines, symptoms of diseases and medical encyclopedia. These websites are listed below. Prior to scraping the terms from the websites, we ensured that our web crawlers were not forbidden on these websites by checking the robots.txt files and the terms and conditions of each domain.

- <https://de.wiktionary.org/wiki/Verzeichnis:Deutsch/Medizin>
- <https://de.wikiquote.org/wiki/Kategorie:K%C3%B6rperteile>
- <https://de.wiktionary.org/wiki/Verzeichnis:Latein/K%C3%B6rperteile>
- <https://www.taschenhirn.de/mensch-und-natur/haeufigste-krankheiten-der-welt/>
- <https://www.taschenhirn.de/mensch-und-natur/organe-des-menschen/>
- <https://www.netdokter.de/krankheiten/#Lexikon>
- <https://www.gelbe-liste.de/krankheiten>
- https://flexikon.doccheck.com/de/Liste_von_Medizinprodukten
- https://de.wikipedia.org/wiki/Liste_medizinischer_Fachgebiete
- https://www.onmeda.de/krankheiten/krankheiten_az.html

We then compared string similarity between each search query from the search query dataset and the list of scraped medical terms. We used the `fuzzywuzzy` Python module for fuzzy string matching (<https://github.com/seatgeek/fuzzywuzzy>). The module allows the estimation of differences between sequences of characters based on Levenshtein Distances.

Manual inspection of a random sample of matches between the search queries and the list of medical terms revealed that using a sorted token approach with a partial match ratio of 0.7 resulted in a reasonably low number of false positives. That is, we classified those search queries as health-related where the match ratio between a search query and any of the entries from the list of medical terms was at least 0.7. All other search queries were classified as not health-related. With this approach, we classified 9,278 out of 1,197,421 search queries (0.76%) as health-related. 763 users out of the 1,656 users who used a search engine at least once (that is, 46.07%) searched for health-related information at least once.

Findings

In this section, we present the main results of our study.

Health-related internet and app use

1,662 (84.84%) out of 1,959 participants in the web logs dataset visited any health-related domain. [Table 4](#) shows the five most popular subcategories of health-related domains across

Table 4. Top subcategories of health-related domains and apps.

<i>Subcategory</i>	<i>Number of participants</i>
Web Logs Dataset	1,959 (100%)
<i>Any Health Domain Visited</i>	1,662 (84.84%)
Exercise and Weight Loss	1,050 (53.60%)
Vitamins and Food Supplements	943 (48.14%)
Alternative Medicine	809 (41.30%)
Psychology/Psychiatry	588 (30.02%)
Dermatology	555 (28.33%)
Apps Dataset	1,328 (100%)
<i>Any Health App Used</i>	494 (37.20%)
Health Tracking	200 (15.06%)
Workout and Exercise	145 (10.92%)
Weight Loss	105 (7.91%)
Women's Health	100 (7.53%)
<i>Frequent Health App Use</i>	224 (16.87%)
Health Tracking	98 (7.38%)
Workout and Exercise	47 (3.54%)
Weight Loss	45 (3.39%)
Women's Health	20 (1.51%)

Web logs and apps datasets.

<https://doi.org/10.1371/journal.pone.0234663.t004>

participants in the sample. Participants seem predominantly interested in information about exercising, losing weight and food supplements, but also in alternatives to traditional medicine. Moreover, mental health and dermatology are popular health topics. A comparison with Table 2 demonstrates that the subcategories with the highest numbers of unique visitors are also among those with the highest number of unique domains.

494 (37.20%) out of 1,328 participants in the apps dataset used a health app at least once. Using the definition of *frequent* health app use (see Table 3), we find that only 224 (16.87%) out of 1,328 participants frequently use health apps. Regarding the most popular subcategories of health apps measured via the number of users, Table 4 shows that users are predominantly interested in monitoring and tracking their health. This finding is likely due to the popularity of wearables that allow, for example, the monitoring of one's heart rate. Similarly, many participants used an app for exercising and workout, but also apps that focus on weight loss. Furthermore, apps for women's health are popular. Manual inspection revealed that this subcategory mainly consists of period and ovulation tracker apps. Interestingly, the same categories are also the most popular ones when we restrict the analysis to frequent app users. In addition, the popularity of apps among participants seems to match the popularity of subcategories measured by the number of unique apps in each subcategory (Table 3). Overall, however, more people seem to browse the internet rather than use apps for health-related purposes.

To better understand who uses the internet and apps for health-related purposes, we estimate logistic regression models. Results are shown in Table 5. The models predict a binary variable indicating whether a participant visited any health-related domain, used any health app, used a health app frequently (that is, app use ≥ 30 minutes) or made any health-related search query. Predictors are socio-demographic characteristics, health conditions and lifestyle indicators.

Table 5. Odds ratios of logistic regression models predicting health-related online activities.

	<i>Health domains</i>	<i>P value</i>	<i>Health apps (freq. use)</i>	<i>P value</i>	<i>Health apps</i>	<i>P value</i>	<i>Health searches</i>	<i>P value</i>
Intercept	3.79 (1.98-7.25)	<.001	0.71 (0.38-1.30)	.27	0.26 (0.12-0.60)	<.001	0.50 (0.31-0.82)	.01
Age	1.01 (1.00-1.02)	.26	0.99 (0.98-1.00)	.01	0.98 (0.97-1.00)	.02	1.00 (0.99-1.01)	.67
Female	1.45 (1.11-1.89)	.01	1.17 (0.92-1.49)	.20	1.46 (1.05-2.01)	.02	1.31 (1.08-1.60)	.01
Basic sec. school (reference)	—	—	—	—	—	—	—	—
Extensive sec. school	1.11 (0.78-1.57)	.57	1.25 (0.88-1.77)	.21	1.09 (0.68-1.75)	.71	0.88 (0.68-1.15)	.35
High school	1.13 (0.75-1.68)	.56	1.36 (0.92-1.99)	.12	1.35 (0.82-2.23)	.24	1.05 (0.78-1.43)	.74
University degree	1.35 (0.88-2.07)	.17	1.14 (0.76-1.70)	.52	0.97 (0.57-1.66)	.91	1.53 (1.13-2.08)	.01
Employed	0.88 (0.65-1.19)	.40	0.95 (0.71-1.26)	.70	0.88 (0.60-1.27)	.49	1.03 (0.83-1.28)	.80
Income: ≤€999 (reference)	—	—	—	—	—	—	—	—
Income: €1,000-€1,999	0.98 (0.68-1.40)	.90	1.12 (0.81-1.55)	.50	1.04 (0.68-1.59)	.84	0.93 (0.72-1.20)	.57
Income: ≥€2,000	0.91 (0.62-1.34)	.63	0.96 (0.67-1.38)	.83	1.15 (0.72-1.82)	.56	0.83 (0.63-1.11)	.21
Back problems	0.95 (0.70-1.30)	.77	1.12 (0.85-1.49)	.42	0.95 (0.66-1.38)	.79	0.90 (0.72-1.12)	.34
Allergies	1.24 (0.91-1.70)	.17	1.11 (0.85-1.44)	.44	1.38 (0.99-1.91)	.06	0.94 (0.76-1.17)	.58
High blood pressure	1.20 (0.83-1.73)	.34	0.85 (0.60-1.20)	.35	0.89 (0.56-1.43)	.64	1.31 (1.01-1.70)	.04
Sleeplessness	1.28 (0.86-1.90)	.23	1.21 (0.89-1.72)	.30	0.86 (0.55-1.35)	.50	1.06 (0.81-1.39)	.68
Depression	1.45 (0.95-2.22)	.09	1.17 (0.83-1.65)	.36	1.29 (0.83-2.00)	.26	1.38 (1.05-1.82)	.02
Obesity	1.33 (0.82-2.14)	.25	1.37 (0.95-1.99)	.09	1.30 (0.81-2.09)	.27	1.17 (0.86-1.58)	.32
Smoking	0.69 (0.53-0.89)	.01	0.73 (0.58-0.93)	.01	0.54 (0.39-0.74)	<.001	0.89 (0.73-1.08)	.23
Do any sports	0.88 (0.64-1.19)	.40	1.11 (0.84-1.48)	.47	1.17 (0.80-1.71)	.43	0.97 (0.78-1.21)	.80
<i>Number of participants</i>	1,959		1,328		1,959			

95% Confidence intervals in parentheses. Frequent use: ≥ 30 minutes use time in total by user.

<https://doi.org/10.1371/journal.pone.0234663.t005>

Columns two and three indicate that women are more likely to browse the internet for health-related purposes, while those who smoke are less likely to seek health information online. Regarding health-related app use (fourth and fifth column), we find that the likelihood of using a health app decreases with age and users who smoke are also less likely to do so. Replacing the dependent variable *any* health app use with *frequent* health app use (columns six

and seven), we find similar results. Frequent health app use decreases with age and smoking, but we also find that women are more likely to be frequent health app users. For all three models, we do not find that medical conditions (such as back problems or having a high blood pressure) play a substantial role. The last two columns of [Table 5](#) show results regarding the question who is most likely to *search* for health information using a search engine. Among all internet users in our data, female users are more likely to search health information online as well as users with a university degree. Moreover, those who have reported a high blood pressure and depression are more likely to search information via search engines.

Next, we analyze whether the frequency and duration of using the internet and apps for health-related purposes shows similar patterns. The fine-grained records of online behavior and app use allow us to calculate for each individual how often and how long they used the internet/apps for health-related purposes. We consider the sum of all health-related online activity (that is, the sum of internet use, app use and search engine use). To account for different base levels of online and app activities, we divide each individual's overall online health activity by the same individual's overall online and app activity. [Table 6](#) shows the results from the linear regression models. Again, female users spent more of their total online and app activities with health-related activities. Moreover, those individuals who report high blood pressure spent more of their total activity with health-related activities. Similar to the results shown in [Table 5](#), smokers invest less time into health-related online and app activities.

Health-related search queries

[Fig 1](#) shows the most frequently used words across participants in the search query dataset after removing stopwords. Many participants searched for pharmacies. One motivation may be to find the closest pharmacy, one that is open or to order from an online pharmacy. Moreover, popular terms cover, for example, symptoms, pain and therapy, but also medicine and several organs or body parts (such as, skin and chest). Furthermore, women's health as well as children and babies were important topics.

In addition to single words, we also considered the most frequently used bigrams (a sequence of two adjacent words) across users ([Table 7](#)). Analyzing bigrams allows a better understanding of which topics users search for. Similar to the most frequently used single words, pharmacies are among the most popular bigram words. Moreover, information about specific diseases, but also about rather general health issues (dry skin) and nutrition are sought.

Discussion and conclusion

We used a unique combination of records of individuals' online and app activities with socio-demographic and health information in this study. This dataset provided detailed and fine-grained insights into participants' internet and app use for health-related purposes. In addition, we observed not only if participants used the internet and/or apps at all, but also to what extent. We studied which health aspects users are most interested in when browsing the internet and using apps. Previous literature had to rely on inaccurate and often incomplete self-reports from surveys. Such self-reports suffer from bias due to users not being able to accurately recall when and how long they used the internet or apps and what exactly they did [35]. Moreover, we also overcame limitations in sample size, while previous research often had to rely on samples limited in size.

Analyzing health-related search queries made to search engines allowed us to study *how* users obtain health information on the internet. Previous studies often had to rely on openly available, but aggregated and less detailed search query data (see, for example, [44]). Obtaining

Table 6. Linear regression models predicting intensity of health-related online activities.

	Frequency	P value	Duration	P value
Intercept	1.512 (0.436)	<.001	1.578 (0.477)	<.001
Age	-0.013 (0.007)	.06	-0.013 (0.007)	.08
Female	0.536 (0.176)	.002	0.687 (0.193)	<.001
Basic sec. school (reference)	—	—	—	—
Extensive sec. school	0.067 (0.237)	.78	0.125 (0.259)	.63
High school	0.069 (0.272)	.80	-0.167 (0.297)	.57
University degree	0.052 (0.278)	.85	0.034 (0.303)	.91
Employed	-0.025 (0.199)	.90	0.071 (0.218)	.75
Income: ≤€999 (reference)	—	—	—	—
€1,000-€1,999	0.001 (0.230)	.99	-0.056 (0.252)	.82
≥€2,000	0.125 (0.252)	.62	0.093 (0.276)	.74
Back problems	-0.024 (0.202)	.91	-0.157 (0.221)	.48
Allergies	-0.025 (0.196)	.90	-0.118 (0.214)	.58
High blood pressure	0.542 (0.232)	.02	0.570 (0.254)	.03
Sleeplessness	0.092 (0.244)	.71	0.367 (0.267)	.17
Depression	0.145 (0.250)	.56	0.207 (0.274)	.45
Obesity	-0.073 (0.275)	.79	0.072 (0.301)	.81
Smoking	-0.289 (0.172)	.09	-0.385 (0.189)	.04
Do any sports	-0.081 (0.200)	.69	-0.002 (0.219)	.99
Number of participants	1,959		1,959	

Standard errors in parentheses.

<https://doi.org/10.1371/journal.pone.0234663.t006>

search query data on the user level, for example from search engine providers, is difficult [40] and data, although large in size, do not come with additional information about users' socio-demographics and health. Obtaining access to data like those used in our study, however, is relatively easy through commercial vendors. While our data and our findings are specific to Germany, similar data are available from providers in many other countries (for example, in the U.S., U.K., Spain and France) which would allow for cross-cultural studies.

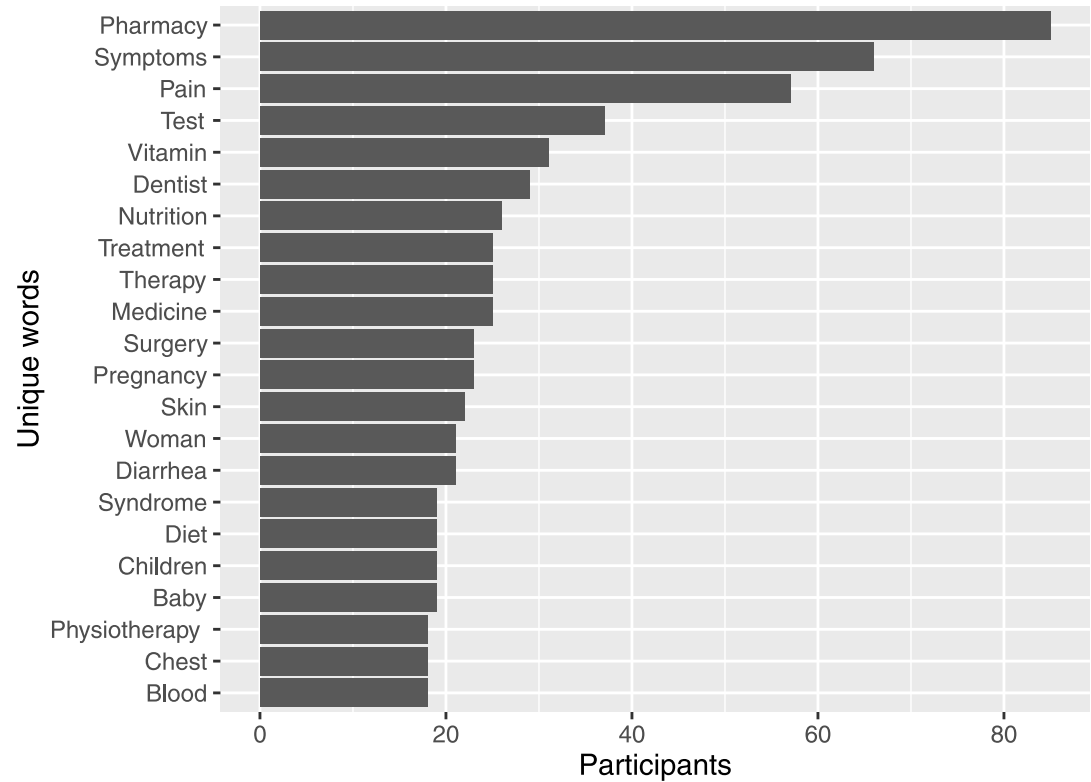


Fig 1.

<https://doi.org/10.1371/journal.pone.0234663.g001>

Socio-demographic correlates of internet and app use

Our results on usage patterns across societal groups confirm findings from previous work. First, women are more likely to browse and search for health content online and spend more of their total online activities looking for health information than men. Moreover, women are more likely to be frequent app users. One explanation for these findings may be that women are more often concerned with child rearing than men [52]. That is, women do not only seek information for themselves, but also for their children. Therefore, they spend more time as they have to seek information for more people and people with different needs. Another

Table 7. Most popular bigram words, across users.

<i>Bigram words</i>	<i>Number of participants</i>
Online pharmacy	38
Ulcerative colitis	8
Crohn's disease	6
Multiple sclerosis	6
Dry skin	6
Vitamin B12	6
Cataract	5
Ankylosing spondylitis	5
Rheumatoid arthritis	5

Translation by authors. Some German bigrams translate to a different number of words in English.

<https://doi.org/10.1371/journal.pone.0234663.t007>

explanation holds that women are more reactive to deviations from health [17]. Likewise, men tend to wait longer before seeking professional help with their health [53]. Regarding the finding that women are more likely to be frequent health app users, it seems that the popularity of apps for women's health (such as period trackers, ovulation diaries, and apps concerned with pregnancy) explains this finding. That is, although we found that frequent health app use is rather limited, women seem to profit more than men from technological innovations in the mobile health sector.

Second, we found that younger people and users with a university education are more likely to use health apps and search for health information online. We believe that the effect is driven by better technology literacy among younger and people with a higher education. However, as the likelihood of developing (multiple) chronic health conditions increases with age, older people may actually profit more from using health apps and searching for health information on the internet [54]. Unfortunately, these results do not indicate that internet use helps reducing shortcomings in health knowledge for societal groups with low education [17].

Third, our results show that smoking significantly decreases both the likelihood and the intensity of using the internet and apps for health-related purposes. One explanation for this finding may be that smokers are also more likely to be of lower socio-economic status (that is, lower education and lower income, for example) and more often male [55, 56]. Thus, smokers may be less likely to engage in health-related activities due to the correlation of smoking with other determinants of health-related online activities. However, it is also possible that people who smoke are in general more risk tolerant and thus less concerned with their health [57].

Contents of internet and app use

The analysis of subcategories of health-related domain reveals that exercising, weight loss and nutrition are among the most popular topics. That is, users seem especially interested in obtaining information about how to live a healthy life. Unsurprisingly, the same categories in addition to health tracking and women's health are also popular among health apps. These results seem to speak to increasing desires for the 'quantified-self', self-optimization through tracking and the analysis of one's body using health trackers, apps and the like [58].

Another popular topic is mental health. This finding is important as mental health is often associated with stigma and mental health consumers often feel discriminated [59]. Web logs provide new ways of measuring and studying mental health that may be less affected by self-reporting bias (though we note that our measure of depression and sleeplessness rely on self-reported survey data). That is, observing digital traces may offer new means for understanding who may be in need of support and how people may be reached. Moreover, understanding how users gather information may help inform and guide the design of targeted interventions to support those seeking professional help [60]. The popularity of the mental health category also adds to the debate about deficiencies regarding mental health support (such as insufficient availability of psychotherapy) in Germany [61].

Furthermore, we found that alternative medicine is a popular topic among users. This finding seems to confirm reports documenting that Germans are, more than residents from other European countries, particularly susceptible to home remedies [62]. Moreover, it may also support notions of decreasing trust in evidence-based healthcare and medical experts as expressed through anti-vaccination movements, for example [18, 63]. Against this background, studying how, where and why users turn to alternatives to evidence-based medicine and what content they consume may help develop campaigns aiming to fight the dissemination of inaccurate and potentially dangerous information on the web. Moreover, lay-people often do not have the necessary skills to evaluate the quality of medical information found

online. While content may cover professionally or peer-reviewed information, it may also include lay-people's opinions and anecdotes that might potentially harm users [19, 64]. As mentioned earlier in this paper, the rise of the anti-vaccination movement and similar phenomena is facilitated through information spread through the internet [18]. To study who is susceptible to medical misinformation, future work should, for example, examine the content of medical information consumed by users on the internet. Using detailed records of users' online activities in combination with scraping and analyzing contents of health websites allows researchers to answer questions that were difficult to study before.

Regarding the analyses of health-related search queries, we note that users often use search engines to seek information about health professionals (such as dentists and physiotherapy) or health institutions (for example, pharmacies). That is, reasons for searching seem primarily functional and once again underline the important role of the internet in facilitating everyday life through the ubiquitous offer of information. In addition, we found treatment-related terms to be among the most popular (such as treatment, therapy and surgery). This finding is somewhat confirmed by the analysis of bigrams, which shows that users often seek information regarding pharmacies, but also about specific diseases. That is, users seem especially interested in obtaining information about specific health issues, but likely also information on treatment options and remedies for conditions they experience. Just as the popularity of health apps, the popularity of treatment-related terms demonstrates users' growing demand for understanding their body and for pro-actively surveying and managing their own health.

Limitations

Our analyses of search queries in this paper could only scratch the surface of what we believe is possible with more advanced natural language processing algorithms. However, we also note that the automated extraction, categorization and understanding of health-related search query data is challenging, especially when it comes to inferring user intentions. For example, a person searching for "ulcerative colitis" may seek information about the disease because she has never heard of it, she may worry that she has the disease due to specific symptoms she experiences or she may have the disease and seek, for example, information on treatments. Understanding which intentions users have when searching for health-related information may, for example, require studying whole browsing sessions. In addition, studying developments of anti-vaccination positions, for example, will likely require the observation of even longer time periods. Moreover, analyzing specific health topics may also require observing larger samples as the number of observations will rapidly decrease if those topics are not widespread in the population.

Future work

Besides substantive interest in health research, we believe that studying health-related internet and app will prove important from a privacy perspective. Privacy research documents that website tracking and sharing of sensitive personal information, which includes health information, is a widespread phenomenon in the online and app world [65–69]. For example, more than 76% of popular web pages that offer information and support regarding mental health contain third-party tracking elements (such as cookies) for marketing purposes [66]. Thus, sensitive personal information on users' mental health may be observable for third parties. Yet, information about such tracking practices provided to users of web pages or apps often does not meet the privacy regulations required under the European Union's General Data Protection Regulation or under the ePrivacy law. Some web pages and apps even fall short collecting users' informed consent at all when sharing personal sensitive information with third parties

[65–67]. In addition, users often do not understand what information is collected about them when browsing websites or using health apps. Even if privacy policies are provided, users rarely read them [70]. It is therefore crucial to understand whether third parties may be able to infer users sensitive information like their health status by observing, for example, her online activities through cookies and tracking in apps and on mobile devices [69].

In light of the current COVID-19 pandemic, it seems more important than ever to study where and how people search for health information, what kinds of websites they visit or apps they use, and critically examine the quality of information users find given that plenty of (mis) information is disseminated online. Passively tracked browser and app use data, such as the data used in the present study, might be a promising way to shed more light into this.

Acknowledgments

The authors thank SINUS-Institut in Heidelberg, Germany and respondi AG in Cologne, Germany, for providing access to the data and Mariel M. Leonard for helpful comments on an earlier version of this paper.

Author Contributions

Conceptualization: Ruben L. Bach, Alexander Wenz.

Data curation: Ruben L. Bach, Alexander Wenz.

Formal analysis: Ruben L. Bach, Alexander Wenz.

Funding acquisition: Ruben L. Bach, Alexander Wenz.

Investigation: Ruben L. Bach, Alexander Wenz.

Methodology: Ruben L. Bach, Alexander Wenz.

Project administration: Ruben L. Bach, Alexander Wenz.

Resources: Ruben L. Bach, Alexander Wenz.

Software: Ruben L. Bach, Alexander Wenz.

Supervision: Ruben L. Bach, Alexander Wenz.

Validation: Ruben L. Bach, Alexander Wenz.

Visualization: Ruben L. Bach, Alexander Wenz.

Writing – original draft: Ruben L. Bach.

Writing – review & editing: Ruben L. Bach, Alexander Wenz.

References

1. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. *International journal of medical informatics*. 2006; 75(1):8–28. <https://doi.org/10.1016/j.ijmedinf.2005.07.032>
2. Pang PCI, Chang S, Pearce JM, Verspoor K. Online Health Information seeking Behaviour: Understanding Different Search Approaches. In: *Proceedings of the 2014 PACIS*. Chengdu, China: AIS; 2014. p. 229.
3. Pang PCI, Verspoor K, Chang S, Pearce J. Conceptualising health information seeking behaviours and exploratory search: result of a qualitative study. *Health and Technology*. 2015; 5(1):45–55. <https://doi.org/10.1007/s12553-015-0096-0>
4. Fox S, Duggan M. Health Online 2013; 2013. Available from https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf

5. Accenture. 2018 Consumer Survey on Digital Health; 2018. Available from https://www.accenture.com/t20180306T103559Z_w_/us-en/_acnmedia/PDF-71/accenture-health-2018-consumer-survey-digital-health.pdf
6. Bitkom Research. 2019 Digital Health; 2019. Available from https://www.bitkom.org/sites/default/files/2019-05/190508_bitkom-pressekonzferenz_e-health_prasentation.pdf
7. Fox S, Duggan M. Mobile Health 2012; 2012. Available from https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2012/PIP_MobileHealth2012_FINAL.pdf
8. Statista. Digital Health, eHealth, mHealth & Hospitals in the U.S. 2017; 2017. Available from <https://www.statista.com/study/44892/book-of-tables-for-the-statista-survey-on-digital-health-ehealth-mhealth-and-hospitals-2017/>
9. Aitken M, Lyle J. Patient Adoption of mHealth. Report by the IMS Institute for Healthcare Informatics. Parsippany, NJ: IMS Institute for Healthcare Informatics; 2015.
10. Li, I, Dey, AK, Forlizzi, J. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In: Proceedings of the 13th international conference on Ubiquitous computing. New York: ACM; 2011. p. 405–414.
11. Klasnja, P, Consolvo, S, McDonald, DW, Landay, JA, Pratt, W. Using mobile and personal sensing technologies to support health behavior change in everyday life: lessons learned. In: AMIA Annual Symposium Proceedings. Bethesda, MD: AMIA; 2009. p. 338–342.
12. Fritz, T, Huang, EM, Murphy, GC, Zimmermann, T. Persuasive Technology in the Real World: A Study of Long-Term Use of Activity Sensing Devices for Fitness. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems. CHI'14. New York, NY, USA: ACM; 2014. p. 487–496.
13. White RW, Horvitz E. Cyberchondria: Studies of the Escalation of Medical Concerns in Web Search. *ACM Trans Inf Syst.* 2009; 27(4):23:1–23:37.
14. Ashrafian H, Toma T, Harling L, Kerr K, Athanasiou T, Darzi A. Social networking strategies that aim to reduce obesity have achieved significant although modest results. *Health affairs.* 2014; 33(9):1641–1647. <https://doi.org/10.1377/hlthaff.2014.0370>
15. Wohlers EM, Sirard JR, Barden CM, Moon JK. Smart phones are useful for food intake and physical activity surveys. In: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Piscataway, NJ, US: IEEE; 2009. p. 5183–5186.
16. Conroy DE, Yang CH, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. *American journal of preventive medicine.* 2014; 46(6):649–652. <https://doi.org/10.1016/j.amepre.2014.01.010>
17. Manierre MJ. Gaps in knowledge: tracking and explaining gender differences in health information seeking. *Social Science & Medicine.* 2015; 128:151–158. <https://doi.org/10.1016/j.socscimed.2015.01.028>
18. Kata A. A postmodern Pandora's box: Anti-vaccination misinformation on the Internet. *Vaccine.* 2010; 28(7):1709–1716. <https://doi.org/10.1016/j.vaccine.2009.12.022>
19. Fox S. Online Health Search 2006; 2006. Available from https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/2006/PIP_Online_Health_2006.pdf.pdf
20. Lorence D, Park H. Study of education disparities and health information seeking behavior. *Cyberpsychology & behavior.* 2007; 10(1):149–151. <https://doi.org/10.1089/cpb.2006.9977>
21. Beaudoin CE, Hong T. Health information seeking, diet and physical activity: an empirical assessment by medium and critical demographics. *International journal of medical informatics.* 2011; 80(8):586–595. <https://doi.org/10.1016/j.ijmedinf.2011.04.003>
22. Atkinson N, Saperstein S, Pleis J. Using the internet for health-related activities: findings from a national probability sample. *Journal of medical Internet research.* 2009; 11(1):e5.
23. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: An analysis of health information seeking behavior among US adults. *Cogent Social Sciences.* 2017; 3(1):1302785. <http://doi.org/10.1080/23311886.2017.1302785>
24. Tennant B, Stelfson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. *Journal of medical Internet research.* 2015; 17(3):e70. <http://doi.org/10.2196/jmir.3992> PMID: 25783036
25. Massey PM. Where do US adults who do not use the internet get health information? Examining digital health information disparities from 2008 to 2013. *Journal of health communication.* 2016; 21(1):118–124. <http://doi.org/10.1080/10810730.2015.1058444> PMID: 26166484
26. Houston TK, Allison JJ. Users of Internet health information: differences by health status. *Journal of medical Internet research.* 2002; 4(2):e7. <http://doi.org/10.2196/jmir.4.2.e7> PMID: 12554554

27. Bansil P, Keenan NL, Zlot AI, Gilliland JC. Health-related information on the web: results from the HealthStyles Survey, 2002–2003. *Preventing chronic disease*. 2006; 3(2):1–10.
28. McCully SN, Don BP, Updegraff JA. Using the Internet to Help With Diet, Weight, and Physical Activity: Results From the Health Information National Trends Survey (HINTS). *Journal of medical Internet research*. 2013; 15(8):e148. <https://doi.org/10.2196/jmir.2612>
29. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. *Journal of medical Internet research*. 2017; 19(4):e125. <https://doi.org/10.2196/jmir.5604>
30. Ernsting C, Stühmann LM, Dombrowski SU, Voigt-Antons JN, Kuhlmeier A, Gellert P. Associations of Health App Use and Perceived Effectiveness in People With Cardiovascular Diseases and Diabetes: Population-Based Survey. *JMIR mHealth and uHealth*. 2019; 7(3):e12179. <https://doi.org/10.2196/12179>
31. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. *JMIR mHealth and uHealth*. 2015; 3(4):e101. <https://doi.org/10.2196/mhealth.4924>
32. Xie Z, Nacioglu A, Or C. Prevalence, demographic correlates, and perceived impacts of mobile health app use amongst Chinese adults: cross-sectional survey study. *JMIR mHealth and uHealth*. 2018; 6(4):e103. <https://doi.org/10.2196/mhealth.9002>
33. Shen C, Wang MP, Chu JT, Wan A, Viswanath K, Chan SSC, et al. Health app possession among smartphone or tablet owners in Hong Kong: population-based survey. *JMIR mHealth and uHealth*. 2017; 5(6):e77. <https://doi.org/10.2196/mhealth.7628> PMID: 28583905
34. ORCHA Organisation for the Review of Care and Health Apps. Available from <https://www.orcha.co.uk/>
35. Araujo T, Wonneberger A, Neijens P, de Vreese C. How Much Time Do You Spend Online? Understanding and Improving the Accuracy of Self-Reported Measures of Internet Use. *Communication Methods and Measures*. 2017; 11(3):173–190. <https://doi.org/10.1080/19312458.2017.1317337>
36. Cartright MA, White RW, Horvitz E. Intentions and Attention in Exploratory Health Search. In: *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR'11. New York, NY, USA: ACM; 2011. p. 65–74.
37. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature*. 2009; 457(7232):1012. <https://doi.org/10.1038/nature07634>
38. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in big data analysis. *Science*. 2014; 343(6176):1203–1205. <https://doi.org/10.1126/science.1248506>
39. Abebe R, Hill S, Vaughan JW, Small PM, Schwartz HA. Using search queries to understand health information needs in africa. In: *Proceedings of the International AAAI Conference on Web and Social Media*. vol. 13. Palo Alto, CA, USA: AAAI; 2019. p. 3–14.
40. Fournay A, White RW, Horvitz E. Exploring Time-Dependent Concerns About Pregnancy and Childbirth from Search Logs. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI'15. New York, NY, USA: ACM; 2015. p. 737–746.
41. Cooper CP, Mallon KP, Leadbetter S, Pollack LA, Peipins LA. Cancer Internet search activity on a major search engine, United States 2001–2003. *Journal of medical Internet research*. 2005; 7(3):e36. <https://doi.org/10.2196/jmir.7.3.e36>
42. McHugh SM, Corrigan M, Morney N, Sheikh A, Lehane E, Hill AD. A quantitative assessment of changing trends in internet usage for cancer information. *World journal of surgery*. 2011; 35(2):253–257. <https://doi.org/10.1007/s00268-010-0830-8>
43. Foroughi F, Lam AK, Lim MS, Saremi N, Ahmadvand A. “Googling” for cancer: an infodemiological assessment of online search interests in Australia, Canada, New Zealand, the United Kingdom, and the United States. *JMIR cancer*. 2016; 2(1):e5. <https://doi.org/10.2196/cancer.5212>
44. Fazeli SD, Carlos RC, Hall KS, Dalton VK, et al. Novel data sources for women’s health research: mapping breast screening online information seeking through Google trends. *Academic radiology*. 2014; 21(9):1172–1176.
45. Phillips CA, Leahy AB, Li Y, Schapira MM, Bailey LC, Merchant RM. Relationship between state-level Google online search volume and cancer incidence in the United States: retrospective study. *Journal of medical Internet research*. 2018; 20(1):e6. <https://doi.org/10.2196/jmir.8870>
46. Marcu A, Muller C, Ream E, Whitaker KL. Online Information-Seeking About Potential Breast Cancer Symptoms: Capturing Online Behavior With an Internet Browsing Tracking Tool. *Journal of medical Internet research*. 2019; 21(2):e12400. <https://doi.org/10.2196/12400>
47. Respondi AG. Available from <https://www.respondi.com/>
48. SINUS Markt- und Sozialforschung GmbH. Available from <https://www.sinus-institut.de/>

49. Wakoopa. Available from <https://www.wakoopa.com/>
50. Webshrinker. Available from <https://www.webshrinker.com/>
51. Domain taxonomy. Interactive Advertising Bureau. Available from <https://support.aerserv.com/hc/en-us/articles/207148516-List-of-IAB-Categories>
52. Craig L. Does father care mean fathers share? A comparison of how mothers and fathers in intact families spend time with children. *Gender & society*. 2006; 20(2):259–281. <https://doi.org/10.1177/0891243205285212>
53. Addis ME, Mahalik JR. Men, masculinity, and the contexts of help seeking. *American psychologist*. 2003; 58(1):5. <https://doi.org/10.1037/0003-066X.58.1.5>
54. Wildenbos GA, Peute L, Jaspers M. Aging barriers influencing mobile health usability for older adults: A literature based framework (MOLD-US). *International journal of medical informatics*. 2018; 114:66–75. <https://doi.org/10.1016/j.ijmedinf.2018.03.012>
55. van Loon AJM, Tjihuis M, Surtees PG, Ormel J. Determinants of smoking status: cross-sectional data on smoking initiation and cessation. *European Journal of Public Health*. 2005; 15(3):256–261. <https://doi.org/10.1093/eurpub/cki077>
56. Bottorff JL, Haines-Saah R, Kelly MT, Olliffe JL, Torchalla I, Poole N, et al. Gender, smoking and tobacco reduction and cessation: a scoping review. *International journal for equity in health*. 2014; 13(1):114. <https://doi.org/10.1186/s12939-014-0114-2> PMID: 25495141
57. Jusot F, Khlal M. The role of time and risk preferences in smoking inequalities: a population-based study. *Addictive behaviors*. 2013; 38(5):2167–2173. <https://doi.org/10.1016/j.addbeh.2012.12.011>
58. Williamson B. Algorithmic skin: health-tracking technologies, personal analytics and the biopedagogies of digitized health and physical education. *Sport, Education and Society*. 2015; 20(1):133–151. <https://doi.org/10.1080/13573322.2014.962494>
59. Wahl OF. Mental Health Consumers' Experience of Stigma. *Schizophrenia Bulletin*. 1999; 25(3):467–478. <https://doi.org/10.1093/oxfordjournals.schbul.a033394>
60. Aguilera A. Digital Technology and Mental Health Interventions: Opportunities and Challenges. *Arbor*. 2015; 191(771):a210. <https://doi.org/10.3989/arbor.2015.771n1012>
61. Bundespsychotherapeutenkammer. Ein Jahr nach der Reform der Psychotherapie-Richtlinien: Wartezeiten 2018; 2018. Available from https://www.bptk.de/wp-content/uploads/2019/01/20180411_bptk_studie_wartezeiten_2018.pdf
62. STADA Arzneimittel AG. STADA Group Gesundheitsreport 2019; 2019. Available from https://www.deinegesundheit.stada/media/1314/stada_gesundheitsreport_2019.pdf
63. Huang ECH, Pu C, Chou YJ, Huang N. Public Trust in Physicians—Health Care Commodification as a Possible Deteriorating Factor: Cross-sectional Analysis of 23 Countries. *Inquiry*. 2018; 55:1–11.
64. Tan SSL, Goonawardene N. Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review. *Journal of medical Internet research*. 2017; 19(1):e9. <https://doi.org/10.2196/jmir.5729>
65. Huckvale K, Prieto JT, Tilney M, Benghozi PJ, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. *BMC medicine*. 2015; 13(1):214. <https://doi.org/10.1186/s12916-015-0444-y>
66. Privacy International. Your mental health for sale: How websites about depression share data with advertisers and leak depression test results; 2019a. Available from <https://www.privacyinternational.org/sites/default/files/2019-09/Yourmentalhealthforsale—PrivacyInternational.pdf>
67. Privacy International. No Body's Business But Mine: How Menstruation Apps Are Sharing Your Data; 2019b. Available from <https://privacyinternational.org/long-read/3196/no-bodys-business-mine-how-menstruation-apps-are-sharing-your-data>
68. Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and privacy analysis of mobile health applications: the alarming state of practice. *IEEE Access*. 2018; 6:9390–9403. <https://doi.org/10.1109/ACCESS.2018.2799522>
69. Bach RL, Kern C, Amaya A, Keusch F, Kreuter F, Heinemann J, et al. Predicting voting behavior using digital trace data. *Social Science Computer Review*. 2019; Online first.
70. Dolin C, Weinshel B, Shan S, Hahn CM, Choi E, Mazurek ML, et al. Unpacking Perceptions of Data-Driven Inferences Underlying Online Targeting and Personalization. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI'18. New York, NY, USA: ACM; 2018. p. 493.