Shadow Health-Related Data: Definition, Categorization, and User Perspectives

Yamane El Zein*
CSEM SA
Neuchâtel, Switzerland
yamane.el-zein@csem.ch

Kavous Salehzadeh Niksirat[†]
University of Lausanne
Lausanne, Switzerland
kavous.salehzadehniksirat@unil.ch

Noé Zufferey
ETH Zurich
Zurich, Switzerland
zuffereyn@ethz.ch

Mathias Humbert
University of Lausanne
Lausanne, Switzerland
mathias.humbert@unil.ch

Kévin Huguenin
University of Lausanne
Lausanne, Switzerland
kevin.huguenin@unil.ch

Abstract

Health-related data (HRD) about individuals are increasingly generated and processed. The sources and volume of such data have grown larger over the past years, they include wearable devices, health-related mobile apps, and electronic health records. HRD are sensitive, have important privacy implications, hence hold a special status under existing privacy laws and regulations. In this work, we focus on *shadow* HRD: these HRD are generated and/or processed by individuals by using general-purpose digital tools outside of a professional healthcare information system. Some examples are health-related queries made by individuals on general-purpose search engines and LLM-based chatbots, or medical appointments and contact information of health professionals synced to the cloud. Such data, and the privacy risks stemming from them, are often overlooked when studying digital health. Using information from two focus group sessions (23 participants in total), we identified and categorized a broad variety of user behaviors that, including the aforementioned examples, lead to the creation of *shadow* HRD. Then, informed by this categorization, we designed a questionnaire and deployed it through an online survey (300 respondents) to assess the prevalence of such behaviors among the general public, as well as user awareness of (and concerns about) the privacy risks stemming from their *shadow* HRD. Our findings show that most respondents adopt numerous and diverse behaviors that create *shadow* HRD, and that very few resort to mechanisms to protect their privacy.

CCS Concepts

• **Security and privacy → Human and societal aspects of security and privacy**; • **Applied computing → Health informatics**; • **Human-centered computing → Empirical studies in HCI**.

---

*Also with University of Lausanne.
[†]Also with EPFL.

Keywords

health-related data, privacy, user study

1 Introduction

Recent technologies, such as fitness trackers, health mobile applications, and AI-based tools, have become so popular that health-related data (HRD), previously restrained to the professional healthcare setting, are now ubiquitously processed by a plethora of digital services. Such technologies are useful to end-users for self-managing their health. However, this phenomenon complicates the process of evaluating and mitigating the privacy risks associated with HRD. As HRD provide information about individuals' physical and mental health conditions, they may be used for discriminatory purposes by health insurers or employers, and could even lead to prosecution (particularly in the post-Roe era in the US [14]). As such, HRD are considered sensitive, have important privacy implications, and hold a special status under existing data protection laws (e.g., HIPAA and GDPR). However, HRD created or processed by digital tools primarily designed for health purposes might very well be just the tip of the iceberg. The fact that HRD can also be created and/or processed through general-purpose tools (i.e., not intended for health-related purposes) is often overlooked. In this work, we provide the following formal definition of such HRD in Definition 1, and henceforth refer to them as *shadow* HRD.[1] Our definition of *shadow* HRD draws a parallel to that of shadow IT, defined by Haag and Eckhardt [27] as "hardware, software, or services built, introduced, and/or used for a job, without explicit approval or even knowledge of the organization."

This definition builds on that of HRD, which we derive from the definition of 'data concerning health' in the GDPR (Article 4 (15) and Recital 35) [52], as *"any personal data related to the physical or*

---

[1]The similar term "shadow health records" has been used in [55] to discuss legal aspects surrounding some *shadow* HRD.

DEFINITION 1. **Shadow *health-related data (*Shadow *HRD)** *are health-related data that are generated and/or processed by individuals by using general-purpose digital tools outside of a professional healthcare information system (i.e., digital systems designed to manage healthcare delivery within healthcare ecosystems). Typical examples include health-related queries made by individuals on general-purpose search engines and on LLM-based chatbots, photos of skin conditions, medical appointments, and contact information of health professionals synced to the cloud.*

mental health of a natural person, which reveal information about his or her past, present or future health status."

As *shadow* HRD can be created through a variety of behaviors, they are not as evident to identify as *non-shadow* HRD, which are created within well-defined healthcare information systems or tools explicitly intended for health. Therefore, their privacy implications are often underestimated. From a legal perspective, the enforcement of additional provisions specific to standard health-related data (which belongs to the category of sensitive data) might become more challenging, as the effort for service providers to identify which of the data it processes are health-related, hence sensitive, could be considered disproportionate. While existing types and sources of HRD have been categorized in prior work [3, 25, 45], *shadow* HRD have often been either excluded of such analyses, or only mentioned briefly. Further, many studies have explored users' behaviors and perceptions with respect to the use of technologies that are primarily intended for health purposes and may compromise the privacy of HRD [2, 21], yet none have touched on the general-purpose technologies that could lead to a similar outcome. In this work, we bridge the gap and answer the following questions: **(RQ1)** What are the user behaviors that lead to the creation of HRD, particularly *shadow* HRD? **(RQ2)** How prevalent are these behaviors, and what are the levels of awareness and concern of technology users with respect to the associated privacy risks? Do they take any privacy-preserving measures to mitigate them?

The overall methodology is summarized in Figure 1. To answer RQ1, we hold two focus group sessions (23 participants in total), the first with researchers and practitioners from the computer science, information systems, health IT, medicine, and law fields, and the second with university students from various backgrounds, through which we collect different user behaviors that lead to the creation of *shadow* HRD, and thus obtain an inventory of such behaviors. We then categorize and organize them into a classification of (*shadow*) HRD. Such a classification is useful to highlight the different ways in which pieces of *shadow* HRD are created, which would then enable identifying the associated privacy risks, and in turn, finding countermeasures to mitigate them. Then, to answer RQ2, we design a questionnaire informed by this classification and deploy it through a large-scale user online survey (300 respondents). Through this questionnaire, we evaluate the prevalence of previously identified behaviors that lead to the creation of *shadow* HRD, users' awareness of and concerns with regards to *shadow* HRD, as well as their use (or lack thereof) of measures to protect such data.

Our main findings show that such behaviors are highly prevalent among technology users, who typically do not take any particular protection measures to mitigate the risks associated with their *shadow* HRD. For example, 69% of respondents reported saving the contact of their physician in their digital address book, while 51%

reported including the relevant medical specialty, among other details. Further, our results show that some users engage in behaviors that may create *shadow* HRD about other people (e.g., children). These findings, as well as the classification of *shadow* HRD that we present in this work, can be used to raise awareness among users about the privacy risks associated with *shadow* HRD, inform policymakers about their ubiquity and the importance of enforcing appropriate laws pertaining to them, and encourage privacy researchers to include *shadow* HRD when considering health privacy.

2 Related Work

In this section, we discuss prior work that systematizes and categorizes health-related data and technologies, as well as prior work that, with respect to HRD, assesses behaviors, awareness, concerns, and perceptions of individuals.

2.1 Classifications of Data Sources

Classifications and taxonomies organize and group elements of a given field by finding common patterns and relationships among and between them, presenting them clearly hence enabling the identification of problems and definition of solutions within the field at hand [46, 47]. Multiple classifications and taxonomies of HRD, as well as technologies that lead to their creation, have been proposed. Many of them focus on specific contexts or types of health-related services. For instance, Bashshur et al. [8] propose a taxonomy of the telemedicine domain, with reference to information systems made to facilitate communication among different stakeholders within the healthcare setting, whereas Khaled [32] proposes a taxonomy of the Internet of Medical Things (IoMT), the subfield of IoT targeted for healthcare, describing types of IoMT nodes, the main purposes for their use, and the data that they generate. A taxonomy of pervasive healthcare systems is proposed by Muras et al. [45]. It depicts different properties of such technologies, such as purpose, sources of data, and technology types. Similarly, Alrige and Chatterjee [3] present a taxonomy of wearables used in healthcare. Other classifications and taxonomies focus specifically on mobile health applications [25, 75] by highlighting different dimensions of these applications such as their modes of data capture, and purposes of use.

Whereas the above-mentioned works focus on specific health-related technologies and the creation of HRD in specific health-related contexts, other classifications pertain to broader domains such as mHealth [10], eHealth [9, 48, 50] or digital health [19]. Although broad, these classifications are constrained either to the professional healthcare ecosystem, or to digital services that are primarily designed for health management purposes (i.e., mobile health apps and wearable devices). To the best of our knowledge,
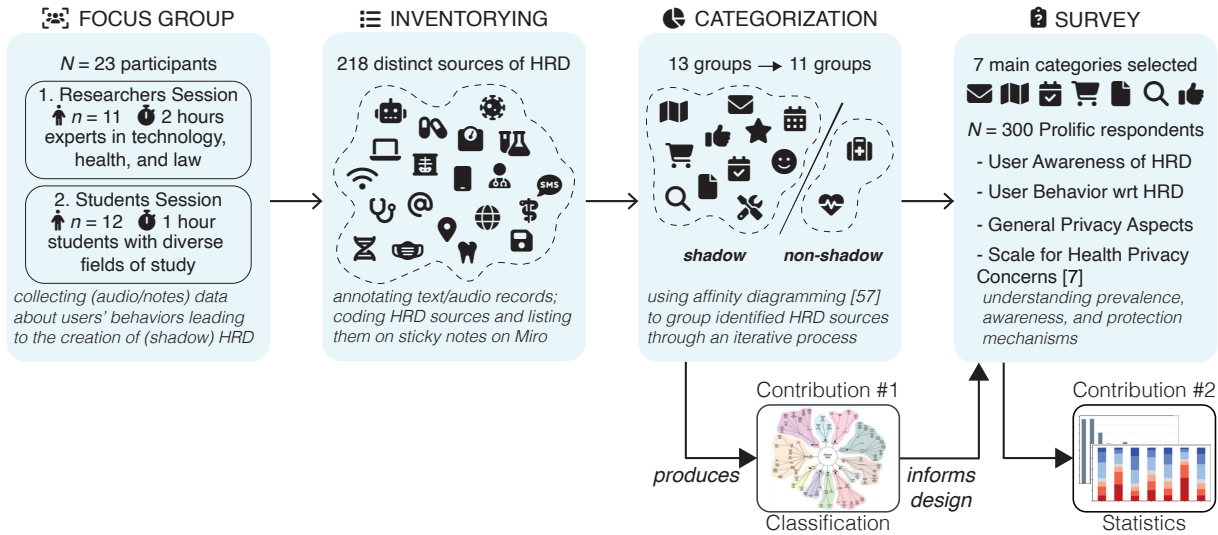
Figure 1: Procedure of the study including (1) two focus group sessions, (2) generation of an inventory of *shadow* HRD, (3) categorization and a classification of *shadow* HRD, and (4) a survey questionnaire insights about behaviors, awareness and concern as outcome.

no prior work has systematically identified and categorized sources of *shadow* HRD as defined in Section 1. In this work, we bridge this gap by proposing a classification of *shadow* HRD sources, focusing on HRD created through the use of general-purpose tools rather than ones primarily designed for health purposes. Moreover, our approach for building the classification of *shadow* HRD differs from the above-cited works from a methodological perspective. While authors of these works develop their classifications mainly through review of the literature, we develop ours through focus groups. This choice is motivated by the fact that our classification encompasses *user behaviors* (that lead to the creation of *shadow* HRD), which we deemed relevant to explore through participation of said users, rather than from the literature.

2.2 User Behaviors, Awareness, and Concerns

A large number of user studies have investigated, through surveys, interviews, or focus groups, the behaviors, awareness and concerns of individuals regarding HRD and technologies that they stem from. Within the clinical setting, several studies have explored patient attitudes, perceptions and awareness concerning electronic health records (EHR) [13, 23, 34, 61, 73]. These studies report that respondents exhibit high trust in the healthcare system with respect to the privacy of their EHR, while still voicing concerns about the sensitivity of the information they contain. Other studies focus on user attitudes towards health information exchange (HIE) [4, 18, 24, 33, 40, 60], assessing patients' willingness to share data from their EHR for scientific research. Outside the clinical setting, a number of studies have evaluated users' behaviors, awareness, and concerns toward the use of off-the-shelf wearable devices (e.g., Fitbit) [1, 2, 21, 36, 43, 66, 69, 77]. Findings from such studies are summarized in a recent literature survey [56], reporting that wearable device users exhibit low knowledge of how data from their wearables is processed, as well as low privacy concerns. Other studies investigated users' perceptions and attitudes toward mobile

phone health apps [16, 41, 42, 53, 58]. Findings from similar studies have been summarized in a literature survey [71], which reports privacy concerns as being one of the recurring critiques by users against mobile health apps in the surveyed studies. Some studies focused on genetic data, exploring individuals' patterns of sharing DNA data, as well as their privacy concerns toward disclosing data to direct-to-consumer genetic testing and analysis platforms, such as 23andMe or GedMatch [5, 6, 22, 44, 57, 65, 70, 72]. HRD sharing is also studied in the context of online health communities (i.e., online social networks revolving around health, connecting patients, health professionals, caregivers, and others who have a shared interest in specific diseases or health conditions), such as Patients-LikeMe [20, 64, 74]. Furthermore, some works study users' concerns about the use of chatbots and have addressed concerns related, among others, to the disclosure of HRD. Specifically, Chametka et al. [15] look at users' perceptions of the security and privacy of mental health chatbots, while others [26, 76] explore users' concerns toward general-purpose chatbots, and investigate, among other concerns, the ones relating to health data being shared with the chatbots. Some studies also explore, with regards to HRD sharing, individuals' privacy concerns, preferences, and awareness of privacy risks, with no particular focus on specific technologies or data types. Trinidad et al. [67] explore individuals' willingness to share health data with commercial third parties, distinguishing between healthcare and business third parties; they found that patients were more willing to share their HRD with the former rather than the latter. Ostherr et al. [49] assessed users' trust and privacy perceptions in regard to health data generated outside of the clinical setting, through several types of personal devices and software designed for tracking personal health. Finally, Bansal et al. [7] proposed a standard scale for evaluating individuals' perceptions of online health information sensitivity, as well as their privacy concerns. We use a subset of their survey questions for our own study (Section 4). Whereas a large amount of research assesses behaviors, awareness, and concern of individuals with respect to HRD, to the

best of our knowledge, no other studies have assessed them with respect to *shadow* HRD.

3 Focus Group and Classification

To answer RQ1, we first identify, through focus groups, different behaviors that lead to the creation of *shadow* HRD. Second, we systematically categorize the resulting behaviors into a structured classification. The methodologies of both are depicted in Figure 1 and explained below.

3.1 Focus Group - Methodology

We conducted two different focus group sessions, the purpose of which was to create a comprehensive inventory of different sources of HRD, focusing mainly on sources of *shadow* HRD rather than on *non-shadow* HRD (i.e., stemming from the mainstream healthcare context). Our goal was to gather different *behaviors* of individuals, in order to create an inventory of *shadow* HRD. Hence, we used focus groups [35, 38] as they enabled us to gather insights from a diverse group of people whose behaviors could differ. They were able to have discussions, through which they thought of additional sources of HRD (unlike when deploying a survey or conducting interviews). To gain insights from lay technology users, we conducted one focus group session with researchers whose technical and health-oriented expertise enables them to identify a wide range of HRD sources and a second session with students from diverse fields of study, who reflected on their daily activities and perspectives. The first focus group session with $n = 11$ participants (i.e., henceforth *researchers session*) involved the research team, in addition to other researchers from different disciplines. The second focus group session with $n = 12$ participants (i.e., henceforth *student session*) involved students enrolled at either the researchers' institution or geographically adjacent ones.

The use of focus groups may depend on participants' awareness, access to, and use of certain tools and technologies. While literature reviews often serve as a suitable method for collecting similar data (see Section 2), the information in the literature with regards to health-related data stemming from digital tools that are not designed for health is limited and sparse. As such, the use of focus group sessions, including participants from various demographic groups and professional backgrounds, results in a richer inventory of *shadow* HRD. However, we also reviewed the literature prior to holding the focus group sessions, and indirectly included *shadow* HRD sources mentioned in the literature through our participation in the first focus group session (see Section 3.1.1, Researchers Session). Before conducting the focus group studies, we obtained ethical approval from the Institutional Review Board (IRB) of our university.

3.1.1 Recruitment and Participants.

Researchers Session. The first focus group session involved 11 researchers, including five from the research team (one Ph.D student, two post-doctoral researchers, two professors). The positions of the research team members during this session were as follows: two led the session, but all five contributed to the inventory of sources of *shadow* HRD, similarly to all other participants. Some of the *shadow* HRD sources contributed by the authors were encountered

by them while reviewing related work. The research interests of the research team members are privacy and security, particularly in relation to health and wearables data, as well as human-computer interaction, particularly with regard to user perspectives on privacy. Three were women, and eight were men. The researchers are affiliated with various institutions and departments and hold different academic and professional positions: four Ph.D. students, two post-doctoral researchers, one senior data scientist, three professors, and one physician. Their fields of research are computer science, data science, information systems, health science, medicine, and law. Nine researchers were present in person, and two joined remotely. Their participation was voluntary. They were not compensated financially. Prior to the session, participants were briefed about its main purpose but were not asked to prepare anything. Their consent for video and audio recording of the session was sought via e-mail, and they were asked to bring their mobile phones with them.

Student Session. The second focus group session involved 12 students recruited through the university's research service. Three members of the research team led the session and provided instructions, however, they did not contribute to the inventory of sources of *shadow* HRD and only attended the session as facilitators. The interested participants filled out a short screener survey to provide their contact details, demographic information, availability, and level of proficiency in English, as well as their consent to participate in the session and to be audio recorded. As we conducted the session in English, we selected participants who reported at least a C1 (Advanced) level. We invited an equal number of women and men, as it is important to capture gender-specific behaviors that create *shadow* HRD. Yet, as registration was on a first-come-first-served basis, we could not further control the gender balance and, unfortunately, our final group was imbalanced in terms of gender: four were women and eight were men. The participants' ages ranged from 18 to 23 ($M = 19.8$, $SD = 1.7$). Participants came from seven different schools within the university and hence had a variety of academic backgrounds. In the invitation to take the screener survey, we included a sentence encouraging people with chronic physical or mental conditions, disorders, and/or disabilities to participate without having to explicitly disclose whether the latter applied to them. We did this to promote a diversity of profiles in the participant group, as such individuals could have distinct behaviors related to the management of their specific condition and a personal interest in participating in a study around HRD. We did not collect sensitive information about them. Participants were remunerated with the equivalent of ~USD 22 in the local currency (for one hour).

3.1.2 Procedure. The purpose of the focus group sessions was to gather ideas leading to the creation of an inventory of HRD, in particular the *shadow* HRD sources, types, and contexts in which they are created. The participants were asked to describe scenarios in which they believe their HRD was created, preferably outside of professional healthcare information systems and favoring *breadth* over *depth* of ideas (i.e., a larger number of HRD sources rather than details about specific ones). As the two sessions followed a similar structure, we primarily describe the procedure for the researchers session and highlight any differences in the student session.

Researchers session. The session lasted two hours. After reminding participants that it was recorded and that we would collect their written notes at the end of it, and having a brief round-table for introductions, the session leader made a presentation defining the term "shadow health-related data," the purpose of the session, the types of ideas we aimed to collect, including some examples (e.g., *"I save the number of my physician in my phone address book, which is synced with my Google Account,"*) and a brief overview of the planned activities and discussions, described below.

Activity 1: We asked participants to open their phones and to note down any sources of *shadow* HRD they could identify, from apps, files, media, built-in features, *etc.* The activity lasted 10 minutes. Each participant worked individually and wrote down the identified sources of HRD on paper. We chose an individual activity because each person's phone has unique apps, files, and settings tailored to their specific needs.

Discussion about Activity 1: Activity 1 was followed by a 40 minutes general discussion, in which each participant was given a few minutes to present the sources of HRD that they noted. Afterwards, we used the remaining time for participants to add any new ideas that came to mind after listening to other participants' contributions.

Activity 2: For the second activity, we asked participants to work in groups of three to four, to think together of any sources and types of *shadow* HRD while favoring HRD sources unrelated to mobile phone and to avoid repeating ideas from Activity 1. To encourage active discussions, we opted for a group activity that could lead to the emergence of new ideas. The activity lasted 20 minutes. Again, each group noted the identified sources of HRD on paper.

Discussion about Activity 2: Activity 2 was followed by a 30 minutes general discussion in which one representative of each group was given five minutes to present the sources of HRD that their group identified. This was followed by an open discussion, allowing for any remaining ideas or thoughts.

The results of the focus group session are (1) audio and video recording of the session, (2) hand-written notes of the participants, and (3) backup notes taken by two of the authors.

Students Sessions. Here, we only highlight the main differences with the researchers session. The students session lasted one hour (rather than two), as we expected students, with less expertise related to the research topic, would have less to share than the researchers. The duration of the activities and discussions were re-allocated as follows: 10 minutes for Activity 1, 20 minutes for the discussion about Activity 1, 10 minutes for Activity 2, and 20 minutes for the discussion about Activity 2. All participants were present in the meeting room (none were online).

3.1.3 Data Analysis. We analyzed the audio recordings of the sessions, as well as the hand-written notes. We provide details of our analysis in subsequent sections.

Coding and Tagging. The first author listened to the audio recordings, starting with the general discussions that followed Activities 1 and 2. Using a digital audio editing software (Audacity), when a participant mentioned a source of HRD, the researcher annotated the corresponding segment with a label spanning it and described the source of HRD as closely as possible to the way it was described by the participant while remaining concise (a.k.a. in vivo coding [37] by retaining only keywords). Afterwards and for completeness, the researcher annotated any sources of HRD that were mentioned during the group discussions of Activity 2 and that were not later conveyed during the general discussion.

The labels, each representing one source of HRD, were added under the form of (digital) sticky notes, in a collaboration platform (Miro). Henceforth, we refer to these labels as "HRD sources." HRD sources from the researchers and student sessions were distinguished by using a different color for each session. Also, a few HRD sources that were included by focus group participants in their written notes, yet not mentioned during the discussion, were extracted and added to the corresponding group of HRD sources. We added HRD sources that the research team thought of, over the duration of the research project, yet were not mentioned during any of the focus group sessions.

For the students session, which took place after the researchers session, we tagged the HRD sources as either *Seen* if already mentioned during the researchers session, *Partially seen* if mentioned with some subtle differences (e.g., *"Blood analysis results sent by e-mail by the lab"* vs. *"Sharing medical reports using e-mail"*), or *Unseen*, if not mentioned at all. Of the HRD sources from student session, 59% were labeled as *Seen*, 25% as *Partially seen*, and 16% as *Unseen*. Given the low proportion of unseen ideas, we did not conduct any additional focus group sessions. For both sessions, we tagged HRD sources as either *Shadow*, *Borderline-shadow* or *Non-shadow*. Both *Shadow* and *Borderline-shadow* HRD sources fall under the category of *shadow* HRD as defined in Section 1, because they both stem from the use of digital tools outside of a professional healthcare information system. However, for this categorization, we tagged HRD that stem from tools primarily designed for health-related purposes (e.g., fitness apps) as *Borderline-shadow*, and those stemming from general-purpose tools as *Shadow*. The rationale behind this distinction is that *Borderline-shadow* HRD are easier to identify as being health-related than *Shadow* ones are. We also use *Structured*, *Semi-structured*, and *Unstructured* tags to depict the level of structure of the identified HRD. We consider HRD that are organized in a fixed format with well-defined fields as *Structured*, those that do not adhere to a fixed format yet contain markers (e.g., keywords, tags) indicating their content as *Semi-structured*, and those that lack formatting and organization as unstructured (e.g., untagged text or images). We also tagged HRD sources for which the HRD relates to individuals other than the one creating them (e.g., *"Child's physician phone number in address book"*) with the *Interdependent privacy* tag [28–31]. Finally, we used the *Far-fetched* tag to indicate HRD sources that are either too futuristic (e.g., *"Card that has your medical information which gives you access to hospital facilities automatically"*) or require a significant effort by an adversary to infer health-related information from (e.g., *"Thermostat of smart homes can indicate health status"*).

Categorization of HRD Sources. First, we merged HRD sources from the researchers session with those from the student session that were tagged as *Unseen* or *Partially seen*, and those continuously identified by the authors during the research project. We discarded HRD sources tagged as *Far-fetched* from further analysis, focusing

on directly health-related data. We used affinity diagramming [59] (a.k.a. the KJ method) to group identified sources of HRD. The goal was to categorize the identified sources of HRD into a classification, summarizing, and organizing different shadow and non-shadow ones. The first iteration involved picking at random an HRD source and making it the first one of its group. Then, every other HRD source was examined and either placed in an existing group, if deemed similar, or used to create a new one. If a source of HRD was deemed similar to multiple groups, it was duplicated and added to all of them. After all HRD sources were added to groups, every formed group was assigned a name. At the end of this step, 23 groups were created. Most of them reflected the type of services associated with the HRD sources in the group (e.g., Communications, Navigation/Maps, etc.). The second and third iteration involved going through each group and creating subgroups of similar HRD sources within it. Some of the initial groups were joined into super-groups, and some were eliminated and their content redistributed. We also ensure that duplicated ideas only appear in one category. When making such decisions, we favor categorizing HRD sources, based on the purpose of the tool they stem from. Figure 1 summarizes the methodology for the classification of sources of *shadow* HRD (Section 3.1) and the user survey (described in Section 4).

3.2 Classification - Results and Discussion

In this section, we present the resulting classification of HRD sources.

3.2.1 Classification of HRD Sources. The resulting classification contains 218 distinct sources of HRD. 61% of them tagged as *Shadow*, 25% as *Borderline-shadow*, and 14% as *Non-shadow*. The classification contains 13 main groups that we subsequently refer to as *categories*. These categories reflect the type (based on purpose) of the digital tools through which HRD can be created. Of the categories, 11 include *Shadow* HRD sources: ☑ Productivity/Organization, ☺ Entertainment, 📅 Reservations, 🛒 Shopping/Finance, 🔍 Information Seeking, 🗺 Navigation/Transport, ✉ Communications, 👍 Social Networks, 📄 Files/Multimedia, ★ Lifestyle and ✂ Tools. One category includes *Borderline-shadow* HRD sources: 💗 Health/Fitness, and one category includes *Non-shadow* HRD sources: 📱 Medical. A representation of this classification can be found in Figure 2. To enhance clarity of the figure and given the focus of this work, *Non-shadow* and *Borderline-shadow* HRD sources are omitted. Table 1 in Appendix A provides examples of *shadow* HRD for non-obvious cases.

3.2.2 Discussion of HRD Sources. This classification, which we use when designing the questionnaire for our online user survey described in Section 4, depicts how numerous and diverse *shadow* HRD sources are, and shows that HRD can be created through many digital tools not primarily designed for health. It emphasizes the need for better protection of *shadow* HRD and can help researchers in the health, privacy, and legal fields, as it encourages consideration of the identified HRD sources when developing solutions for health data protection. Although some of these *shadow* HRD may not seem sensitive, they are considered as such in privacy regulations,[2] and hence must be identified and protected accordingly.

Aside from objective sensitivity, HRD sensitivity is also subjective and depends on the data subject and context. For example, a patient making a call to their general practitioner might be less concerned about their privacy than one making a phone call to their urologist, as more specific—and possibly more embarrassing—disorders may be inferred from the latter. Further, one may not consider inputting the address of their ophthalmologist in a navigation service as sensitive, while someone may consider inputting the address of an OB-GYN clinic providing abortion services as such. The degree to which sensitive information can be inferred from the presented *shadow* HRD, as well as the potential harm that could incur varies on the type of *shadow* HRD, and on the context. As with any HRD, they can reveal the medical conditions of the data subject as well as their general fitness level, which in turn may lead to increased insurance premiums, and more difficulty finding employment, or obtaining loans. Further, access to this information by others may cause embarrassment to the data subject and may have social and psychological repercussions on them [17]. We leave a more quantitative investigation of the extent of inference risks, based on existing inference attacks from the literature, for future work.

Some overarching concepts emerged through analysis of the data from the focus groups. The concept of interdependent privacy [30] appeared eight times (e.g., *"Sharing own and family and friends' health information on messaging systems"*, *"Using LLMs to search for the medical research related to brother's condition"*), spanning four main categories. The concept of syncing the data of apps and services, or files containing HRD to the cloud was also recurrent (e.g., *"Questions for medical appointment in notes app, connected to cloud"*). We include questions about both of these concepts in the questionnaire to assess the prevalence of these practices.

Finally, by examining the tags about structure of HRD, we notice that most *shadow* HRD are either semi-structured or unstructured, whereas most *borderline-shadow* and *non-shadow* HRD are structured. As more effort is required by an adversary to extract unstructured or semi-structured data, we design the questionnaire used for our user survey to assess, for some HRD sources, the level of structure of the HRD that individuals can create through them (e.g., whether or not they use tags when storing health-related photos). Although this classification is extensive, we do not claim it to be exhaustive, as we cannot guarantee that all HRD sources were mentioned during the focus groups or thought of by the research team. This classification is meant to be extensible and adaptable, welcoming the addition of new sources of HRD, in particular those stemming from the emergence of new technologies.

Further, note that we do not consider that all *shadow* HRD included in this classification should be categorized as *sensitive* data (in the legal sense of the term). However, it is important to highlight the ubiquity of *shadow* HRD, as adversaries attempting to infer health information about individuals could benefit from combining *shadow* HRD from multiple of the presented sources, to improve their inference.

4 Online Survey

To answer RQ2, we conduct a survey with users of general-purpose tools in order to assess the prevalence of behaviors that lead to the creation of *shadow* HRD, to understand the users' awareness
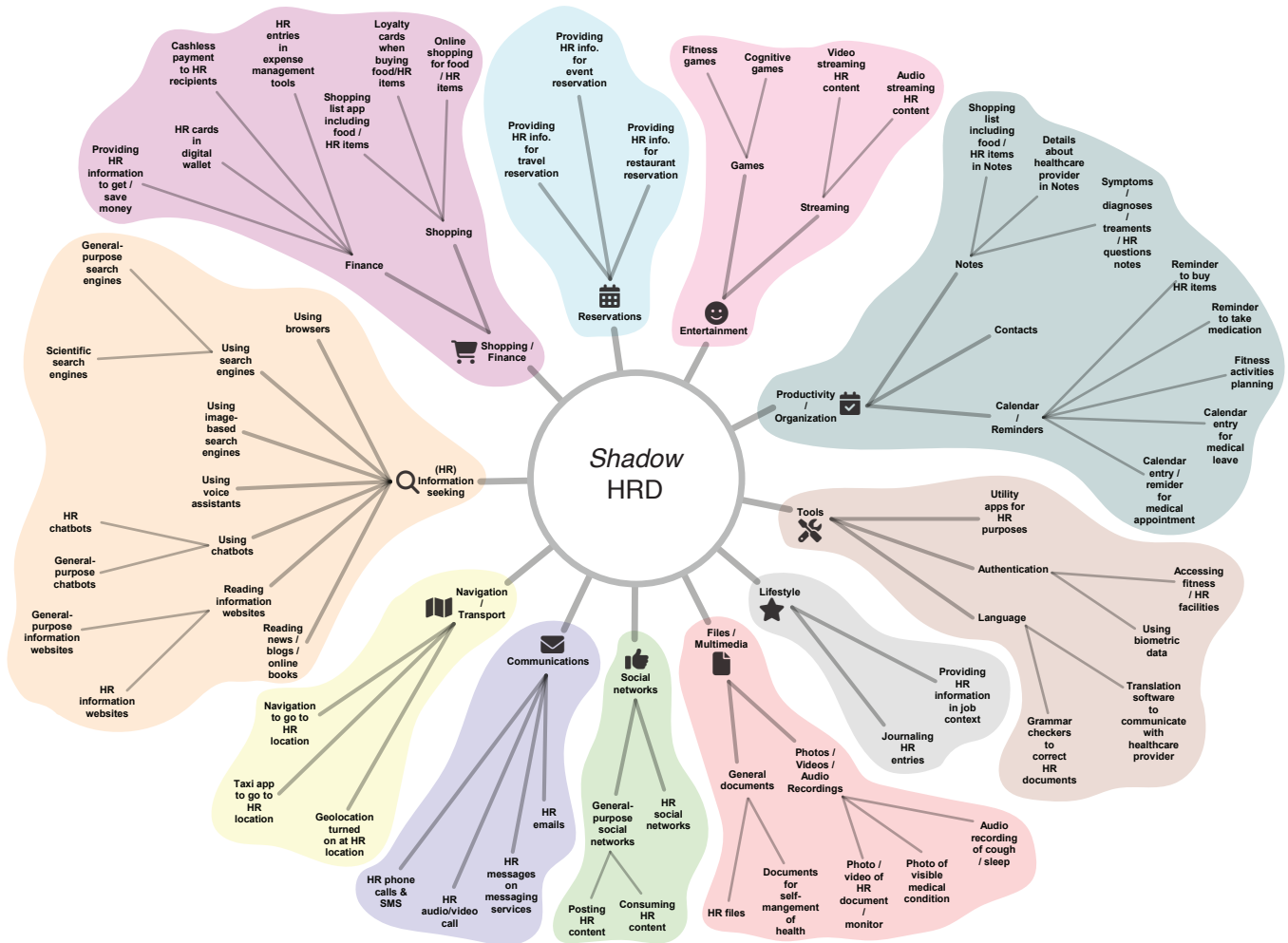
---

[2]https://www.edps.europa.eu/data-protection/our-work/subjects/health_en

Figure 2: Classifications of *shadow* HRD sources inventoried during the focus group study.

of the privacy risks stemming from these behaviors, and to know their concerns and whether they take any measures or use any mechanisms to protect their privacy. We further separated RQ2 into several sub-questions and designed our survey to address these specific questions: **(RQ2.1)** How prevalent are the behaviors that lead to the creation of *shadow* HRD? **(RQ2.2)** To what extent are users aware of the privacy risks associated with the *shadow* HRD created by their behaviors? **(RQ2.3)** Do users, for privacy reasons, purposefully refrain from taking certain actions in order to prevent the creation of *shadow* HRD? **(RQ2.4)** What privacy-preserving measures do users employ when dealing with their *shadow* HRD?

4.1 Survey - Methodology

Figure 1 includes our survey methodology. Due to the exploratory nature of the study, we did not conduct any statistical power analysis to determine the sample size. By referencing similar survey studies (e.g., $N = 227$ [69]; $N = 325$ [54]), we targeted recruitment of approximately 300 respondents. The survey study received approval from our IRB.

4.1.1 Recruitment. We recruited our survey respondents through Prolific, a platform recognized for its reliability in scientific research [51]. We recruited respondents from the United States to ensure cultural and societal relevance, as the US population offers diverse perspectives and behaviors regarding health-data sharing [68]. Given the high technology adoption rates in the US [12], this population is particularly suitable for studies involving online services and digital behavior.

We first conducted a screener survey to select respondents eligible for our main survey. In the survey, we asked respondents which apps or services they typically use, covering seven categories: (i) ✉ Communications (e.g., WhatsApp, Zoom), (ii) 📖 Navigation/Transport (e.g., Google/Apple Maps, Waze), (iii) 📆 Productivity/Organization (e.g., Notes, Calendar), (iv) 🛒 Shopping/Finance (e.g., online stores or banking), (v) 📄 Files/Multimedia (e.g., saved documents or photos), (vi) 🔍 Information Seeking (e.g., Google, ChatGPT, YouTube), and (vii) 👍 Social Networks (e.g., Instagram, Reddit). These categories were determined based on the findings of Section 3. To ensure all respondents were familiar with the categories of apps/services in the main survey, we recruited *only* those

who reported using all seven categories. We collected data from 1006 screener respondents and selected 617 who met this criterion.

4.1.2 Survey Design. We designed the questionnaire to gather information about users' awareness, behaviors, concerns, and the protective mechanisms they use with respect to *shadow* HRD. We selected seven categories of services, based on the classification outlined in Section 3.2. Although the classification covers *shadow* HRD, *borderline-shadow* HRD (e.g., fitness apps), and *non-shadow* HRD (e.g., medical apps), our survey focused primarily on *shadow* HRD categories, thus aligning with the work's main objectives. For the questionnaire to be concise, we excluded categories with niche behaviors, such as ★ Lifestyle or 🗓 Reservations, that are less likely to be applicable to a large group of individuals (e.g., including health-related information in journaling apps). Given that the survey was conducted in the US, we ensured that the practices mentioned were relevant to US residents, and we included examples of popular apps and services commonly used in the US, by searching for the apps with the largest number of US users to illustrate each practice. The questionnaire included 34 items, organized into six sections. The number of items in certain sections varied based on the respondents' earlier responses (i.e., display logic). The survey was intended to take around 15 min to complete. The complete questionnaire can be found in Appendix C. Each section of the survey is detailed below:

Sec. A: Introduction. The survey begins with a consent form. Using a seven-point Likert scale, they were asked to describe their level of engagement with health-related activities (e.g., having regular health checkups), ranging from *"not at all engaged"* to *"extremely engaged"*. For quality control (as recommended by Prolific), respondents were asked to answer the same question from the screener survey regarding the apps or services they typically use. If there was a mismatch in their responses, their participation in the survey was terminated.

Sec. B: User Awareness of HRD. This section includes a single question used to collect the respondents' (self-reported) assessment of the likelihood that various categories of apps or services contain (shadow) HRD. Respondents rated, on a scale from *"extremely unlikely"* to *"extremely likely"*, how likely they believe each of the aforementioned seven categories of apps or services is to contain information about their health. We do not provide specific examples of information about health to the respondents, in order not to prime them, and capture their "true" awareness which drives their behaviors with respect to *shadow* HRD.

Sec. C: User Behavior Regarding HRD. This section consists of seven blocks, each block focusing on one of the seven categories: ✉ Communications, 🗺 Navigation/Transport, 🗓 Productivity/Organization, 🛒 Shopping/Finance, 📄 Files/Multimedia, 🔍 Information Seeking, and 👍 Social Networks. For each category, we posed the following types of questions: (i) ***Shadow* HRD-creating Behavior:** To assess whether respondents engage in behaviors that could lead to the creation of *shadow* HRD. For example, for the category of 🔍 Information Seeking, respondents were asked (see Q20) which service they typically use to look for symptoms or health-related information or to better communicate with healthcare providers, with options such as health information service

websites (e.g., WebMD), online translation tools (e.g., Google Translate), and chatbots (e.g., ChatGPT). (ii) **Refraining Actions:** To identify actions they purposely refrain from to avoid the creation of *shadow* HRD. For example, for the category of 👍 Social Networks, respondents were asked (see Q26) if they purposefully refrained from any actions for privacy reasons, such as posting about their health condition, following social media accounts dedicated to a health condition they have, or engaging in forums or support groups related to a health condition they have. (iii) **Protection Strategies:** To determine actions they take to protect the privacy and anonymity of their HRD. To identify the protection strategies that can be applied on the user side, one PhD student and one professor from the research team conducted literature review and iterative discussions. Respondents were asked if they resorted to these strategies. For example, for 🗺 Navigation/Transport, respondents were asked (see Q9) if they typically take any measures to protect their privacy when going to a health-related appointment, such as using maps in offline mode, deleting their location history after the appointment, or using private or incognito mode. Not all three types of questions were asked for every category; the relevance of the practices and context determined which questions were posed.

We also included questions, by inquiring whether files and media are tagged or given relevant file names, to capture specific behaviors related to the level of detail in the piece of *shadow* HRD, such as the details included when saving a physician's contact in a digital address book (e.g., medical specialty) (see Q12), and in order to understand the level of structure of health-related files and media (see Q19). Finally, concerning protection measures, we verified, through follow-up questions, the reliability of those reported by the respondents in order to ensure that the selected protection mechanisms were indeed applied by the respondents (see Q6 or Q21). For example, when a respondent reported using end-to-end encrypted messaging services for health-related communications, we followed up with a question asking about what specific services they use.

Sec. D: General Privacy Aspects. This section includes three questions about respondents' perceptions and behaviors related to privacy. First, a matrix question assessed respondents' levels of concern on a seven-point Likert scale, from *"Not at all concerned"* to *"Extremely concerned,"* about various entities having access to their HRD (see Q27). We included entities such as intimate partners [39], hackers, employers, and health insurance companies. Second, to gauge the severity of privacy threats (see Q28), respondents were asked which types of personal data are synced to their cloud accounts. To maximize the quality of the responses, we provided the respondents with precise instructions on how to check the information in their phone parameters, for both Android and iOS. Some of the options included contacts, calendars, notes, galleries, passwords, and wallets. Third, one question explored the notion of 'interdependent privacy' [30] and whether respondents engage in behaviors that lead to the creation of *shadow* HRD about others, such as significant other (S/O), child, and parent (see Q29).[3]

---

[3]When wording the question, we provided the examples of S/O, child and parent to respondents, as these examples were mentioned by several participants during the focus group sessions.

Sec. E: Health-Related Privacy Concerns. In this section, we employ a standard scale [7] to assess respondents' concerns regarding the privacy of their HRD. Respondents had to rate three statements on a 10-point Likert scale. They evaluated the *advisability* of submitting health information online, their concerns about potential *abuse* of this information, and the likelihood that submitted health information could be *compromised*, by being shared or sold.

Sec. F: Demographics. This section includes questions to gather demographic information about the respondents. Respondents were asked about their gender identity (following guidelines in [62]) and their field of work or study. The purpose of the latter was to identify how many respondents were from relevant fields such as health, medicine, computer science, and IT. Other demographics (e.g., age) are provided by Prolific.

4.1.3 Procedure and Data Reliability. Before launching the survey, we conducted online cognitive pretests to identify and resolve any potential issues with the questionnaire. We asked two Ph.D. students from our department, not involved in this research project, to take the survey. During each pretest, the first author closely observed the respondents as they completed the survey and asked them to rephrase the questions in their own words. After responding to all questions, the first author and pretest respondent discussed the respondent's understanding and responses. With this cognitive pretest, we verified that the questions were generally clear. Only a few comprehension issues were identified and addressed. For example, we clarified the term "pseudonymous" by adding "(i.e., fake)" and reworded "Never granting apps access" to "Carefully reviewing app permissions" when asking about measures to protect the confidentiality of health-related files.

Although Prolific is a reputable crowdsourcing platform [51], it cannot entirely prevent undesirable behaviors. Therefore, we implemented several strategies to ensure data quality. First, respondents with mismatched responses between the main survey and the screener were excluded. Second, we included three attention checks and excluded the 95 respondents who did not select all three. Third, we excluded two respondents who completed the survey in less than five minutes, categorizing them as *speeders*. Finally, to minimize order effects [63], we randomized the presentation of the seven categories in Section C and the options in multiple-choice questions (MCQs), except for Likert scales.

As a result, out of the 617 eligible respondents, 508 began the main survey. Ultimately, 397 completed the main survey, with *N* = 300 included in the final analysis. On average, it took respondents 14 min, 44 sec to complete the survey (SD: 8 min, Min: 5 min, 59 sec, Max: 51 min, 43 sec). Respondents were compensated GBP 2.25 (∼ USD 2.84).

In an effort to represent the opinions of all genders equally, we also strove to achieve gender balance in our sample. Utilizing Prolific's demographic data, we deployed the survey in multiple batches to control for gender balance.[4]

4.1.4 General Statistics. The average age of the retained respondents was 39.1 (SD: 11 years). For comparison, the average age of

---

[4]According to the latest census [11], the US population is composed of approximately 50.8% of women.
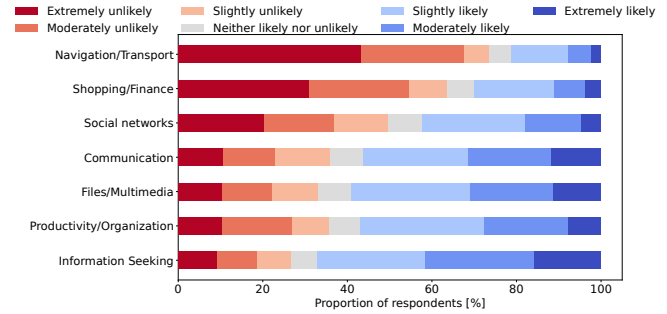
Figure 3: The level of user awareness about the possibility of (shadow) HRD creation across seven categories of apps and services.
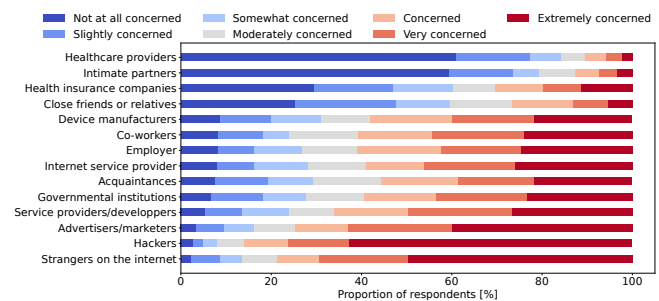
**Figure 4: The level of user concern about HRD being accessed by different types of adversaries**

the respondents to the screener survey was 40.0 (SD: 13.1 years). The inclusion criteria applied through our screener survey did not have an important effect on the age distribution of participants. Figure 9 in Appendix B provides more details with regard to age and depicts the fact that no particular age groups were excluded due to our inclusion criteria. Regarding gender, 50.5% (N=151) identified as women, 47.5% (N=143) as men, 1% (N=3) as non-binary, 1 respondent declared to be questioning, and 1 preferred not to answer. Respondents came from a variety of professional backgrounds. Relevant to this work, we report 16% from the IT/computer science field, and 12% from the health/healthcare field. Regarding health-related privacy concerns, most respondents exhibited moderate concerns about health privacy (Advisability: ($M = 4.5$, $SD = 1.8$), Abuse: ($M = 5.9$, $SD = 1.9$), Compromise: ($M = 7.3$, $SD = 2.2$)). Regarding the engagement with health-related activities, which we evaluated on a 7-point Likert scale ranging from *Not at all engaged* to *Extremely engaged*, we report 45% of respondents reporting being at least *Well engaged*. 21% reported being *Moderately engaged* while 34% reported being at most *Somewhat engaged*. More details can be found in Figure 10 in Appendix B.

4.2 Survey - Results and Discussion

Here, we present the survey findings and discuss their implications.

4.2.1 Users Have Varying Levels of Awareness and Concern.

Awareness. In Q3, we look at respondents' level of awareness regarding the possibility of *shadow* HRD creation. When asked how
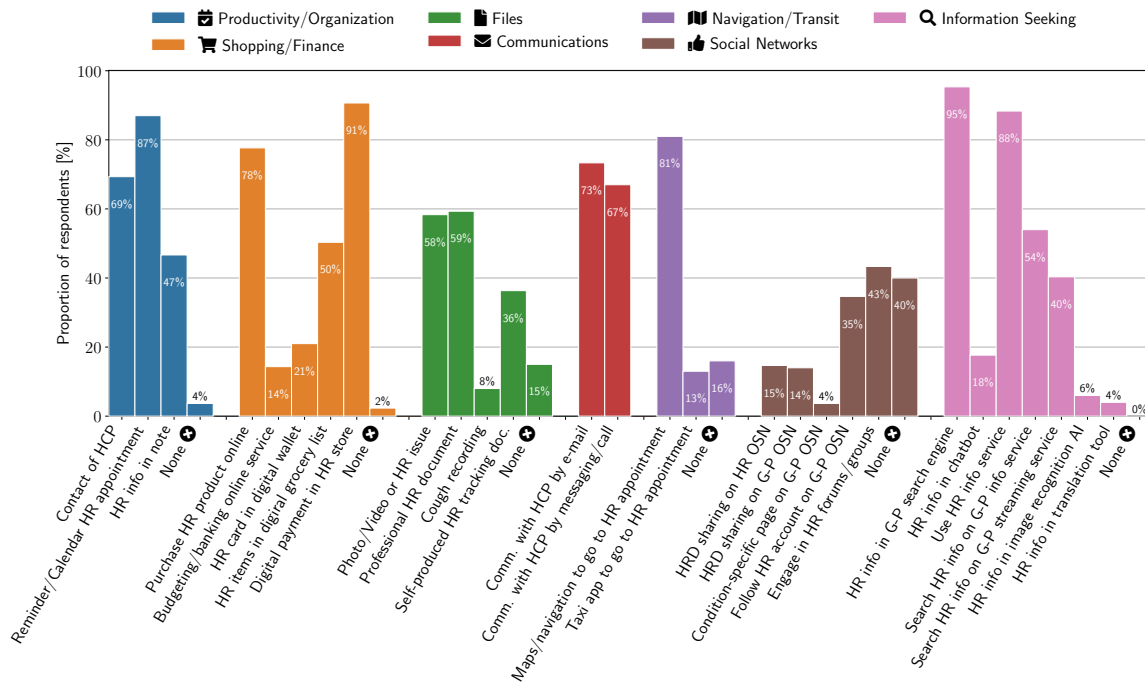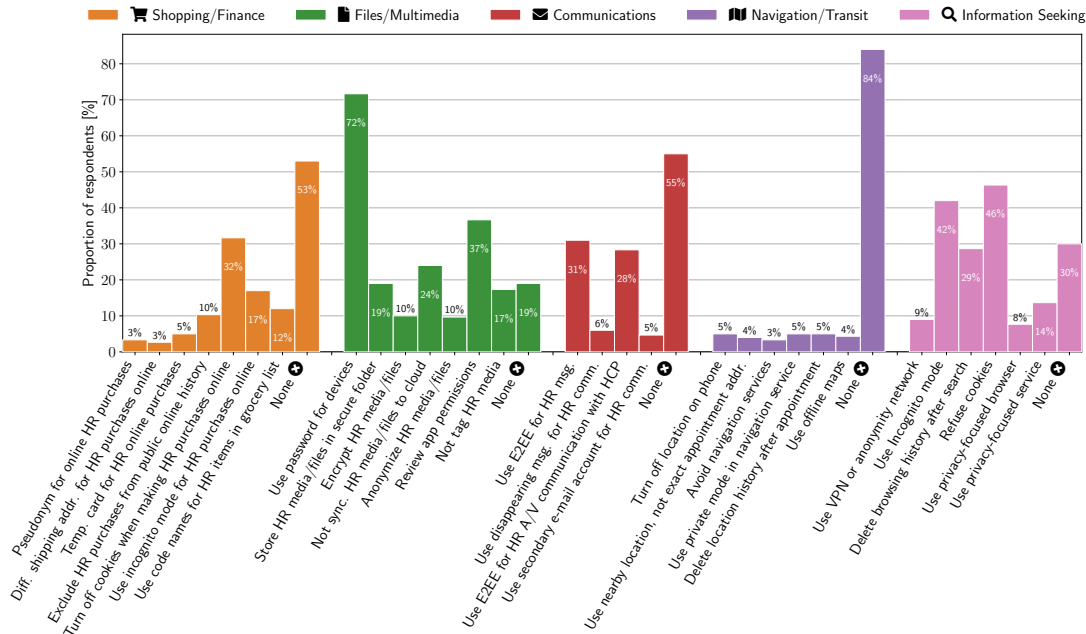
Figure 5: Different types of user behaviors that lead to creation of *shadow* HRD, self-reported across seven categories. Acronyms as follows. *HR*: health-related, *HCP*: healthcare provider, *OSN*: online social networks, *G-P*: general-purpose.

likely they think that (shadow) HRD would be contained in different types of such apps and services, more than half deemed it at least *slightly likely* for HRD to be contained in apps and services belonging to the following categories: ☑ Productivity/Organization, 🔍 Information Seeking, ✉ Communications, and 📄 Files/Multimedia, though they deemed it at least *slightly unlikely* for apps belonging to the 🛒 Shopping/Finance, and 📖 Navigation/Transport categories. The trend is less clear for the 👍 Social Networks category, with 42.3% of respondents deeming such apps and services to contain HRD likely, 49.6% deeming it unlikely, and 8.1% deeming it *neither likely nor unlikely*. The results are summarized in Figure 3. Thus, respondents exhibited a higher level of awareness with regards to *shadow* HRD creation through ☑ Productivity/Organization, 🔍 Information Seeking, ✉ Communications, and 📄 Files/Multimedia apps and services, than for those belonging to the 📖 Navigation/Transport and 🛒 Shopping/Finance categories.

Concerns. We also looked, through Q27, at respondents' concerns regarding their HRD being accessed by adversaries. The results are summarized in Figure 4. A majority of respondents reported little concern about healthcare providers and intimate partners having access to their HRD, with 61.6% and 59.2% reporting being *not at all concerned*, respectively. This result was expected, as individuals typically exhibit high levels of trust towards intimate partners (despite recent studies on privacy infringements by the latter [39]) and healthcare providers. In the latter case, the results can be explained by the trade-off between utility and privacy, and is also exemplified by prior work showing that patients typically trust healthcare institutions with the privacy and security of their EHR. On the contrary, respondents were most concerned about hackers, strangers on the

Internet, and advertisers gaining access to their HRD, with 62.6%, 49.3%, and 39.8% respondents reporting being *extremely concerned* respectively with regard to these adversaries.

4.2.2 Users Engage in Shadow *HRD-Creating Behaviors.* In Figure 5, we summarize the findings about user behavior (Q4, Q8, Q10, Q14, Q17, Q20, Q24). Most notably (i.e., behaviors selected by more than half of respondents), in the ☑ Productivity/Organization category, 69% reported saving the contact of their healthcare provider and 87% reported having reminders or calendar entries for health appointments. In the 🛒 Shopping/Finance category, 78% reported purchasing health-related products online, 50% reported adding health-related items to their shopping list, and 90% reported using digital payments in health-related stores (e.g., pharmacy). In the 📄 Files/Multimedia category, 58% reported taking photos or videos of health-related issues, and 59% reported saving medical files (generated within a professional healthcare context) to their devices. In the ✉ Communications category, 73% respondents reported communicating with healthcare providers via e-mail, and 67% via messaging or calling services (i.e., voice or video over IP). In the 📖 Navigation/Transport category, 81% respondents reported using a map, navigation, or transit service (e.g., Google Maps) to go to health-related appointments. In the 🔍 Information Seeking category, for finding health-related information, 95% respondents reported using general-purpose search engines (e.g., Google), 89% reported using a health-related information service (e.g., WebMD), and 54% reported using a general-purpose information service (e.g., Wikipedia). Even though LLM-based chatbots are an emerging technology, a non-negligible proportion (17%) of respondents reported

Figure 6: Users' protection strategies to avoid creation of *shadow* HRD or to hide own identity, self-reported across five categories. Acronyms as follows. *HR*: health-related, *HCP*: healthcare provider, *E2EE*: end-to-end encryption.

using them when seeking health information. Finally, for the 👍 Social Networks category, neither of the behaviors included in the questionnaire was selected by more than half of the respondents. As such, *shadow* HRD-creating behaviors related to social media are less prevalent than behaviors from other categories. Social dynamics within social media services differ from those of other digital services. Individuals may be more willing to share HRD when they believe access to the HRD is limited to the service provider, as opposed to sharing it on social media, where the audience could be people they personally know, or entities interested in their HRD (e.g., health insurers). Users may (often erroneously) trust service providers to protect their HRD and not share it beyond their servers, while they may not trust that social media posts about their health would stay confined to the platform on which it was shared, complicating control over online HRD. With regards to more specific behaviors (Q12, Q19), a large majority reported including identifying information about healthcare providers when saving their contact details (86% include their name, 88% their phone number). More than half (51%) reported explicitly including the healthcare provider's medical specialty (e.g., nephrology), and a non-negligible proportion (21%) reported including health-related notes. Given such level of detail, digital address books can be revealing of health conditions of individuals, as they make the data richer and easier to exploit by adversaries with access to address book data. The indicated specialty may be used as a keyword to infer one's physical or mental health conditions with more precision (e.g., having kidney disease). However, when asked how often they tag, categorize, or give meaningful names to health-related files and media, 64% reported either never or rarely doing so, thus making it more

difficult for adversaries to identify *shadow* HRD contained in files and media.

4.2.3 Users Rarely Use Protection Mechanisms. Our findings regarding refraining from engaging in *shadow* HRD-creating behaviors or using protection measures to mitigate privacy risks relating to *shadow* HRD are summarized in Figure 6 and presented in detail here. Most of the respondents, for all categories except 👍 Social Networks, either never refrain from engaging in *shadow* HRD-leaking behaviors or do so rarely. These results are aligned with patterns of users of wearable devices, who also rarely refrain from removing their fitness tracker in privacy-sensitive situations [69]. As for the 👍 Social Networks category, the responses were more diverse, with 37% reporting that they never refrain from *shadow* HRD-creating behaviors on social networks, and 26% reporting that they always or very often refrain.

With regard to the use of protection mechanisms, for three out of five of the categories for which such questions were asked (🛒 Shopping/Finance, ✉ Communications, 📖 Navigation/Transport), most respondents reported using none of the proposed protection mechanisms. For the 📄 Files/Multimedia category, 72% reported using passwords for their devices to protect their files and media, among other data. However, such a measure protects against only adversaries in close physical proximity. The 🔍 Information Seeking category was the one for which people used more protection mechanisms. Finally, with regard to the 👍 Social Networks category (Q25), when asked whether they used anonymization techniques, such as face blurring or using pseudonyms when sharing HRD on social networks or forums, 51% reported either never or rarely doing so, whereas 17% reported either always or very often doing so.
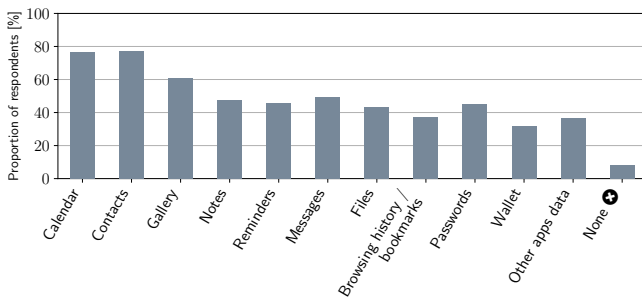
Figure 7: Shadow HRD types synchronized to Cloud
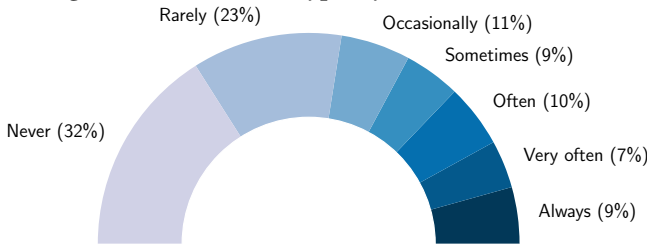
Figure 8: Frequency of engaging in HRD-creating behaviors with consequences for others, such as S/O, child, parents. (Interdependent privacy [30])

4.2.4 Severity of Privacy Threat. To evaluate the severity of the privacy threat, we investigated which application data was synchronized to cloud services, and whether or not respondents engage in behaviors that create *shadow* HRD concerning others (Q29). The results are summarized in Figure 7 and Figure 8 respectively. Most respondents synchronize their calendars, contacts, and gallery (i.e., photos and videos). For the remaining apps and services, a non-negligible proportion (consistently above 30%) reported synchronizing them to the cloud. These results show that, although some of the *shadow* HRD-creating behaviors identified in this paper pertain to local apps or data (e.g., reminders for health-related appointments), these *shadow* HRD are often propagated beyond local devices, and could consequently be accessed by online service providers and third parties.

Although the majority of respondents reported not engaging in behaviors creating *shadow* HRD about others, a non-negligible proportion of 26% reported doing so at least often. This suggests that the privacy risks surrounding *shadow* HRD not only pertain to the individual creating them but can also affect others around them. Interdependent privacy risks are not well covered by data protection laws, as the data generator is not the sole data subject.

5 Conclusion

Shadow HRD are sensitive health data that potentially evade being treated as such with regards to the enforcement of data protection laws, since they do not stem clearly from health-related technologies or contexts, and are hence more difficult to identify. In this work, we generate, through two focus group sessions, an inventory of *shadow* HRD sources, and categorize them to highlight the technologies through which they can be created. The resulting classification is meant to be extensible, towards developing a comprehensive

taxonomy, with the long-term purpose of covering as many possible sources of *shadow* HRD, thus raising awareness of users about privacy risks to their HRD, associated with the use of seemingly non-health-related technologies. Another one of its purposes is to encourage privacy researchers and law practitioners to consider *shadow* HRD when developing solutions for health-related data protection. Our assessment, through a large-scale online survey, of user behaviors, awareness and concerns with respect to *shadow* HRD shows that their creation is widespread among technology users, who rarely take protective measures to mitigate privacy risks relating to them, but rather engage in practices such as synchronizing these data, either intentionally or inadvertently, to the cloud, thus exacerbating the privacy risks. Further, we show that these risks are not constrained to the individual creating them, but could apply to others through interdependent privacy-related practices. We envision deploying our survey and extending our analysis to respondents outside the US, to capture a diversity of possibly socioeconomically-dependent *shadow* HRD-creating behaviors. We also aim to conduct in-depth interviews with some survey respondents, to further investigate their thought process with respect to their reported behaviors, and understand the privacy-utility trade-offs at hand. Finally, this work is focused on *shadow* HRD from the technology users perspective, but does not specifically consider *shadow* HRD-creating behaviors of health practitioners. As future work, we intend to address the latter, although this endeavor could be more challenging due to the legal frameworks around such practices.

References

[1] Angeliki Aktypi, Jason R.C. Nurse, and Michael Goldsmith. 2017. Unwinding Ariadne's Identity Thread: Privacy Risks with Fitness Trackers and Online Social Networks. In *Proceedings of the 2017 on Multimedia Privacy and Security*. ACM, Dallas Texas USA, 1–11. https://doi.org/10.1145/3137616.3137617

[2] Abdulmajeed Alqhatani and Heather Richter Lipford. 2019. "There is nothing that I need to keep secret": Sharing Practices and Concerns of Wearable Fitness Data. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. USENIX Association, Santa Clara, CA, 421–434. https://www.usenix.org/conference/soups2019/presentation/alqhatani

[3] Mayda Alrige and Samir Chatterjee. 2015. Toward a Taxonomy of Wearable Technologies in Healthcare. In *New Horizons in Design Science: Broadening the Research Agenda*, Brian Donnellan, Markus Helfert, Jim Kenneally, Debra VanderMeer, Marcus Rothenberger, and Robert Winter (Eds.). Springer International Publishing, Cham, 496–504.

[4] Jessica S. Ancker, Alison M. Edwards, Melissa C. Miller, and Rainu Kaushal. 2012. Consumer Perceptions of Electronic Health Information Exchange. *American Journal of Preventive Medicine* 43, 1 (July 2012), 76–80. https://doi.org/10.1016/j.amepre.2012.02.027

[5] Khadija Baig, Reham Mohamed, Anna-Lena Theus, and Sonia Chiasson. 2020. "I'm hoping they're an ethical company that won't do anything that I'll regret": Users Perceptions of At-home DNA Testing Companies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu

HI USA, 1–13. https://doi.org/10.1145/3313831.3376800

[6] Khadija Baig, Daniela Napoli, and Sonia Chiasson. 2023. A comparison of users' and non-users' perceptions of health and ancestry at-home DNA testing. In *Proceedings of the 2023 European Symposium on Usable Security*. ACM, Copenhagen Denmark, 48–67. https://doi.org/10.1145/3617072.3617107

[7] Gaurav Bansal, Fatemeh "Mariam" Zahedi, and David Gefen. 2010. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision Support Systems* 49, 2 (May 2010), 138–150. https://doi.org/10.1016/j.dss.2010.01.010

[8] Rashid Bashshur, Gary Shannon, Elizabeth Krupinski, and Jim Grigsby. 2011. The Taxonomy of Telemedicine. *Telemedicine and e-Health* 17, 6 (July 2011), 484–494. https://doi.org/10.1089/tmj.2011.0103

[9] Emiel A Boogerd, Tessa Arts, Lucien JLPG Engelen, and Tom H van de Belt. 2015. "What Is eHealth": Time for An Update? *JMIR Res Protoc* 4, 1 (March 2015), e29. https://doi.org/10.2196/resprot.4065

[10] Adele Botha, Martin Weiss, and Marlien Herselman. 2018. Towards a Taxonomy of mHealth. In *2018 International Conference on Advances in Big Data, Computing and Data Communication Systems (icABCD)*. IEEE, Durban, South Africa, 1–9. https://doi.org/10.1109/ICABCD.2018.8465427

[11] US Census Bureau. 2020. Age and Sex Composition: 2020. https://www.census.gov/library/publications/2023/decennial/c2020br-06.html Section: Government.

[12] US Census Bureau. 2023. Three Results From Recent Research on Advanced Technology Use and Automation. https://www.census.gov/newsroom/blogs/research-matters/2023/09/advanced-technology-use-and-automation-results.html Section: Government.

[13] K. Caine and R. Hanania. 2013. Patients want granular privacy control over health information in electronic medical records. *Journal of the American Medical Informatics Association* 20, 1 (Jan. 2013), 7–15. https://doi.org/10.1136/amiajnl-2012-001023

[14] Jiaxun Cao, Hiba Laabadli, Chase H Mathis, Rebecca D Stern, and Pardis Emami-Naeini. 2024. "I Deleted It After the Overturn of Roe v. Wade": Understanding Women's Privacy Concerns Toward Period-Tracking Apps in the Post Roe v. Wade Era. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–22. https://doi.org/10.1145/3613904.3642042

[15] Paulina Chametka, Sana Maqsood, and Sonia Chiasson. 2023. Security and Privacy Perceptions of Mental Health Chatbots. In *2023 20th Annual International Conference on Privacy, Security and Trust (PST)*. IEEE, Copenhagen, Denmark, 1–7. https://doi.org/10.1109/PST58708.2023.10320174

[16] Amy Hai Yan Chan and Michelle L. L. Honey. 2022. User perceptions of mobile digital apps for mental health: Acceptability and usability - An integrative review. *Journal of Psychiatric and Mental Health Nursing* 29, 1 (Feb. 2022), 147–168. https://doi.org/10.1111/jpm.12744

[17] Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule, Board on Health Sciences Policy, Board on Health Care Services, and Institute of Medicine. 2009. *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*. National Academies Press, Washington, D.C. https://doi.org/10.17226/12458 Pages: 12458.

[18] Pouyan Esmaeilzadeh and Tala Mirzaei. 2018. Comparison of consumers' perspectives on different health information exchange (HIE) mechanisms: an experimental study. *International Journal of Medical Informatics* 119 (Nov. 2018), 1–7. https://doi.org/10.1016/j.ijmedinf.2018.08.007

[19] Houda Fakhkhari, Bouchaib Bounabat, and Ismail Kassou. 2023. Digital Health Taxonomy. In *Proceedings of the 6th International Conference on Networking, Intelligent Systems & Security (NISS '23)*. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3607720.3607780

[20] Jeana Frost, Ivar E Vermeulen, and Nienke Beekers. 2014. Anonymity Versus Privacy: Selective Information Sharing in Online Cancer Communities. *Journal of Medical Internet Research* 16, 5 (May 2014), e126. https://doi.org/10.2196/jmir.2684

[21] Sandra Gabriele and Sonia Chiasson. 2020. Understanding Fitness Tracker Users' Security and Privacy Knowledge, Attitudes and Behaviours. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–12. https://doi.org/10.1145/3313831.3376651

[22] Deborah Goodman, Catherine O. Johnson, Deborah Bowen, Megan Smith, Lari Wenzel, and Karen Edwards. 2017. De-identified genomic data sharing: the research participant perspective. *Journal of Community Genetics* 8, 3 (July 2017), 173–181. https://doi.org/10.1007/s12687-017-0300-1

[23] L. Goodwin, K. Courtney, J.D. Kirby, M.A. Iannacchione, and T. Manley. 2002. A Pilot Study: Patients' Perceptions About the Privacy of Their Medical Records. *Online Journal of Nursing Informatics* 6 (Sept. 2002), 1–16. https://ojni.org/1002/courtney.htm

[24] David Grande, David A. Asch, Fei Wan, Angela R. Bradbury, Reshma Jagsi, and Nandita Mitra. 2015. Are Patients With Cancer Less Willing to Share Their Health Information? Privacy, Sensitivity, and Social Purpose. *Journal of Oncology Practice* 11, 5 (Sept. 2015), 378–383. https://doi.org/10.1200/JOP.2015.004820

[25] Maike Greve, Tim-Benjamin Lembcke, Stephan Diederich, Alfred Benedikt Brendel, and Lutz M. Kolbe. 2020. Healthy by App - Towards a Taxonomy of Mobile Health Applications. In *24th Pacific Asia Conference on Information Systems, PACIS 2020*. AIS, Dubai, UAE, 217. https://aisel.aisnet.org/pacis2020/217

[26] Ece Gumusel. 2024. A literature review of user privacy concerns in conversational chatbots: A social informatics approach: An Annual Review of Information Science and Technology (ARIST) paper. *Journal of the Association for Information Science and Technology* Early view (May 2024), 1–34. https://doi.org/10.1002/asi.24898

[27] Steffi Haag and Andreas Eckhardt. 2017. Shadow IT. *Business & Information Systems Engineering* 59, 6 (Dec. 2017), 469–473. https://doi.org/10.1007/s12599-017-0497-x

[28] Mathias Humbert, Erman Ayday, Jean-Pierre Hubaux, and Amalio Telenti. 2017. Quantifying Interdependent Risks in Genomic Privacy. *ACM Transactions on Privacy and Security* 20, 1 (Feb. 2017), 1–31. https://doi.org/10.1145/3035538

[29] Mathias Humbert, Didier Dupertuis, Mauro Cherubini, and Kévin Huguenin. 2022. KGP Meter: Communicating Kin Genomic Privacy to the Masses. In *2022 IEEE 7th European Symposium on Security and Privacy (EuroS&P)*. IEEE, Genoa, Italy, 410–429. https://doi.org/10.1109/EuroSP53844.2022.00033

[30] Mathias Humbert, Benjamin Trubert, and Kévin Huguenin. 2020. A Survey on Interdependent Privacy. *Comput. Surveys* 52, 6 (Nov. 2020), 1–40. https://doi.org/10.1145/3360498

[31] Valérie Junod and Sami Salihu. 2020. Interdependent Privacy & Medical Information. *Life Science Rechts (LSR)* 2020, 4 (Nov. 2020), 195–204. https://lsr.recht.ch/de/artikel/02lsr0420auf/interdependent-privacy-medical-information

[32] Ahmed E. Khaled. 2022. Internet of Medical Things (IoMT): Overview, Taxonomies, and Classifications. *Journal of Computer and Communications* 10, 08 (2022), 64–89. https://doi.org/10.4236/jcc.2022.108005

[33] Katherine K Kim, Jill G Joseph, and Lucila Ohno-Machado. 2015. Comparison of consumers' views on electronic data sharing for healthcare and research. *Journal of the American Medical Informatics Association* 22, 4 (July 2015), 821–830. https://doi.org/10.1093/jamia/ocv014

[34] Michio Kimura, Jun Nakaya, Hiroshi Watanabe, Toshiro Shimizu, and Kazuyuki Nakayasu. 2014. A Survey Aimed at General Citizens of the US and Japan about Their Attitudes toward Electronic Medical Data Handling. *International Journal of Environmental Research and Public Health* 11, 5 (April 2014), 4572–4588. https://doi.org/10.3390/ijerph110504572

[35] Richard A. Krueger and Mary Anne Caey. 2014. *Focus Groups: A Practical Guide for Applied Research*. Sage Publ, 2455 Teller Road Thousand Oaks, CA 91320. https://us.sagepub.com/en-us/nam/focus-groups/book243860

[36] Karen Lamb, Hsiao-Ying Huang, Andrew Marturano, and Masooda Bashir. 2016. Users' Privacy Perceptions About Wearable Technology: Examining Influence of Personality, Trust, and Usability. In *Advances in Human Factors in Cybersecurity*, Denise Nicholson (Ed.). Vol. 501. Springer International Publishing, Cham, 55–68. https://doi.org/10.1007/978-3-319-41932-9_6 Series Title: Advances in Intelligent Systems and Computing.

[37] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. Chapter 11 - Analyzing Qualitative Data. In *Research Methods in Human Computer Interaction (Second Edition)*, Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser (Eds.). Morgan Kaufmann, Boston, 187–228. https://doi.org/10.1016/B978-0-12-805390-4.00008-X

[38] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. Chapter 8 - Interviews and focus groups. In *Research Methods in Human Computer Interaction (Second Edition)*, Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser (Eds.). Morgan Kaufmann, Boston, 187–228. https://doi.org/10.1016/B978-0-12-805390-4.00008-X

[39] Karen Levy and Bruce Schneier. 2020. Privacy threats in intimate relationships. *Jour. of Cybersecurity* 6, 1 (Jan. 2020), 1–13. https://doi.org/10.1093/cybsec/tyaa006 Publisher: Oxford Academic.

[40] Rasha Mahmoud, Alan R. Moody, Moran Foster, Natasha Girdharry, Loreta Sinn, Bowen Zhang, Mariam Afshin, Thayalasuthan Vivekanandan, Samantha Santoro, and Pascal N. Tyrrell. 2019. Sharing De-identified Medical Images Electronically for Research: A Survey of Patients' Opinion regarding Data Management. *Canadian Association of Radiologists Journal* 70, 3 (Aug. 2019), 212–218. https://doi.org/10.1016/j.carj.2019.04.002

[41] Lisa Mekioussa Malki, Ina Kaleva, Dilisha Patel, Mark Warner, and Ruba Abu-Salma. 2024. Exploring Privacy Practices of Female mHealth Apps in a Post-Roe World. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–24. https://doi.org/10.1145/3613904.3642521 event-place: , Honolulu, HI, USA,.

[42] Maryam Mehrnezhad and Teresa Almeida. 2023. ""My sex-related data is more sensitive than my financial data and I want the same level of security and privacy\": User Risk Perceptions and Protective Actions in Female-oriented Technologies". In *Proceedings of the 2023 European Symposium on Usable Security*. ACM, Copenhagen Denmark, 1–14. https://doi.org/10.1145/3617072.3617100

[43] Michael Methlagl, Friederike Michlmayr, and Valentina Perillo. 2023. Technological Trust Perceptions in Wearable Fitness Technology: A Person-Centred Approach. *Journal of Technology in Behavioral Science* 8, 4 (May 2023), 392–401. https://doi.org/10.1007/s41347-023-00320-7

[44] Anna Middleton, Richard Milne, Mohamed A. Almarri, Shamim Anwer, Jerome Atutornu, Elena E. Baranova, Paul Bevan, Maria Cerezo, Yali Cong, Christine Critchley, Josepine Fernow, Peter Goodhand, Qurratulain Hasan, Aiko Hibino,

Gry Houeland, Heidi C. Howard, S. Zakir Hussain, Charlotta Ingvoldstad Malmgren, Vera L. Izhevskaya, Aleksandra Jędrzejak, Cao Jinhong, Megumi Kimura, Erika Kleiderman, Brandi Leach, Keying Liu, Deborah Mascalzoni, Álvaro Mendes, Jusaku Minari, Nan Wang, Dianne Nicol, Emilia Niemiec, Christine Patch, Jack Pollard, Barbara Prainsack, Marie Rivière, Lauren Robarts, Jonathan Roberts, Virginia Romano, Haytham A. Sheerah, James Smith, Alexandra Soulier, Claire Steed, Vigdís Stefànsdóttir, Cornelia Tandre, Adrian Thorogood, Torsten H. Voigt, Anne V. West, Go Yoshizawa, and Katherine I. Morley. 2020. Global Public Perceptions of Genomic Data Sharing: What Shapes the Willingness to Donate DNA and Health Data? *The American Journal of Human Genetics* 107, 4 (Oct. 2020), 743–752. https://doi.org/10.1016/j.ajhg.2020.08.023

[45] Joanna Alicja Muras, Vinny Cahill, and Emma Katherine Stokes. 2006. A Taxonomy of Pervasive Healthcare Systems. In *2006 Pervasive Health Conference and Workshops*. IEEE, Innsbruck, 1–10. https://doi.org/10.1109/PCTHEALTH.2006.361680

[46] Robert C Nickerson, Upkar Varshney, and Jan Muntermann. 2013. A method for taxonomy development and its application in information systems. *European Journal of Information Systems* 22, 3 (May 2013), 336–359. https://doi.org/10.1057/ejis.2012.26

[47] Robert C. Nickerson, Upkar Varshney, Jan Muntermann, and Henri Isaac. 2009. Taxonomy development in information systems: Developing a taxonomy of mobile applications. In *17th European Conference on Information Systems, ECIS 2009*, Susan Newell, Edgar A. Whitley, Nancy Pouloudi, Jonathan Wareham, and Lars Mathiassen (Eds.). AIS, Verona, Italy, 1138–1149. http://aisel.aisnet.org/ecis2009/388

[48] Hans Oh, Carlos Rizo, Murray Enkin, Alejandro Jadad, John Powell, and Claudia Pagliari. 2005. What Is eHealth (3): A Systematic Review of Published Definitions. *Journal of Medical Internet Research* 7, 1 (Feb. 2005), v7i1e1. https://doi.org/10.2196/jmir.7.1.e1

[49] Kirsten Ostherr, Svetlana Borodina, Rachel Conrad Bracken, Charles Lotterman, Eliot Storer, and Brandon Williams. 2017. Trust and privacy in the context of user-generated health data. *Big Data & Society* 4, 1 (June 2017), 205395171770467. https://doi.org/10.1177/2053951717704673

[50] Claudia Pagliari, David Sloan, Peter Gregor, Frank Sullivan, Don Detmer, James P Kahan, Wija Oortwijn, and Steve MacGillivray. 2005. What Is eHealth (4): A Scoping Exercise to Map the Field. *Journal of Medical Internet Research* 7, 1 (March 2005), e9. https://doi.org/10.2196/jmir.7.1.e9

[51] Stefan Palan and Christian Schitter. 2018. Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance* 17 (March 2018), 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

[52] European Parliament and Council of European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). http://data.europa.eu/eli/reg/2016/679/oj Publication Title: OJ, Volume: L 119, p. 1–88.

[53] Wei Peng, Shaheen Kanthawala, Shupei Yuan, and Syed Ali Hussain. 2016. A qualitative study of user perceptions of mobile health apps. *BMC Public Health* 16, 1 (Dec. 2016), 1158. https://doi.org/10.1186/s12889-016-3808-0

[54] Jamie Pinchot and Donna Cellante. 2021. Privacy Concerns and Data Sharing Habits of Personal Fitness Information Collected via Activity Trackers. *JISAR* 14, 2 (June 2021), 4. https://jisar.org/2021-14/n2/JISARv14n2p4.html

[55] W. Nicholson Price, Margot E. Kaminski, Timo Minssen, and Kayte Spector-Bagdady. 2019. Shadow health records meet new data privacy laws. *Science* 363, 6426 (Feb. 2019), 448–450. https://doi.org/10.1126/science.aav5133

[56] Kavous Salehzadeh Niksirat, Lev Velykoivanenko, Noé Zufferey, Mauro Cherubini, Kévin Huguenin, and Mathias Humbert. 2024. Wearable Activity Trackers: A Survey on Utility, Privacy, and Security. *Comput. Surveys* 56, 7 (July 2024), 1–40. https://doi.org/10.1145/3645091

[57] Saskia C. Sanderson, Kyle B. Brothers, Nathaniel D. Mercaldo, Ellen Wright Clayton, Armand H. Matheny Antommaria, Sharon A. Aufox, Murray H. Brilliant, Diego Campos, David S. Carrell, John Connolly, Pat Conway, Stephanie M. Fullerton, Nanibaa' A. Garrison, Carol R. Horowitz, Gail P. Jarvik, David Kaufman, Terrie E. Kitchner, Rongling Li, Evette J. Ludman, Catherine A. McCarty, Jennifer B. McCormick, Valerie D. McManus, Melanie F. Myers, Aaron Scrol, Janet L. Williams, Martha J. Shrubsole, Jonathan S. Schildcrout, Maureen E. Smith, and Ingrid A. Holm. 2017. Public Attitudes toward Consent and Data Sharing in Biobank Research: A Large Multi-site Experimental Survey in the US. *The American Journal of Human Genetics* 100, 3 (March 2017), 414–427. https://doi.org/10.1016/j.ajhg.2017.01.021

[58] Tanja Schroeder, Maximilian Haug, and Heiko Gewald. 2022. Data Privacy Concerns Using mHealth Apps and Smart Speakers: Comparative Interview Study Among Mature Adults. *JMIR Formative Research* 6, 6 (June 2022), e28025. https://doi.org/10.2196/28025

[59] Raymond Scupin. 1967. The KJ Method: A Technique for Analyzing Data Derived from Japanese Ethnology. *Human Organization* 56, 2 (1967), 233–237. https://www.jstor.org/stable/44126786

[60] T. G. Smith, M. E. Dunn, K. Y. Levin, S. P. Tsakraklides, S. A. Mitchell, L. V. Van De Poll-Franse, K. C. Ward, C. L. Wiggins, X. C. Wu, M. Hurlbert, and N. K. Aaronson. 2019. Cancer survivor perspectives on sharing patient-generated health data with central cancer registries. *Quality of Life Research* 28, 11 (Nov. 2019), 2957–2967. https://doi.org/10.1007/s11136-019-02263-0

[61] Hiral Soni, Adela Grando, Anita Murcko, Sabrina Diaz, Madhumita Mukundan, Nassim Idouraine, George Karway, Michael Todd, Darwyn Chern, Christy Dye, and Mary Jo Whitfield. 2020. State of the art and a mixed-method personalized approach to assess patient perceptions on medical record sharing and sensitivity. *Journal of Biomedical Informatics* 101 (Jan. 2020), 103338. https://doi.org/10.1016/j.jbi.2019.103338

[62] Katta Spiel, Oliver L. Haimson, and Danielle Lottridge. 2019. How to do better with gender on surveys: a guide for HCI researchers. *Interactions* 26, 4 (June 2019), 62–65. https://doi.org/10.1145/3338283 Place: New York, NY, USA Publisher: Association for Computing Machinery.

[63] Fritz Strack. 1992. "Order Effects" in Survey Research: Activation and Information Functions of Preceding Questions. In *Context Effects in Social and Psychological Research*, Norbert Schwarz and Seymour Sudman (Eds.). Springer, New York, NY, 23–34. https://doi.org/10.1007/978-1-4612-2848-6_3

[64] Shiwei Sun, Jin Zhang, Yiwei Zhu, Mian Jiang, and Shuhui Chen. 2022. Exploring users' willingness to disclose personal information in online healthcare communities: The role of satisfaction. *Technological Forecasting and Social Change* 178 (May 2022), 121596. https://doi.org/10.1016/j.techfore.2022.121596

[65] Holly K. Tabor, Jacquie Stock, Tracy Brazg, Margaret J. McMillin, Karin M. Dent, Joon-Ho Yu, Jay Shendure, and Michael J. Bamshad. 2012. Informed consent for whole genome sequencing: A qualitative analysis of participant expectations and perceptions of risks, benefits, and harms. *American Journal of Medical Genetics Part A* 158A, 6 (June 2012), 1310–1319. https://doi.org/10.1002/ajmg.a.35328

[66] Sanjit Thapa, Abubakar Bello, Alana Maurushat, and Farnaz Farid. 2023. Security Risks and User Perception towards Adopting Wearable Internet of Medical Things. *International Journal of Environmental Research and Public Health* 20, 8 (April 2023), 5519. https://doi.org/10.3390/ijerph20085519

[67] M. Grace Trinidad, Jodyn Platt, and Sharon L. R. Kardia. 2020. The public's comfort with sharing health data with third-party commercial companies. *Humanities and Social Sciences Communications* 7, 1 (Nov. 2020), 149. https://doi.org/10.1057/s41599-020-00641-5

[68] The Pew Charitable Trusts. 2021. Most Americans Want to Share and Access More Digital Health Data. https://pew.org/3B9DJXq

[69] Lev Velykoivanenko, Kavous Salehzadeh Niksirat, Noé Zufferey, Mathias Humbert, Kévin Huguenin, and Mauro Cherubini. 2021. Are Those Steps Worth Your Privacy?: Fitness-Tracker Users' Perceptions of Privacy and Utility. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (Dec. 2021), 1–41. https://doi.org/10.1145/3494960

[70] Miranda E. Vidgen, Sid Kaladharan, Eva Malacova, Cameron Hurst, and Nicola Waddell. 2020. Sharing genomic data from clinical testing with researchers: public survey of expectations of clinical genomic data management in Queensland, Australia. *BMC Medical Ethics* 21, 1 (Dec. 2020), 119. https://doi.org/10.1186/s12910-020-00563-6

[71] VanAnh Vo, Lola Auroy, and Aline Sarradon-Eck. 2019. Patients' Perceptions of mHealth Apps: Meta-Ethnographic Review of Qualitative Studies. *JMIR Mhealth Uhealth* 7, 7 (July 2019), e13817. https://doi.org/10.2196/13817

[72] Jake Weidman, William Aurite, and Jens Grossklags. 2019. On Sharing Intentions, and Personal and Interdependent Privacy Considerations for Genetic Data: A Vignette Study. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 16, 4 (July 2019), 1349–1361. https://doi.org/10.1109/TCBB.2018.2854785

[73] Elissa R Weitzman, Skyler Kelemen, Liljana Kaci, and Kenneth D Mandl. 2012. Willingness to share personal health record data for care improvement and public health: a survey of experienced personal health record users. *BMC Medical Informatics and Decision Making* 12, 1 (Dec. 2012), 39. https://doi.org/10.1186/1472-6947-12-39

[74] Paul Wicks, Michael Massagli, Jeana Frost, Catherine Brownstein, Sally Okun, Timothy Vaughan, Richard Bradley, and James Heywood. 2010. Sharing Health Data for Better Outcomes on PatientsLikeMe. *Journal of Medical Internet Research* 12, 2 (June 2010), e19. https://doi.org/10.2196/jmir.1549

[75] Alan Yang and Upkar Varshney. 2022. Mobile health evaluation: Taxonomy development and cluster analysis. *Healthcare Analytics* 2 (Nov. 2022), 100022. https://doi.org/10.1016/j.health.2022.100022

[76] Zhiping Zhang, Michelle Jia, Hao-Ping (Hank) Lee, Bingsheng Yao, Sauvik Das, Ada Lerner, Dakuo Wang, and Tianshi Li. 2024. "It's a Fair Game", or Is It? Examining How Users Navigate Disclosure Risks and Benefits When Using LLM-Based Conversational Agents. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–26. https://doi.org/10.1145/3613904.3642385

[77] Noé Zufferey, Kavous Salehzadeh Niksirat, Mathias Humbert, and Kévin Huguenin. 2023. "Revoked just now!" Users' Behaviors Toward Fitness-Data Sharing with Third-Party Applications. *Proceedings on Privacy Enhancing Technologies* 2023, 1 (Jan. 2023), 47–67. https://doi.org/10.56553/popets-2023-0004

A Examples of *shadow* HRD

Table 1: Examples of *shadow* HRD. Only non-obvious cases are included. More self-explanatory categories, like "general-purpose chatbots" or "reminder to take medication," are not further elaborated.

| Category	Sub-category	Example
	Fitness activities planning	setting reminders for exercise routines
	Reminder to buy health-related items	reminder to buy medication, vitamins, medical aids
	Healthcare provider's contact in Contacts	saving medical specialty of the physician in Contacts
	Shopping list including food/health-related items in Notes	entering the list of medicine to buy in Notes
	Audio streaming health-related content	listening to a health-related podcast on Spotify
	Cognitive games	playing Lumosity or Elevate to enhance memory
	Fitness games	playing Pokémon GO on the phone or Beat Saber on VR to increase physical activity
	Providing health-related info. for restaurant reservation	providing food allergies or dietary restrictions when booking restaurant online
	Use of loyalty cards in store when buying food/health-related items	using the loyalty card of pharmacy when buying medicine
	Health-related cards in digital wallet	having health insurance or organ donation card in Apple Wallet
	Providing health-related information to get/save money	filling out surveys with health-related questions when applying for bank loans
	Using browsers	having health-related websites in search history
	Scientific search engines	using Google Scholar to search health-related papers
	Using image-based search engines	using Google Lens to identify visual symptoms
	Health-related information websites	using websites like WebMD to read health-related advice
	Geolocation turned on at health-related location	visiting a hospital without turning off the GPS on the phone
	Health-related audio/video call on online services	using WhatsApp to make video calls with a family doctor
	Posting health-related content	posting a photo while at the hospital on Instagram
	Consuming health-related content	following a rare disease-specific social media account
	Health-related social networks	using PatientsLikeMe to connect with others who have similar health conditions
	Documents for self-mangement of health	using Excel files to log medical test results for comparison over the years
	Photo of visible medical condition	using the phone camera to take pictures of injured parts of the body that cannot be seen easily (e.g., the scalp)
	Journaling health-related entries	using Daylio (a diary app) to record symptoms
	Providing health-related information in job context	filling out surveys with health-related questions when applying for job
	Translation software to communicate with healthcare provider	using DeepL to accurately convey symptoms when traveling in a foreign country
	Using biometric data	using biometric authentication to unlock the phone
	Accessing fitness/health-related facilities	using a smart card to access the gym facilities
	Utility apps for health-related purposes	using Clock app on the smartphone to set regular bedtime and wake-up alarms

B   Additional statistics and results
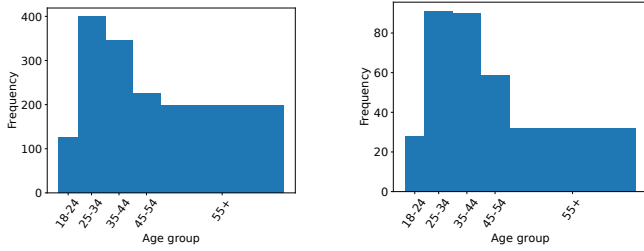
(a) Distribution of age of respondents to the screener survey, prior to filtering based on screener survey responses.

(b) Distribution of age of respondents to the main survey, after filtering based on screener survey responses.

Figure 9: Comparison of the age distribution before and after the screener. The distributions are similar, suggesting that our screener did not lead to the exclusion of any particular age groups from the analysis.
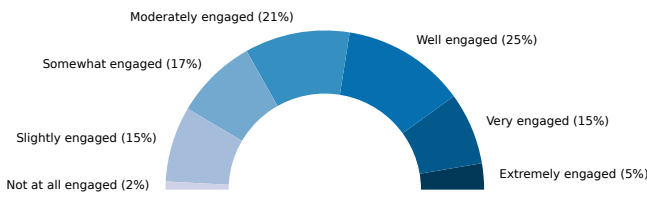
Figure 10: The level of self-reported engagement in health-related activities

C   Survey Transcript

Table 2: Survey Transcript Summary

Survey sections	Question numbers	Section description
Sec. A	Q1, Q2	Introduction
Sec. B	Q3	User Awareness of HRD
Sec. C ⤨		User Behavior Regarding HRD
Subsec. 1	Q4, Q5, Q6, Q7	✉ Communication
Subsec. 2	Q8, Q9	🗺 Maps/Navigation
Subsec. 3	Q10, Q11, Q12, Q13	🗓 Productivity/Organization
Subsec. 4	Q14, Q15, Q16	🛒 Shopping/Finance
Subsec. 5	Q17, Q18, Q19	🗎 File/Multimedia
Subsec. 6	Q20, Q21, Q22, Q23	🔍 Information Seeking
Subsec. 7	Q24, Q25, Q26	👍 Social Networks/Forums
Sec. D	Q27, Q28, Q29	General Privacy Aspects
Sec. E	Q30, Q31, Q32	Health-related Privacy Concerns
Sec. F	Q33, Q34	Demographics

Note: Coding rules are colored in gray (not visible to respondents)

Sec. A.   Consent

STUDY

You are invited to participate in a study on behaviors, concerns, and awareness of users regarding health-related data sharing. You have been invited to this survey because, a few days ago, you filled out a pre-screener questionnaire on Prolific, matching our research criteria. We greatly appreciate your participation!

This study is conducted and financed by the Information Security and Privacy lab of Prof. Kévin Huguenin and the Cybersecurity lab of Prof. Mathias Humbert at the University of Lausanne (UNIL), Switzerland.

OUTLINE OF THE STUDY PROCEDURE

If you consent, we ask you to complete a questionnaire. The questionnaire consists of about 30 questions regarding your digital habits, behaviors, and concerns when dealing with health-related data in your day-to-day life. You will need about 15 minutes to complete the questionnaire.

REMUNERATION

At the end of the study, you will be awarded £2.25 ($2.84) for your participation

CONFIDENTIALITY AND DATA PROCESSING

Your answers will be recorded in a confidential and secure way. They will only be accessible to the researchers and authorized personnel from the University of Lausanne and ETH Zurich.

In the case where the answers are shared with the scientific community to promote open science (open data), they will be anonymized and/or aggregated.

YOUR RIGHTS

Your participation in this study is entirely voluntary. You have the right to refuse to participate or to withdraw from the study at any time. If you withdraw from the study, your data will be deleted and you will not be remunerated.

If you have any questions about the study, please feel free to contact the research team using the Prolific messaging service.

CONSENT

By giving your consent, you acknowledge that you are at least 18 years old. You also acknowledge that you have read the above information and that you agree to it.

◯ I consent
◯ I do not consent

What is your Prolific ID?
Please note that this response should auto-fill with the correct ID
pre-filled open text field

Q1. How would you describe your level of engagement in health-related activities? (e.g., having regular health checkups, engaging in wellness activities, using health management technologies such as health or fitness trackers / mobile apps)

◯ Not at all engaged
◯ Slightly engaged
◯ Somewhat engaged
◯ Moderately engaged
◯ Well engaged
◯ Very engaged
◯ Extremely engaged

Q2. Do you typically use apps or services that belong to one of the following categories? Select all that apply.
Please note that this question is identical to one that was asked in the pre-screener that you answered a few days ago.

☐ Productivity and organization (Notes, Calendar, Reminder, Contacts)
☐ Shopping (e.g., online stores) or finance (e.g., banking, payment, wallet)
☐ Search engines (e.g., Google), chatbots (e.g., ChatGPT), or streaming (e.g., YouTube, Spotify)
☐ Social networks (e.g., Instagram) or forums (e.g., Reddit)
☐ Communication (e.g., phone, SMS, Messenger, WhatsApp, Zoom)
☐ Maps and navigation (e.g., Google / Apple Maps, CityMapper, Moovit, Transit, Waze)
☐ Files and multimedia (e.g., PDF documents saved on your computer, photos / videos / audio recordings saved on your phone)
☐ None of the above

Sec. B.   General awareness question

Q3. In your opinion, from "Extremely unlikely" to "Extremely likely", how likely are each of these types of apps or services to contain information about your health?

– Productivity and organization (Notes, Calendar, Reminder, Contacts)
– Shopping (e.g., online stores) or finance (e.g., banking, payment, wallet)
– Search engines (e.g., Google), chatbots (e.g., ChatGPT), or streaming (e.g., YouTube, Spotify)
– Social networks (e.g., Instagram) or forums (e.g., Reddit)
– Communication (e.g., phone, SMS, Messenger, WhatsApp, Zoom)

– Maps and navigation (e.g., Google / Apple Maps, CityMapper, Moovit, Transit, Waze)
– Files and multimedia (e.g., PDF documents saved on your computer, photos / videos / audio recordings saved on your phone)

○ Extremely unlikely
○ Moderately unlikely
○ Slightly unlikely
○ Neither likely nor unlikely
○ Slightly likely
○ Moderately likely
○ Extremely likely

Sec. C. User Behavior Regarding HRD

Next, you will be presented with 7 short blocks of questions, each relating to one of the 7 categories of apps and services described in the previous questions.

These 7 short blocks will be followed by 2 short blocks of general questions, not relating to any specific category of apps and services, as well as 1 block of demographics questions.

Subsec. 1. Communication

The following block of questions revolves around your use of communication apps and services for health-related purposes.

Q4. Which of the following do you typically do? Select all that apply.
☐ Sending / receiving an e-mail including health-related information (e.g., to / from health insurance company, healthcare provider, loved one)
☐ Interacting with a healthcare professional through a text / voice / video communication service (e.g., phone call, Messenger, WhatsApp, Zoom)
☐ None of the above

Q5. Which e-mail service(s) do you / would you use to communicate with healthcare professionals? Please list all of them, one per line.
open text field

Q6. Which messaging service(s) do you / would you use to communicate with healthcare professionals? Please include both general-purpose messaging services (e.g., WhatsApp, Messenger) and services specific to your healthcare center(s) or clinic(s). Please list all of them, one per line.
open text field

Q7. Which of the following measures do you typically take to protect the confidentiality of your health-related communications? Select all that apply.
If a measure is not applicable to you (e.g., if you do not interact with healthcare providers using communication apps or services), please select it if you would use it in the future.
* End-to-end encryption (E2EE) is a type of messaging that keeps the content of the conversation private from everyone, including the messaging service.
☐ Ensuring health-related data that you share on messaging services is end-to-end encrypted *
☐ Using disappearing messages (e.g., in Messenger / WhatsApp) for health-related communications
☐ Ensuring any health-related data that you share via e-mail is end-to-end encrypted *
☐ Ensuring audio/video communications with a healthcare provider are end-to-end encrypted *
☐ Using a pseudonymous (i.e., fake) secondary e-mail address or account for health-related communications
☐ None of the above

Subsec. 2. Maps/Navigation

The following block of questions revolves around your use of maps and navigation apps and services for health-related purposes.

Q8. When going to a health-related appointment, which of the following do you typically do, either before or during the journey? Select all that apply.
☐ Using a connected digital map or transit / navigation service to get to the appointment (e.g., Google/Apple Maps, Maps, CityMapper, Moovit, Transit, Waze)
☐ Using a taxi app to get to the appointment (e.g., Uber, Lyft)
☐ None of the above

Q9. Which of the following measures do you typically take to protect your privacy when going to a health-related appointment (e.g., medical appointment)? Select all that apply.
If a measure is not applicable to you (e.g., if you do not use navigation apps or services to get to health-related appointments), please select it if you would use it, if needed in the future.
☐ Turning off your location on your phone
☐ Setting the destination in the navigation service to a nearby location instead of the exact appointment address
☐ Avoiding navigation and maps services altogether
☐ Please select this option to show that you are paying attention

☐ Using private / incognito mode in the maps or navigation apps or services
☐ Deleting your location history after the appointment
☐ Using maps in offline mode
☐ None of the above

Subsec. 3. Productivity/Organization

The following block of questions revolves around your use of productivity and organization apps and services for health-related purposes.

Q10. Which of the following do you typically do? Select all that apply.
☐ Saving a healthcare provider's contact details in a digital address book (e.g., phone numbers, address)
☐ Adding a reminder or calendar entry for a health-related appointment
☐ Writing health-related information (e.g., condition names, symptoms, medication names, questions) in a digital note (e.g., in Notes app)
☐ None of the above

Q11. Which of these details do you typically include when adding a reminder or calendar entry for a health-related appointment? Select all that apply.
If you do not add reminders or calendar entries for health-related appointments, please select the details you would include, if you were to do so in the future.
☐ Name
☐ Phone number
☐ E-mail address
☐ Street address
☐ Medical specialty (e.g., nephrologist)
☐ Notes relating to health condition (e.g., condition names, symptoms, medication names, questions)
☐ None of the above

Q12. Which of these details do you typically include when saving the contact of a healthcare provider in your digital address book? Select all that apply.
If you do not save the contacts of healthcare providers, please select the details you would include, if you were to do so in the future.
☐ Name
☐ Phone number
☐ E-mail address
☐ Street address
☐ Medical specialty (e.g., nephrologist)
☐ Notes relating to health condition (e.g., condition names, symptoms, medication names, questions)
☐ None of the above

Q13. Do you purposefully do any of the following, for privacy reasons?
- Refrain from saving the contact of a healthcare provider in your digital address book
- Refrain from adding a reminder or calendar entry for a health-related appointment Omit / conceal information when doing either of the above?
○ Never
○ Rarely
○ Occasionally
○ Sometimes
○ Often
○ Very often
○ Always

Subsec. 4. Shopping/Finance

The following block of questions revolves around your use of shopping and finance apps and services for health-related purposes.

Q14. Which of the following do you typically do? Select all that apply.
* Health-related items refer to any items used for preventing, diagnosing, treating, or managing symptoms of a physical / mental health condition (e.g., diabetes, anxiety) or natural physiological process (e.g., menstruation, pregnancy). They do not include basic hygiene products such as soap or shampoo.
☐ Buying health-related items * (e.g., medication, vitamins, medical aids) on online shops
☐ Inputting/labelling health-related expenses in a budgeting or banking online service
☐ Storing health-related cards (e.g., health insurance card) in your digital wallet (e.g., Apple / Google Wallet)
☐ Adding health-related items * (e.g., medication, vitamins, medical aids) in a digital grocery list
☐ Buying health-related items * (e.g., medication, vitamins, medical aids) in physical stores using your debit / credit card or electronic payment methods (e.g., Paypal, Venmo, Cash App, Apple / Google Pay)
☐ None of the above

Q15. Which of the following measures do you typically take to protect the confidentiality of your health-related purchases? Select all that apply.
If a measure is not applicable to you (e.g., if you do not make health-related online purchases), please select it if you would use it, if needed in the future.

☐ Using a pseudonymous (i.e., fake) account on online shops when making health-related online purchases

☐ Using a virtual card (e.g., Privacy) or temporary card number to hide your personal information when paying for health-related online purchases

☐ Using a different shipping address than yours (e.g., pickup point) when making health-related online purchases

☐ Excluding health-related purchases from your public purchase history on online shops (e.g., on Amazon)

☐ Turning off unnecessary cookies when making health-related online purchases

☐ Please select this option to show that you are paying attention

☐ Using private browsing or incognito mode when making health-related purchases online

☐ Adding health-related items to your digital grocery list without explicitly stating the product type or name (e.g., placeholder, pseudonym)

☐ None of the above

Q16. Do you purposefully refrain from any of the actions stated in the question above (including both the ones you selected and those you did not), for privacy reasons?
○ Never
○ Rarely
○ Occasionally
○ Sometimes
○ Often
○ Very often
○ Always

Subsec. 5. File/Multimedia

The following block of questions revolves around your use of files and multimedia for health-related purposes.

Q17. Which of these types of files or media do you have on your devices (e.g., mobile phone, computer)?
☐ Photo or video of a health-related issue (e.g., skin issue, eye issue), captured by yourself, a friend, or a relative
☐ Digitized or original health-related documents, produced by a healthcare provider (e.g., blood test results, echocardiogram, ultrasound)
☐ Audio recording of a health-related issue (e.g., cough), captured by yourself, a friend, or a relative
☐ Documents (not mobile app) to self-track health-related information (e.g., spreadsheet with your weight)
☐ None of the above

Q18. Which of the following measures do you typically use to protect the confidentiality of your health-related files? Select all that apply.
If a measure is not applicable to you (e.g., if you do not have health-related files on your devices), please select it if you would use it, if needed in the future.
☐ Using a password for your devices
☐ Storing health-related media and files in a dedicated password-protected secure folder
☐ Encrypting health-related media and files
☐ Not synchronizing health-related media and files to the cloud
☐ Anonymizing health-related media or files by excluding or blurring identifying information (e.g., face, personal information)
☐ Please select this option to show that you are paying attention
☐ Carefully reviewing app permission requests before allowing access to your phone's or computer's file storage or gallery.
☐ Not tagging health-related media or using pseudo-tags that are unrelated to health
☐ None of the above

Q19. How often do you tag, give meaningful names to, or categorize (e.g., sort into specific folders) health-related files and media?
○ Never
○ Rarely
○ Occasionally
○ Sometimes
○ Often
○ Very often
○ Always

Subsec. 6. Information Seeking

The following block of questions revolves around your use of different technologies for looking up health-related information.

Q20. Which of the following do you typically use to look up symptoms or health-related information or to better communicate with healthcare providers? Select all that apply.
☐ A general-purpose search engine (e.g., Google)
☐ A chatbot (e.g., ChatGPT)

☐ A health information service website (e.g., WebMD, Medline)
☐ A general information service website (e.g., Wikipedia)
☐ A general information videos or podcasts website (e.g., YouTube, Spotify)
☐ An online translation tool (e.g., Google translate, DeepL) or writing assistant (e.g., Grammarly)
☐ None of the above
☐ An AI-based image recognition service (e.g., Google Lens for identifying visual symptoms)

Q21. When looking up your symptoms or information related to your health online, which browser(s) do you / would you use? Select all that apply.
☐ Chrome / Chromium
☐ Safari
☐ Edge
☐ Firefox
☐ Samsung Internet
☐ Opera
☐ Avast browser
☐ Brave
☐ Bromite
☐ DuckDuckGo
☐ Tor browser
☐ Ecosia
☐ Epic
☐ I2P
☐ Vivaldi
☐ Other(s), please specify (one browser per line) *open text field*

Q22. When looking up your symptoms or information related to your health online, which of the following measures do you typically use to protect your privacy? Select all that apply.
If a measure is not applicable to you (e.g., if you do look up your symptoms online), please select it if you would use it, if needed in the future.
☐ Using a VPN or an anonymity network (e.g., TOR)
☐ Using Incognito mode
☐ Deleting your browsing history afterwards
☐ Refusing unnecessary cookies
☐ Using a privacy-focused browser (e.g., TOR Browser)
☐ Using a privacy-focused service (e.g., DuckDuckGo, ChatBot running locally)
☐ None of the above

Q23. Do you purposefully refrain from inputting health-related information and keywords in one of the apps or services listed in the question above, due to privacy reasons?
○ Never
○ Rarely
○ Occasionally
○ Sometimes
○ Often
○ Very often
○ Always

Subsec. 7. Social Networks/Forums

The following block of questions revolves around your use of social networks for health-related purposes.

Q24. Which of the following do you typically do? Select all that apply.
☐ Sharing data from health-related apps and services (e.g., fitness) on social networks (e.g., Instagram)
☐ Making a post or status update revealing explicitly a health condition that you have
☐ Having a social media account or blog dedicated to a health condition that you have
☐ Following social media accounts dedicated to a health condition that you have
☐ Engaging in forums or support groups related to a health condition that you have (e.g., Reddit, PatientsLikeMe)
☐ None of the above

Q25. Have you ever used any of the following anonymization techniques to protect your privacy, when either engaging in health-related forums / support groups or making a health-related post or status update on a social media account?
- Using a pseudonym
- Face blurring
If a measure is not applicable to you (e.g., if you do not engage in health-related support groups), please select it if you would use it, if needed in the future.
○ Never
○ Rarely
○ Occasionally

○ Sometimes
○ Often
○ Very often
○ Always

Q26. Do you purposefully refrain from any of the actions stated in the question above (including both the ones you selected and those you did not), for privacy reasons?
○ Never
○ Rarely
○ Occasionally
○ Sometimes
○ Often
○ Very often
○ Always

Sec. D. General (cloud, interdependent privacy, concern)

Q27. From "Not at all concerned" to "Extremely concerned", how concerned would you be if each of these entities had access to your health-related data?
– Device manufacturers (e.g., Apple, Samsung, Garmin)
– Online service / cloud providers or OS / app developers (e.g., Facebook, Google, Dropbox, Android, Apple, WhatsApp)
– Internet service provider (home and mobile)
– Hackers
– Governmental institutions
– Health insurance companies
– Strangers on the internet
– Healthcare providers
– Advertisers and marketers
– Employer (including future employer(s))
– Co-workers
– Acquaintances
– Close friends or relatives
– Intimate partners

○ Not at all concerned
○ Slightly concerned
○ Somewhat concerned
○ Moderately concerned
○ Concerned
○ Very concerned
○ Extremely concerned

Q28. Which of the following personal data is synced to your cloud account(s)? Select all that apply.
Below are standard instructions to check what is synced on iOS and Android. Please also check and select what is synced to any other cloud storage accounts you have (e.g., Dropbox, OneDrive)
<u>Instructions for iOS users</u>
1. Go to Settings > your name.
2. Tap iCloud
3. Tap Show All to see the full list
4. Check for which of the options sync is enabled ("On" or enabled toggle button)
5. Go back to Settings
6. Go to Mail > Accounts
7. For each of the listed accounts, except iCloud, click on the account (e.g., Gmail, Exchange)
8. Check for which of the options sync is enabled ("On" or enabled toggle button)
<u>Instructions for Android users</u>
1. Go to Settings > Accounts and Backup
2. Under Google Drive, tap "Back up data" > "Google Account data"
3. Check if the sync is enabled for Photos & videos ("On")
4. Tap Google Account data
5. Check for which of the options sync is enabled (enabled toggle button)
6. Additional instructions for Samsung users:
a. Go back to Settings > Accounts and Backup
b. Under Samsung Cloud, tap "Back up data"
c. Check for which of the options sync is enabled (enabled toggle button)
☐ Calendar
☐ Contacts
☐ Gallery (photos, videos)
☐ Notes
☐ Reminders
☐ Messages
☐ Files
☐ Browsing history and/or bookmarks
☐ Passwords
☐ Wallet

☐ Other apps data
☐ None of the above

Q29. Previous questions have covered behaviors that may leak your health-related data. To what extent do you engage in similar behaviors for the health of others (e.g., significant other, child, parent)?
E.g., saving the contact details of a child's pediatrician in your digital address book, or storing the test results of a parent on your mobile phone
○ Never
○ Rarely
○ Occasionally
○ Sometimes
○ Often
○ Very often
○ Always

Sec. E. General - standard scale

Q30. I believe that submitting health information on the Internet is:
○ 1 - Not advisable at all
○ 2
○ 3
○ 4
○ 5
○ 6
○ 7
○ 8
○ 9
○ 10 - Highly advisable

Q31. Health information on the Internet, once submitted:
○ 1 - Will not be abused at all
○ 2
○ 3
○ 4
○ 5
○ 6
○ 7
○ 8
○ 9
○ 10 - Will be abused for sure

Q32. Health information on the Internet, once submitted:
○ 1 - Will not be compromised at all
○ 2
○ 3
○ 4
○ 5
○ 6
○ 7
○ 8
○ 9
○ 10 - Could be shared or sold to others

Sec. F. Demographics and follow-up

Q33. Which gender(s) do you identify with?
☐ Woman
☐ Man
☐ Non-binary
☐ Prefer to self-describe *open text field*
☐ Prefer not to disclose

Q34. What is your field of work / study?
○ Business administration / management
○ Finance
○ IT / Computer Science
○ Marketing / Advertising
○ Environmental / Agricultural
○ Science / Mathematics
○ Building / Construction
○ Arts / Design
○ Beauty
○ Education
○ Health / Healthcare
○ Hospitality / Tourism
○ Retail / Customer service
○ Transport / Logistics
○ Manufacturing
○ Law
○ Other, please specify *open text field*
○ Prefer not to say