User Attitudes On Direct-to-Consumer Genetic Testing

Debjani Saha, Anna Chan, Brook Stacy, Kiran Javkar, Sushant Patkar*, and Michelle L. Mazurek
*University of Maryland, *University of Maryland and National Cancer Institute, NIH*
{*dsaha, chana532, bstacy, kjavkar, spatkar, mmazurek*}*@umd.edu*

***Abstract*—Advances in biotechnology now allow users to obtain their genetic information, including ancestry and predisposition to various diseases and health issues, with relative ease. With these new commercial services come a host of privacy concerns with respect to data sharing and access. User data is being sold to third parties, including pharmaceutical and biotechnology companies, and may be accessed by law enforcement in accordance with proper legal procedures. Moreover, many users of these services go on to deposit the data they obtain into online, public repositories that are fully accessible to anyone with an internet connection. The full extent of the risks they face may not be apparent to users. This paper reports on a semi-structured interview study (n=24) examining user concerns regarding these tests, what information they believe they are revealing, and what they think companies are doing with their data. We find that users are concerned with privacy, and understand at a basic level the nature of the data they are revealing. However, their privacy concerns are often insufficient to deter them from taking such a test, and many have difficulty grasping some of the implications of sharing their genetic information with commercial entities.**

***Index Terms*—DTC genetic testing, human factors, privacy, semi-structured interviews**

1. Introduction

Direct-to-consumer (DTC) genetic testing has emerged as a popular service allowing consumers to learn about their genetic information. Services like 23andMe and AncestryDNA partially sequence customers' DNA and return results about their heritage, susceptibility to genetic disease, and other personal information, allowing the general public to obtain information about and from their genomes.

However, the practice of DTC genetic testing raises severe privacy concerns. An individual's DNA contains a wealth of deeply personal information, not only about them, but also their close relatives, and genetic information has proven difficult if not impossible to anonymize [1]–[7].

Regulations on the handling and use of genetic data in a commercial setting are limited, and the potential for abuse or mishandling of this sensitive data is high. In fact, both 23andMe and AncestryDNA have provided information to third parties for biomedical research [8], [9], and genetic information has been subpoenaed by law enforcement as part of criminal investigations [10], [11]. Further, genetic data stored at these commercial entities may be subject to data breaches, insider attacks, and other common threats.

Despite these documented issues, DTC genetic tests continue to grow in popularity. As of the beginning of 2019, an estimated 26 million people had used a DTC genetic test, with a projected growth to 100 million within 2 years [12].

Our contribution. Thus far, little is known about how consumers understand and evaluate the potential risks of taking DTC genetic tests. In this work, we aim to address this gap by evaluating user awareness of the personal information revealed through consumer DNA tests, as well as user knowledge and concern about the associated privacy risks. We do this through a series of semi-structured interviews (n=24), asking about participant experiences with DTC genetic tests, the information they believe they are revealing, and their privacy concerns (or lack thereof). We address the following research questions:

RQ1. What information do users believe is revealed by their genetic data?

RQ2. What concerns, if any, do users have with respect to DTC genetic testing?

RQ3. How do users' concerns influence their decisions to participate in DTC genetic testing?

RQ4. How do users believe their genetic information is used by DTC genetic testing companies?

We find that while our participants are aware in general of potential privacy concerns related to DTC genetic testing, they do not necessarily understand subtle nuances of what their genetic data can reveal. Further, in many cases participants do not see these concerns as relevant to themselves personally, expressing resignation that violation of their privacy is inevitable whether they take a DTC genetic test or not. Participants express comfort with data sharing "for the public good," such as for biomedical research, as long as precautions are taken to protect the user whose data is being shared. Overall, participants tend to believe that the benefits from using DTC genetic testing outweigh the risks, but express a desire for more transparency and control in the process. We discuss how these results do (not) mirror privacy concerns in other contexts and provide some recommendations for addressing users' concerns as well as key gaps in their understanding.

2. Background and Related Work

In this section, we first provide background on DTC genetic testing and associated areas of concern. We next discuss previous work that aims to understand user perspectives on DTC genetic testing. Finally, we briefly review work exploring user privacy attitudes related to other kinds of health data.

Genome sequencing and genetic testing. An individual's

DOI 10.1109/EuroSP48549.2020.00016

genetics can provide an abundance of information about their health, heritage, and more. In his survey paper, Lander describes the major impacts of the sequencing of the human genome, as well as areas for future work [13]. Moreover, technologies enabling long-term sample preservation at room temperatures [14] and low-cost sequencing methods [15] have brought down the overall cost of DTC genetic testing sufficiently to support a mass market. Services such as 23andMe and AncestryDNA serve this niche. These commercial services typically do not do whole-genome sequencing, but rather genotyping, i.e. focusing on particular loci known to contain markers for specific traits such as disease and ancestry [16].

Uniquely identifying information. DNA contains inherently personal, uniquely identifying information that cannot currently be effectively anonymized. Multiple studies have demonstrated the ease of re-identification of genetic and genetic-like data in otherwise anonymized datasets [1]–[7].

Importantly, because closely related individuals share large portions of their genomes [17], re-identification of genetic information affects family members as well. Erlich, et al. claim that to produce a familial match of a third cousin or closer relative, the underlying genetic database need only cover 2% of the targeted population [4]. As the number of people who have taken genetic tests increases, the accuracy of re-identification attacks increases accordingly [12]. Ayday discusses privacy threats associated with familial inference as well as possible cryptographic solutions for handling and sharing genomic data securely [18].

Data breaches. Like all companies, those offering direct to consumer genetic services are subject to data breaches and attacks. In 2017, a breach at MyHeritage exposed account credentials for all 92 million of its users [19]. While MyHeritage says no other information was exposed, this does not rule out the possibility of a future attack in which user genetic information is directly accessed, or stolen account information is used to gain access to genetic information.

Policy, regulation, and user protections. In the U.S., regulations protecting commercially-collected genetic data are limited. The Health Insurance Portability and Accountability Act (HIPAA) includes genetic data only in the context of "covered entities" that are normally associated with the dissemination of health care, such as hospitals, doctors, and insurance companies.[1] This means companies like 23andMe and Ancestry are not subject to these guidelines [20]. The Genetic Information Nondiscrimination Act (GINA) prohibits U.S. employers from collecting genetic information of their employees and discriminating against employees based on genetic information [21].

There is, however, no current comprehensive legislation dealing with individuals' genetic privacy. Fendrick identifies the pressing legal and ethical need for privacy law in genetic research [22]. Phillips considers similar legal issues, advising consumers to consider risks such as the exploitation, selling, or sharing of user data with third parties, and the re-identification of individuals based on their genetic data [23], [24].

1. https://www.hhs.gov/hipaa/for-professionals/faq/354/does-hipaa-protect-genetic-information/index.html

Caulfield and McGuire investigate worldwide policy responses to DTC genetic testing [25]. They report that Germany has banned DTC genetic testing, while medical organizations including the Australian Medical Association, the American Medical Association, and the American College of Medical Genetics also support measures to limit or ban DTC testing. In 2010, only 7 of 32 studied DTC companies offered comprehensive consumer privacy policies. They note that in the U.S., DTC companies have been criticized for deceptive advertisements and claims, engaging in unlicensed practice of medicine, and selling medical devices without appropriate regulatory oversight.

Several U.S. government agencies have issued warnings to consumers emphasizing the misleading nature and limited utility of the results obtained from DTC genetic tests [26], [27], as well as warning DTC companies that they may be providing clinical or diagnostic information without required approval [28]. In 2013, 23andMe's health reports were suspended pending FDA approval; they were reinstated in 2017 [29], [30]. As of late 2019, Ancestry also offers health reports [31].

Transparency and alternative uses. Several consumer genetic testing companies share genetic data with third parties in various ways. 23andMe has sold access to its database to multiple pharmaceutical companies [8], [9]. AncestryDNA has partnered with the biotechnology firm Calico to research human longevity [9]. These alternative uses are not clearly laid out for consumers: "23andMe customers have to wade through pages of fine print before finding out that their information may be 'shared with research partners, including commercial partners.' AncestryDNA's contract claim[ed] a 'perpetual, royalty-free, worldwide, transferable license to use your DNA'" [9].

Public databases where users deposit genetic information obtained from DTC testing are also a source of potential concern. Ney et al. found that it is relatively easy to extract and reidentify information from GEDmatch, a popular public database [3]. Sweeney et al. found that participants in the public Personal Genome Project (PGP) were willing to share information because they believed it was anonymous; however, the researchers were able to reidentify 84% of participants [5]. Further, queries conducted on a public database were able to accurately return a long-range familial match 60% of the time [4].

Further, genetic testing companies are often required by law to comply with subpoenas, search warrants, and court orders for information on their users. A suspect in the decades-old Golden State Killer case was apprehended based on genetic information obtained through commercial tests [32]. In 2017, the FBI subpoenaed information on a possible suspect from the genetic testing company FamilyTreeDNA, although ultimately this data was not used to apprehend the suspect. Users of FamilyTreeDNA who do not want their information to be searchable by the FBI or other law enforcement can opt out of familial matching, but this means they can no longer use the database to find relatives, undermining one of the core offerings of this service [10]. There is no similar mitigation for law enforcement access to public databases.

In response to mounting concerns about alternative uses, in 2018 Ancestry, 23andMe, and other companies adopted new guidelines for protecting privacy. These

guidelines pledge greater transparency in sharing practices and require companies to disclose the number of law enforcement requests they get each year. However, adherence to these rules is strictly voluntary [11].

User perspectives on DTC genetic testing. As DTC genetic testing is still relatively new, user perceptions thereof is an emerging area of inquiry. In an early survey of European university students as potential consumers of DTC genetic tests, participants expressed high expectations for accuracy and usefulness, but preferred national health services to commercial offerings [33]. The study discusses ethical questions related to sex selection and health risk prediction, but does not address data privacy.

A somewhat later literature review found limited awareness of DTC testing services, and even more limited work studying people who had actually used them [34]. Primary concerns among those who were aware included data privacy and test reliability. More recently, however, Roberts, et al. find that individuals who have taken a DTC genetic test were overwhelmingly confident in the quality and accuracy of the results [35].

Other more recent work has begun to explore data privacy perceptions. In semi-structured interviews examining whole-genome-sequencing applications, De Cristofaro identified key concerns including trust, how data would be used, and a strong desire for personal control [36]. Other studies have found that perceived benefits of genetic testing outweigh privacy risks. Cheung, et al. found that early adopters of health technologies, including participants in the public Personal Genome Project, tend to be "unconcerned" or "pragmatic" according to Westin's privacy framework [37], [38]. Overall, participants suggest that the public health benefits of contributing genetic information outweigh concerns about privacy, discrimination, and control of personal data. Relatedly, a survey of university students found that desire to contribute to scientific advancement, as well as to learn personal disease risk, was more salient than privacy concerns [39]. We observe a similar valuation of trade-offs; however, we expand on these works by delving further into the unique risks of compromised genetic data, as well as exploring what third party uses participants find acceptable.

Christofides, et al. focus on data policies of personal genetic testing companies by first reviewing these policies, then surveying users about them [40]. They find that while most companies do provide some privacy information, this information largely goes unread; participants who did read the policies often found that they allowed unexpected data uses. We similarly found that users often gloss over privacy notices and have expectations that are at odds with company policy; however, we also found resignation to the idea that companies will use the data in undesirable ways.

Some studies have focused solely on individuals who used a DTC genetic test. Hausermann, et al. found that participants in openSNP, a public genetic repository, were aware of privacy risks but largely viewed privacy as a lost cause [41]. Further, participants felt that people higher on the "social gradient" face less overall risk from privacy breaches and therefore should contribute more to scientific advancement than those with less privilege. Haeusermann notes, however, that this could disadvantage sensitive populations due to unrepresentative data [41]. Our findings — incorporating both people who have taken tests and those who have only considered it — may corroborate this notion: we find some evidence to suggest that participants from minority populations are more likely to be concerned with institutionalized harm resulting from access to personal genetic information.

In concurrent work, Baig et al. examine a cohort of Canadian DTC genetic test users, yielding many findings similar to our own (e.g. privacy resignation, unawareness of family implications, sharing data for greater good, need for greater transparency) [42]. However, our cohort seems generally more aware of relevant privacy concerns. We suspect this is because our study included both individuals who had taken and those who only considered taking such tests, whereas Baig et al. focused on the former.

Patient perspectives on health data. Finally, we briefly note that many researchers have explored patient perspectives related to other kinds of health data, such as electronic health records (EHR), using card sorting tasks, semi-structured interviews, and surveys [43]–[46]. Shen et al. provide a systematic review of studies exploring these patient perspectives [47]. These works suggest that patients' attitudes toward privacy in an EHR context are complex, and many desire transparency and fine-grained control; these attitudes are echoed by our participants. Researchers also note that privacy concerns sometimes lead patients to behaviors that may place their health at risk, including avoiding their regular doctor, withholding medical history information, or avoiding tests altogether [45], [48]. This may be a particular concern for members of marginalized or stigmatized groups [46]. We observe a related trend in which participants from marginalized groups appear to express greater reservations about the consequences of genetic testing.

3. Methods

We investigated our research questions using a semi-structured interview study. This work was approved by University of Maryland's Institutional Review Board.

3.1. Recruitment

We recruited participants from the Washington, DC metropolitan region. We primarily recruited using an advertisement on Craigslist, but also posted flyers in public areas in and around University of Maryland (UMD) and advertised on various UMD email lists. All recruitment efforts instructed interested individuals to fill in a screening survey, including an online consent form (administered through Qualtrics, see Appendix A). To be selected for the interview, we required that respondents be at least 18 years of age, and have either taken or contemplated taking a DTC genetic test. We did not restrict our population to those who had taken a test, as we were interested in uncovering the reasons for which people choose to proceed with the test (or not). We did not mention that we were studying privacy attitudes in our advertisement, instead stating: "The study is targeted at understanding the motivations and overall perceptions of existing and prospective DNA toolkit users with respect to DNA and its testing technologies." We recruited and interviewed participants until we achieved response saturation [49].

3.2. Interview Protocol

We invited screening survey respondents meeting eligibility requirements to participate in the in-person interview. Interviews occurred between May 2018 and August 2019. The primary interviewer was constant for all participants, while the secondary interviewer rotated between the other researchers. The semi-structured interview was audio recorded for later transcription and analysis purposes. Participants were compensated with $20 for their time, along with $6 to reimburse participants who paid to park on our campus.

The interview study took place on our campus. The whole process took about 45 minutes per participant and included (1) an initial introduction, (2) obtaining written consent (see Appendix B.1), (3) a semi-structured interview (see Appendix B.2), and (4) a post-interview questionnaire to collect basic demographic information (administered through Qualtrics, see Appendix B.3). The entire process is laid out in detail in Appendix B. We designed the interview in four sections to assess different aspects of participants' experiences, opinions, and knowledge regarding DTC genetic tests:

Section 1: Experience. In this section we assessed the participant's experience with DTC genetic testing thus far. We asked if they had taken a test, and if so to specify which one(s). If they had taken a test, we asked them to detail their experience. If they had not, we simply asked for their opinions on these tests. We then asked them why people generally may choose to take these sorts of tests, and to clarify their reasons for participating (or not).

Section 2: Benefits and drawbacks. Here, we asked participants what they believed to be the benefits and concerns that may be associated with DTC genetic tests. We then asked them to specify if any of the concerns they listed were personal concerns, and if so how serious they were. If they had taken a test, we asked them to walk us through their reasoning to take the test in spite of any listed personal concerns. If they had not taken the test, we asked them to elaborate on which concerns had prevented them from doing so thus far. Note that after running six of these interviews, it became clear that the financial cost associated with these tests was a significant deterrent, so we added a question at the end of this section explicitly asking if cost played a role in their decision. We also added some financial questions to the end of the post-interview demographic questionnaire (see Appendix B.3).

Section 3: Knowledge. We first examined the participant's knowledge of the procedures used in DTC genetic testing to get from a user-submitted sample to results. We then asked what they knew about DNA, what it says about you, and where you get it from. We followed by asking them to clarify what kind of inferences can be made based on someone's genetic data. If they mentioned that DNA contains information about one's relatives in addition to oneself, we asked whether anyone else should be involved in an individual's decision to take such a test; otherwise, we first explicitly discussed the issue of family inference, and then asked about others' involvement.

Section 4: Privacy. We purposefully completed this section last to see if participants organically brought up privacy concerns. We first asked participants about the data handling practices that DTC genetic testing companies engage in. Specifically, we asked what a company's responsibilities toward their customers are upon data collection, and followed up by asking what participants thought companies were actually doing. We then asked participants to detail specific practices that they thought companies either should or should not engage in, and why. Possible follow-up points included longevity of data storage, the possibility of in-house research, the possibility of sharing and/or selling of the data to third parties, what form the data is shared in, and who the relevant third parties might be.

We then asked participants to evaluate the appropriateness of different uses of the collected genetic data in the context of sharing. We did not specify any uses, so participants had to come up with possible uses on their own and then evaluate them. Finally, we asked participants about their feelings toward a handful of predetermined hypothetical scenarios where their data might be shared, including their doctors, bosses, friends, insurance providers, and pharmaceutical and/or medical device companies.

3.3. Data Analysis

Interviews were transcribed and then qualitatively analyzed using open and then axial coding [50]. MAXQDA software[2] was used to aid in this process. Two researchers used an iterative coding approach to create a final codebook. In each round, 3 transcripts were independently coded, after which differences in coding and changes to the codebook were settled by discussion amongst the two researchers. The initial codebook was developed after 3 such rounds, and fine-tuned in subsequent rounds. After completing this process for all 24 transcripts, we obtained an inter-rater reliability value of Cohen's $\kappa = 0.697$, which is characterized by Landis and Koch as "substantial" [51]. Disagreements were then resolved to 100% for each transcript, yielding the final set of codes. Inter-rater reliability was calculated using: http://dfreelon.org/utils/recalfront/recal2/#doc. After completing the open coding phase, three researchers used axial coding to identify larger themes stemming from the more fine-grained codes. In our results (§4), counts of codes are provided for the reader's convenience, but should not be interpreted as generalizable prevalence.

3.4. Limitations

Like all human-centered studies, the findings of this work are subject to limitations. We cast a wide net for recruitment (through Craigslist, and around the university) for the purposes of obtaining a diverse population, but we only partially achieved this goal. First, the education level of the average participant was quite high. All participants had at least a high school education, and 16 had at least a bachelor's degree. Additionally, the vast majority of our participant pool was female (n=20). These distributions are not representative of the general population, and may affect the generalizability of these findings.

2. https://www.maxQDA.com

Moreover, this study was conducted in the U.S., in the context of historical (and ongoing) prejudices against racial and ethnic minorities, a mostly privatized health care system, and fewer digital privacy and consumer protections compared to other nations. As such, our findings may not be generalizable to other cultural contexts; however, the U.S. is and likely will remain the largest market for DTC genetic tests, making it an important population of interest [52], [53].

Further, we selected participants who had only *considered* taking a DTC genetic test, as well as those who had taken one, because we were interested in understanding the reasons for both decisions. However, participants who have only considered but not taken the test may be less informed about the process than those who have taken it. Also, participants who have so far opted not to take a test may at some point change their minds. Our results should be interpreted in this context.

4. Results

We present our findings as follows. First, we provide demographic details of the participants (§4.1). Next, we give an overview of the benefits, drawbacks, and knowledge expressed by our participants in Sections 2 and 3 of the interview (§4.2). Finally, we discuss the dominant themes extracted from the fine-grained codes through axial coding (§4.3 - §4.6). These themes are further distilled in §5 into hypotheses to be addressed in future work.

4.1. Participants

A total of 24 participants were recruited and interviewed. Of these, eight had taken a DTC genetic test at the time of the interview, as shown in Table 1. Three had taken multiple tests: T003 used 23andMe, AncestryDNA, uBiome, and an Alzheimer's-specific test; T006 used 23andMe and AncestryDNA; T020 used AncestryDNA and FamilyTreeDNA. N007 was not interested in taking the test herself, but rather for her son, in order to gain insight on his medical condition. Detailed demographic information for each participant can be found in Table 1.

TABLE 1. PARTICIPANT DEMOGRAPHICS.

ID	Source	Gen.	Age	Eth.	Edu.	Area of study
23andMe						
T005	UMD	M	30-39	O	MS	CS
T008	CL	F	60-69	M	MS	Public Health
T018	CL	F	18-29	W	HS	-
AncestryDNA						
T009	CL	F	30-39	M	PhD	PA/E
T010	CL	F	18-29	B	BS	Finance
Multiple						
T003	UMD	M	50-59	W	PhD	Physics
T006	UMD	F	50-59	W	SC	-
T020	CL	F	30-39	M	Prof.	LIS
None						
N001	UMD	F	18-29	AHP	SC *	Food Science
N002	UMD	F	18-29	B	HS *	BioE
N004	UMD	M	18-29	W	BS *	CS
N007	CL	F	50-59	W	MS	Psychology
N011	CL	F	18-29	B	BS	IA/PR
N012	CL	F	18-29	W	BS *	Public Policy
N013	CL	F	40-49	W	MS	School Psych.
N014	CL	F	40-49	-	BS	Engineering
N015	CL	F	-	-	Prof.	-
N016	CL	F	50-59	B	HS	-
N017	CL	F	60-69	B	MS	-
N019	CL	F	60-69	W	MS	Economics
N021	CL	F	70+	W	HS	-
N022	CL	F	60-69	B	AS	Accounting
N023	CL	M	40-49	W	T	Info. Systems
N024	CL	F	18-29	AHP	BS	AS

Source: UMD = Recruited on campus, CL = Recruited on Craigslist. **Ethnicity:** AHP = Asian, Native Hawaiian, or Pacific Islander; B = Black or African American; M = Multiethnic; O = Other; W = White. **Education:** HS = High school graduate, diploma or the equivalent (for example: GED); SC = Some college credit, no degree; T = Trade/technical/vocational training; AS = Associate's degree; BS = Bachelor's degree; MS = Master's degree; PhD = Doctorate degree; Prof. = Professional degree; * = Student at time of interview ($n = 4$). **Field of Study:** CS = Computer Science, PA/E = Public Administration/Education, LIS = Library and Information Science, BioE = Bioengineering, IA/PR = International Affairs and Public Relations, AS = American Studies.

4.2. Relatives Tested, Benefits, Concerns, and Knowledge of Testing Procedures

In this section we provide an overview of participants' responses to Sections 1–3 of the interview, namely:

- What was the participant's experience with DTC genetic testing?
- What are the benefits that participants associate with DTC genetic tests?
- What are the concerns that participants associate with DTC genetic tests?
- How much knowledge do participants have regarding the testing procedures used by DTC genetic testing companies?

This information was pulled directly from the codes assigned during open coding (see Fig. 1). A more detailed discussion of the information in this figure follows. Note that generally we do not list participant counts for specific codes when they are available in the figure; however, where relevant, we do give a breakdown of participant counts for total (vs. unprompted) concerns. Unprompted concerns are those that arose prior to Section 4 of the interview, which specifically referenced privacy and data security (see § 3.2).

4.2.1. Relatives tested. Participants were largely aware of family who had taken a DTC genetic test; 11 said one or more relatives who had done so (see Fig. 1), and several specifically differentiated blood relatives from relatives by marriage. Notably, these 11 included all participants who had taken a test ($n = 8$), but only three who had not. This suggests a possible familial-social aspect to taking these tests. Indeed, a number of our participants indicated as much. N013, for example, states "my brother in-law ... [has] taken one of the ones that's more about what your heritage is ... after I heard about that I thought oh that might be interesting for me to take."

4.2.2. Benefits. The primary benefits listed by participants were the ability to uncover more about their ancestral background (**discover ancestry**) and to learn about rele-

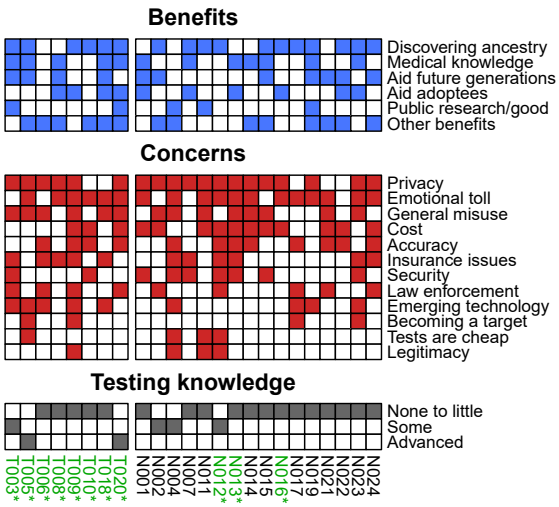

Figure 1. Benefits, concerns, and participant knowledge of testing procedures. On the left are participants who took one or more DTC genetic tests (8), and on the right are those who did not (16). Benefits and concerns are ordered by prevalence (highest on top, least on bottom) with the exception of "Other benefits," and testing knowledge is ordered from top to bottom by increasing knowledge. Participants in green with an asterisk (*) indicated that at least one relative had also taken a test.

vant medical knowledge such as predisposition to specific diseases (**medical knowledge**). This is in line with the fact that these are also the primary offerings of most genetic testing companies. The knowledge derived from these two primary benefits informed the next most-cited benefits, i.e., **aid future generations** and **aid adoptees**.

Most participants who cited **aid future generations** did so in the context of medical knowledge: information gained regarding predisposition to diseases and other health problems could be used to ascertain what kinds of medical issues potential children might have and how to combat them, and even to determine whether or not to have children at all. N019, for example, discussed a family member who developed Huntington's disease: "she knew she had a 50:50 chance of getting the disease and therefore she and her husband made a decision not to have children [but adopted two]. You know, you need to know those kinds of things, those kinds of risks in your family."

In addition, one participant who listed **aid future generations** did so in the context of ancestral knowledge. N024 discussed being the child of immigrants, and saw learning more about her ancestry as a way for her (and her sister) to preserve her culture and heritage for children and grandchildren who will have fewer ties to her country of ethnic origin: "Because ties like the language . . . our relationships with our extended family . . . those are the things that tend to fade away the fastest. So, having a family archive would at least help, I guess, the two of us, but also our children and grandchildren, to feel a greater sense of . . . a better idea of what their roots are like."

Participants who cited **aid adoptees** emphasized the ability to gain otherwise unknowable knowledge by adopted individuals. N022 stated, "especially people [that] have been adopted . . . I'm sure that they're really interested in what their exact background is if they don't have any information on where they were from."

The last benefit that was commonly cited was the ability to contribute to **public research/good**. T003, for example, has deposited his information in an online, public repository of genetic information in the hopes of "contribut[ing] to scientific research." T020 posited that the data gathered through these tests can help genealogists in their research endeavors.

Other benefits (grouped together in the last row of benefits in Fig 1) that were less commonly cited included the accessibility of the tests (N002: "It's easy. They mail it to you, so it's convenient"), finding unknown relatives (N014: "There's a benefit to finding, I guess, relatives that you didn't know existed too"), and confirmation of existing knowledge about your ancestors and ancestry (N019: "Just to find out if what you've always been told is actually true").

4.2.3. Concerns. The three most-cited concerns were **privacy**, **emotional toll**, and **general misuse**. Nineteen (17 unprompted) participants explicitly mentioned **privacy** as a possible concern associated with DTC genetic tests. N014 expressed concerns that were common in the participant pool overall: "who has my data, right, and when is it going to be used, and what is it gonna be used for?" Interestingly, although many participants cited privacy as a concern in general, this does not necessarily translate to personal concern for their own privacy. This phenomenon, along with other aspects of privacy concerns, are discussed in more detail in §4.3.

Emotional toll was cited by 18 participants in three specific contexts. First, several expressed concern over the fact that the medical information obtained from a DTC genetic test may cause anxiety. N013 said that it is "kind of scary because do you really want to know what you have a risk for? I mean what if you can't do anything about it? Then that would be kind of scary." Second, participants discussed the possibility of a user uncovering information about their families that was hidden from them, perhaps purposely. T008 listed a handful of such cases: "People that were never told that they were adopted, discovered that their father is not their father . . . mother had an affair while she was married." Finally, some participants mentioned issues with regards to racial history that could be uncovered by genetic testing. A number of African American participants explicitly mentioned that they had no desire to uncover traces of European ancestry due to the high likelihood of this resulting from assault or abuse in the context of American slavery and discrimination. For example, T010 stated, "I was just concerned about how much European ancestry would be there . . . More than likely rape during slavery." Clearly, the **accuracy** of the tests is relevant in all three of the above contexts.

Fourteen (11) participants brought up the concern of **general misuse**, such as mishandling of data or general distrust of the company. N002 stated, "I guess your DNA or genetic information can be mishandled or possibly replicated in some weird way." N004 painted this possibility more concretely: "I think that the fact that the founder of it [23andMe] was the former wife of Google's person, I can't imagine they don't have a plan further than just reporting everyone's individual DNA sequences." N023 offered a similar sentiment: "You're dealing with a large organization with a lot of resources, and the potential for abuse or misuse is there."

This common refrain of a lack of trust for the companies and their practices, combined with general privacy concerns, was echoed in many of the other concerns brought up by smaller numbers of participants. These included the **security** of the data in the face of breaches - 9 (5); the possibility of **insurance issues**, i.e. that health insurance providers will increase premiums for individuals with genetic predisposition to health conditions - 11 (7); and **law enforcement** accessing the genetic data - 9 (7). Moreover, the fact that the **tests are cheap**, given the historic costs of genetic testing, is a point of worry - 4 (3). N004 opined, "historically, these tests were something in the range of tens of thousands of dollars. My perspective is that they scaled by cutting down quality", or that "they potentially have $500 they're planning to recoup. Or $5000, or $5 million" through, for example, invasive or inappropriate data practices.

Further, the **legitimacy** of the company themselves was brought into question - 4 (2). N004 noted that unlike the medical field, the space of DTC genetic testing "is not relatively established." A concern derived from this issue is the fact that "privacy laws don't typically keep up with technology so I feel like you have to have a lot of people that are just giving away their DNA material not fully knowing how it's going to be used" (T009).

This last point of contention from T009 goes hand-in-hand with the concern that genetic testing remains an **emerging technology** - 9(3). As N023 stated, "I think this is obviously a new kind of emerging, evolving field, if you will. The applications are almost limitless." T003 elaborated, "there's probably lots more of those things that we don't understand yet, because the data doesn't exist." N017 took this a step further, saying "it doesn't seem bad, but to me it seems like one of those things, especially being a minority in this country, a lot of things that you wouldn't think of on the face of it that's any danger to it, but then suddenly if the statistics are turned against you."

N017's concern about **becoming a target** was echoed by 3 other participants (T005, T009, and N023) - 4 (3). Of these 4 participants, 3 were minorities, leading us to the observation that this particular concern may be more prevalent among populations which are more likely to be affected by discrimination.

Finally, financial **cost** stood out as the only short-term concern regarding DTC genetic testing. Despite the drop in the price of genetic testing in recent years, 14 participants cited cost not only as a concern, but as a concrete barrier to taking such a test. Of these participants, 3 did end up taking a test (T009, T010, and T020). T009's test was purchased for her, and the others explicitly mentioned waiting for the price to go down due to discounts (T010), or over time (T020). Among the remaining 11, a common theme was waiting for a sale to get such a test at a discounted price.

4.2.4. Knowledge of testing procedures. We asked participants to explain the testing procedure involved in going from a physical sample (e.g. saliva) to the data that is reported back to customers. The vast majority of participants (18) had **none to little** knowledge of these procedures. Four had **some** knowledge, and only 2 had **advanced** knowledge. These categories were defined as follows:

None to little: No knowledge, or participant only talks about user-stage procedures (e.g. mailing a sample back)

Some: Knowledge of basic procedures (e.g. sequencing, use of genomic databases)

Advanced: Advanced understanding of methodology (e.g. use of markers that are correlated with disease, comparisons are made only against sample populations)

4.3. Awareness of Privacy Concerns does not Translate into Action

While privacy concerns do appear to be on participants' minds, this awareness does not always translate into behaviors to safeguard privacy. We examine this phenomenon in greater detail in this section.

4.3.1. Concerns must be internalized to have an impact. Participants are clearly aware of the privacy concerns surrounding DTC genetic testing. However, this awareness does not necessarily translate into a personal concern regarding privacy. Though 19 (17 unprompted) participants indicated that privacy was a possible concern, only 6 of them (N004, T005, N007, N011, N016, and N024) went on to say that privacy was a *personal* concern that would affect their decision to take such a test. Indeed, while the presence of general privacy concerns was not clearly associated with who goes on to take a test, the presence of personal privacy concerns did seem to be. While our qualitative methodology does not support drawing a firm conclusion, we do see a clear trend (shown in Table 2) that may be worthy of future investigation.

TABLE 2. PRIVACY CONCERNS OF PARTICIPANTS.

	Personal	General only
Tested	1	7
No test	5	11

Of the six participants who listed privacy as a personal concern, only one had taken the test (T005). The remaining 18 participants did not list privacy as a personal concern. Of these 18 participants, seven had taken the test (T003, T006, T008, T009, T010, T018, and T020). The 11 who did not take the test had some other primary reason for not having done so. For eight participants, it was the financial cost; for the other three, it was just a matter of making the time to get to it.

4.3.2. Privacy resignation. Some of the general security and privacy attitudes expressed by participants may help explain *why* general privacy concerns do not necessarily translate to personal ones. These attitudes are well documented in the literature in numerous privacy contexts (see §5). Eight (3) participants made the claim that **nothing is secure**, i.e., that the protocols used by different companies are never impregnable, and that most methods are equally insecure. T020 thought that "it's just a matter of time before a hacker hacks into one of these databases." T008 claimed, "at some point almost nothing is private anymore."

Six (3) participants also cited the notion that their **information is already out there**. T003 discussed the relative gravity of having genetic information leaked versus medical records. He concludes that the latter is more

sensitive, and that he is "more concerned about that sort of information getting out, and it's all over the place." T006 talked about the prevalence of data breaches, and concludes "I'm losing personal information right and left every place else [so] I'm not that paranoid."

Three (2) participants also claim that they are simply **not important enough** to be concerned. T006 claimed "I do not have enough wealth, status, anything else, to worry about my actual personal identification information being stolen through computer hacks of stores or banks or something, because I don't have anything." T018 simply stated, "I'm not that interesting."

Another attitude commonly seen in other security contexts is the idea that the user has **nothing to hide**. This sentiment is echoed by three (2) of our participants. N001 outright said that she "personally wouldn't be too concerned because . . . I mean I don't really have anything to hide." T008 asserted, "I have not committed any crimes. I had to be fingerprinted for various jobs. There's no one I know from my family that has committed any crimes that they need to be hiding from." Similarly, N002 commented "I guess I don't indulge in any bad things so I don't know if it [law enforcement] would ever come my way."

Finally, T018 cited security and privacy **fatigue**, i.e., the constant need to be vigilant of privacy issues has worn her down to the point of no longer caring. She said, "I have no faith in any company to do anything so I just assume that when I give them my information, they're going to do terrible things with it . . . my phone listens to me all day, my targeted ads and whatever, that I just . . . I know I should care because we should all care that we have no privacy anymore, but I just don't . . . I mean a lot of it is I grew up here, and I mean you live in proximity to D.C. right, there's the constantly monitored thing happening, and I grew up here so I'm like used to it."

4.3.3. Benefits outweigh the risks. Another factor explaining why participants go through with testing despite being aware of privacy and other concerns is that they believe that the benefits of DTC genetic tests outweigh the risks. Eight (2) of our participants voiced such a stance. T006 simply stated "I'm willing to take what I consider to be a reasonable risk." N014 elaborated on the matter a little further: "Initially it was shocking that we're in this world of people wanting to share all this information, but then at the same time, you know, things are moving along, and somehow the benefits or the perceived benefits prove to be more than the risks." N007 provided an interesting perspective in this context. While she expressly was not interested in taking the test herself due to some of her concerns, she had no such qualms over getting the test done for her son. Due to his medical condition, she says "he's going to be on Medicaid, forever." As a result, the benefits outweigh risks she might otherwise be concerned about "because nothing's going to happen to my child's health insurance no matter what kind of information I get."

4.4. Sharing Can be Acceptable

We asked participants for their opinions on the prospect of DTC genetic testing companies sharing user data with third parties. This possibility is viewed negatively overall. However, follow-up questions suggest participants hold a more nuanced, contextual view. We asked participants about the kinds of organizations that might be third party recipients of DTC genetic testing companies, what they considered to be appropriate uses of shared data, and their feelings on specific sharing scenarios. We find that while many participants emphasize that sharing should not harm the user, it does not necessarily have to be for the user's direct benefit either. Note that remarks in this section were largely derived from prompted responses.

4.4.1. Third party access is acknowledged and disliked. When asked whether or not DTC genetic testing companies sell or share the data they collect from users, the overwhelming majority of participants acknowledged that this was probably the case (n=19). Thirteen participants said this explicitly. Three participants expressed that they hoped this was not the case, but acknowledged that it probably is. T010, for example, said, "I guess it would be naive of me to think that they're not sharing the data." Three other participants suggested that users essentially forfeit all the rights to their data when they elect to do one of these tests, so what happens to it from there on is up to the company, including sharing. N004 assumed "anyone who has taken one of these tests has fully relinquished all control."

Moreover, of these 19 participants, 10 brought up the point that these DTC genetic testing companies are for profit. With this comes the assumption that these companies will try to make money off their customers' data. According to N012, "I think that's just how a lot of companies operate today. The market for people's data is really hot and profitable. I would wonder if that's part of the reason, they can offer the service widely and for a much cheaper price because they have revenue coming from that too."

Along with conceding the high likelihood of sharing and/or selling the data, participants exhibited largely negative sentiment towards such transactions (17). T018 commented, "Well ideally I think it shouldn't be shared." N004 said, "There's probably nefarious things to use people's genetic information for, and sharing with third parties, especially with financial incentive, would allow that to happen as smoothly as possible." This response also distinguishes between data being sold to third parties versus simply being shared. Some participants did not make a distinction between the two and simply disliked both, as in the case of N022: "I don't like either of them." However, some participants focused on the financial aspect. For example, N021 stated that "if they're in it more for the money, they're not really in it for your . . . anybody's benefit but their own. So then it just kind of changes that whole scenario." Regardless, most participants, including those expressing negative sentiment towards selling and sharing, present a more nuanced view upon further discussion: their attitude toward third party access is context dependent.

4.4.2. Sharing in the public interest is appropriate. The context in which sharing was overwhelmingly deemed appropriate was for genomic and medical research purposes, with 21 participants reporting this sentiment. Sometimes, this was in the context of finding cures or therapies for health issues. N017 said, "I guess if it's something that

is designed to save lives, like you know if we share this data with XYZ they think if they study enough of these you know, they're trying to come up with this miraculous cure for diabetes. Something like that, you know." Others looked at the situation from a more preventative standpoint. N002 for instance remarked on how this data could be used for "detecting diseases, or genetic disorders . . . I guess I'm thinking more in terms of like the medical realm."

Medical research was not the only sort of research that was viewed positively. N012 supported using this data to "further collective knowledge of history or immigration." N024 similarly remarked that "understanding migration movements from however long ago for educational or scientific purposes . . . could be appropriate."

4.4.3. Inappropriate uses. However, not all research was viewed positively. Unauthorized genetic engineering, including cloning, was listed as an inappropriate use by 6 participants. Another class of research that was unpopular was any sort of weaponization of a user's genetic information for the purposes of inflicting harm. This sentiment was reported by 12 participants. T005 opined that third party recipients should not be "companies who create weapons, for example." N011 disliked the fact data sharing in such a scenario would implicitly involve her in harming others: "I would hate to learn that something I provided had anything to do with harming someone else." N017 honed in more on the idea of institutionalized harm that may result from this kind of sharing, such as "anything that would work as a negative for individuals or particular groups." T020 gave a more concrete example of such a scenario: "Searching the data of all African Americans to find crime, or, to solve a crime."

Seven participants commented that utilizing user data to support targeted marketing and advertising was inappropriate. N001 gave the following example: "If I just log into my social media accounts and find out, 'Hey! We found out that you like salty food through your Ancestry.com. Here are the links to buy chips.' I would be pretty mad about that." N012 stated, "First of all, I am just tired of the targeted ads and random companies that I've never done business with knowing things about me. It just feels weird and invasive and this in particular feels highly personally since your DNA is the building blocks of who you are."

4.4.4. Specific scenarios: Ask first and do no harm. We asked participants for their feelings on a number of hypothetical scenarios where their genetic data might be shared. The general consensus was that sharing this information can occur only if two conditions were met: (1) sharing their data must not harm them, and (2) their consent must be solicited in advance of sharing.

In particular, the idea of their doctor having their genetic information was considered acceptable. (This was a unanimous sentiment, though N019 stipulated that she would only share a hard copy that she could then take with her, rather than leaving it at the doctor's office. This remark was prompted by her understanding that a high frequency of identity theft occurs in medical offices.) T020 said she was "comfortable with them receiving my genetic information because they are taking care of me." N022 emphasized the importance of consent: "I think if it was my choice to send it, like if I disclosed it to them then yeah. I don't think it should be an automatic sharing of information." N017 was somewhat different in that she wanted to limit her doctor from learning certain specific information. In particular, she did not want her doctors to receive any evidence that she might develop a degenerative disease, such as "Parkinson's, Alzheimer's, what have you," because this sort of information would be a "cloud laying over your head that's just this time bomb going off . . . wait, I did such and such, does that mean I have Alzheimer's now?" As a result, she had no interest in knowing about any such predisposition, or risking that her doctor would communicate it to her if they knew.

In contrast, the idea of one's employers seeing their genetic information did not sit will with participants. All 24 agreed this is overwhelmingly a poor idea for two specific reasons. Some argued that this would be an invasion of privacy. For example, N021 simply stated "It's too personal." Others argued that there was a potential for harm to the user. N023 offered the following hypothetical example: "it could definitely be used in a discriminatory manner. If you have two employees, and one has a bunch of health issues that you can easily identify or see, and the second one doesn't . . . possibly you're going to discriminate and pick the one that doesn't. It would certainly give certain people an unfair advantage."

Insurance companies seeing user genetic information also elicited a unanimous negative response. N011 argued, "too much goes into insurance already and I don't want my DNA making my rates higher for some reason." N002 provided more detail, saying "if you're predisposed to something that you didn't know of, and — yeah that would cause your insurance to go up and all that other stuff." N016 agreed that insurance companies should not be able to see their customers' genetic information, but questioned whether or not this is already the case: "No, I don't think they need to see it, but I don't know if they would see it anyways, I don't know how that works as far as the company goes." Confusion as to whether or not insurance companies already have this kind of data was echoed by a handful of participants.

Four participants also voiced slightly different stances. N004 and T005 posited that some users might in fact benefit from insurance companies seeing their genetic data on account of them being healthy, i.e., that their premiums would actually decrease rather than increase. T003 and T009 said that they would consider sharing their data if they would receive some kind of benefit as a result (either a discount, or more coverage, respectively). Despite these possibilities, all four participants still agreed that insurance access to genetic information would not be a good idea in general because of the potential for harm.

Pharmaceutical and medical device companies were viewed as a sort of grey area. Fourteen participants expressed an overall negative opinion regarding their data being shared with pharmaceutical companies. However, seventeen participants acknowledged the role that these sorts of companies play in developing new drugs and treatment options, i.e., research in the public interest. For this reason, they were okay with their information being shared, but (a) the data should be subject to certain constraints such as consent and anonymization (see §4.5.3), and (b) it should only be used for the purposes of research

and development of therapies. T005 drew a distinction between mere production and scientific contribution: "some companies just take a formula and produce that medicine — they don't really produce new medicine. So, companies who [do] research and tries to improve the state of the art are welcome to use the data."

There was mixed reaction to the notion that pharmaceutical companies may use genetic data for marketing purposes. Six participants were okay with this, whereas nine were not. N014 remarked on the possibility of personalized recommendations: "I feel like there's such a limited view of what's available out there [in terms of drugs and treatment options]. So the more I'm exposed to things in a way that's more targeted and helpful for me, I feel like that's beneficial for me." N021, in contrast, was wary of the financial incentives that are inherent in marketing: "anything that they're trying to benefit with money . . . kind of changes the whole thing." N013 added to this with reference to the notion of doing public good: "the research department would do good with it where the marketing department is using it to sell something."

A handful of participants did not distinguish between the research and marketing arms of pharmaceutical companies on account of them both being parts of the same entity. As such, the negative connotations of marketing defined their stance on pharmaceutical companies as a whole. When explicitly asked if the distinction between research and marketing mattered, N015 simply responded, "No. It's the same family." N024 echoed this sentiment, stating "I would consider it to be all the same, especially assuming that even though these are different parts of the company, they're all part of the same legal entity."

4.5. User Desires

Overall, participants feel that the users of DTC genetic tests should be the priority, since they are the ones paying for a service. Deference should be given to the user's wishes, rather than the company's interests. Remarks in this section were largely derived from prompted responses.

4.5.1. Communication should be clear on all fronts.

When asked about the responsibilities that DTC genetic testing companies have towards their customers (Section 4 of the interview, see § 3.2), participants emphasized that DTC genetic testing companies have obligations to (a) communicate any information that may affect the user to the user, and (b) ensure that these communications are accessible to the average user.

Informed consent. Eighteen participants brought up the importance of obtaining informed consent from customers. N011 stated it is the company's responsibility to inform customers "how they plan to use the data once we've gotten what we signed up for. So, I'm signing up for this information; are you destroying the data, are you storing it somewhere, are you selling it to someone else?" N023 discussed consent in the context of data sharing specifically: "unless you've consented to it, they should not be able to share that information with other entities, government or research entities. Unless you specifically consent to that type of participation, the information shouldn't be shared with anybody, any partners, any other organizations, obviously, because it's obviously, it's sensitive personal data . . . " Participants emphasized that consent is a two-way street, not limited to the company simply telling customers what they plan to do; rather, users should be able to opt in or out of particular practices. N001 addressed the prospect of DTC genetic testing companies doing in-house research: "I think people should have an option to not be used if they don't want to."

Company transparency. Related to consent, another property of communications that participants desired was transparency (n=15). T009 stated that these companies "do need to tell you how your DNA is going to be used, they need to inform you if they plan to change it." However, the manner in which this information is communicated is important. T009 also said "I hate to say that they need to inform their customers how their DNA will be used, because it's probably listed in that, you know, really long page of, you know, the consent forms, usually it's in legalese and you know many, many pages."

The shortcomings of the unintelligible and overly complicated terms of service provided by these companies was echoed by multiple participants. N014, for example, said "I think a lot of companies, in the fine print, disclose what their responsibilities are but I don't think that's very clear to a lot of people . . . not even clear to the company itself. You know, I think that emphasis needs to be also made clear in the commercials, on their websites." Moreover, broad statements regarding the company's right to change their terms at any time, while technically transparent, were disfavored. N024 remarked, "Last minute, whole 'terms and conditions can change whenever' thing is something that doesn't really sit well with me. It feels . . . even though I'm sure that legally covers them from any liability, that, to me, feels kind of dishonest."

Transparency was also desired in contexts outside of data handling. On the subject of the scientific procedures used to make claims about ancestry, N024 commented "I'm not sure to what extent they're open about what specific procedures they use to evaluate where people are from, but I would want them to be very transparent about that as well." When asked to elaborate, she echoed the sentiment that long, fine-print documents in legalese were not an ideal means of communication. Specifically, she "would want that [scientific procedures] to be expressed very clearly, and in layman's terms, and not just a wall of text that they know that most people wouldn't have the tools or resources to understand."

N013 was also concerned with the transparency of the reported results: "I guess I feel like they also should fully explain well what the results mean or like, I don't know, not require that you should talk to a physician about it or geneticists when you get it but strongly encourage that. I don't know. I have heard some weird things where people get a thing that says you have double the risk of a certain disease than somebody else but the actual risk of getting that disease is 0.001%. So, it's really not that big a deal that its twice the risk but people see that and panic. So, I think there is some responsibility that companies have [to] make sure the people understand clearly what it is that they're reading when they get the report."

4.5.2. Protection of data (and samples) is important.

When asked about the data storage and handling practices that DTC genetic testing companies engage in (Section 4

of the interview, see § 3.2), participants largely focus on customer protection.

Privacy and security measures. Twenty-one participants mentioned that one of the company's responsibilities towards their customers is "privacy;" 17 mention "security." Four participants mentioned that the data should remain anonymized within the company. Another four emphasized the need for limited access within the company itself, e.g., there should be permission structures for how company employees can access user information. Three participants used the term "encryption" specifically. Finally, two participants were of the opinion that these companies' data practices should be subject to regulations from an external party (e.g., the government).

Data longevity. Participants had a wide range of views when it came to how long the company should have access to user data. Four participants asserted that the data should be destroyed immediately following the transmission of test results to the customer. (Three participants also addressed the fate of the physical samples, saying that they should be destroyed once the results of the genetic analysis are obtained.) Eleven participants desired some sort of term limit (generally ranging between 1 and 10 years). Two participants said it should depend on the company's status, e.g., the company should retain the data as long as they remain in business (one of these participants mentioned the additional constraint that the user must maintain an account with the company). In all cases, the justification for these restrictions was to minimize risk of exposure. As N013 put it, "I think that would take the some of the risk out of somebody getting it."

The remaining seven participants said the company can retain the data forever, however this was predicated on the assumption that the data would be stored securely; they did not elaborate much on what exactly this meant.

Data integrity. Eight participants also voiced wishes regarding the integrity of their data (and samples). N002 addressed the physical samples, saying companies should be "making sure . . . that they don't mix up or mishandle samples, or, I guess, make sure it's sterile, and just properly organized." N015 emphasized the importance of "having the assurance that if you're submitting this [sample] that it won't be taken elsewhere or misinterpreted." N017 remarked that the companies should be "making sure that the results they give me are my results, that it's not confused with somebody else's." T003 discussed the storage of user data, commenting that "the repository should be so designed that there's a very low probability that there will be an uncontrolled change to the data that's in the primary repository."

4.5.3. Constrained data sharing. Even in the case of appropriate uses of consumer genetic data (see §4.4), participants voiced support for limiting how data is shared. Twenty participants said they would like for companies to remove any personally identifying information (PII) from their data prior to sharing. Six participants also mentioned the idea that the data should be shared in aggregate rather than as individuals. Participants asked for these constraints in order to protect user privacy, though the efficacy of such measures is debatable, as discussed in §2, §4.6, and §5.

4.6. Consequences are not Fully Grasped

Though participants are able to identify concerns surrounding DTC genetic testing (often unprompted), and are aware of information relevant to those concerns, they do not always understand the larger consequences that emerge as a result. This gap in understanding manifests in two specific contexts: the implications of family inference, and the highly personal nature of DNA.

4.6.1. Implications of family inference are unclear. The vast majority of participants (n=22) were aware that DNA is inherited from our parents. However, slightly less than half of these extended this idea to the realization that DNA contains information that can be used to infer details about one's family (n=9). Even among these nine, not all made the connection that taking one of these DTC genetic tests ultimately reveals information to the company not only about the user (who did consent), but also the user's family (who did not). N004 was keenly aware of this issue, commenting "if you take [the test], you're getting all information about all my siblings, all my parents, all my future kids for the next 20 generations." In contrast, T005 came to this realization as he was doing our interview: "Revealing this information will also reveal information about many other people; that's something that I am now realizing after this interview." T018 remarked, "I can see how that would be a privacy concern for your siblings. Huh. I really never thought of that that way."

User autonomy. Over the course of the interview we eventually unpacked this issue of family inference for all the participants, and asked them whether this necessitated the involvement of others (e.g., family) in a user's decision to take a DTC genetic test. The overwhelming majority (n=22) of participants maintained that this decision was ultimately a personal one. N023 framed this in the context of a cost-benefit analysis: "Even though it does reveal information about siblings or close relatives, at the end of the day, the potential benefits of getting the information outweighs your close relatives having some issue with you getting the test done." He also mentioned that "getting buy-in from close relatives or siblings about whether or not they're okay with it" would be a difficult task. N015 frames the issue in terms of personal independence: "Because it's personal, it's your DNA . . . because it's your control over your genetics, it's your control over your body history. So I don't think I need to get permission from someone."

Two participants (N021 and N024), however, were of a different mind. N024 framed this in the context of consent, since it is "your family's information that's being given out, and if they don't consent to that, then it's really just kind of wrong to do that." N021 emphasizes that family should have a say in the matter because it is "for their own privacy." Though these two participants were in the minority as far as the degree of family involvement, many (n=12) conceded that the decision to take such a test may merit a conversation amongst family, while noting that the final decision was theirs alone. Two participants (N014 and T018) noted that this decision may even merit the involvement of a professional, such as a doctor or geneticist, since the information revealed by such a test

may have important health and family implications such as disease proclivity or discovery of unknown relatives.

4.6.2. Genetic information cannot be anonymized. Participants, once prompted, cited the need for anonymization of user data both in the context of storage within a DTC genetic testing company, and in the context of sharing with third parties. In fact, several participants perceived the ability to anonymize data as a necessary condition for data sharing, even for uses that were deemed to be appropriate, such as medical research (see §4.5.3). Participants were also aware of the wealth of information that DNA contains, including parental source (n=22), disease information (n=22), physical characteristics (n=14), lineage information (n=10), geographical and racial/ethnic origins (n=9), family inference (n=9), and DNA's use in forensics (n=7). Four participants even noted that DNA is unique from person to person. Twenty participants made some mention of the fact that DNA says or tells who you are, or that it contains the genetic makeup of an individual. However, not a single participant realized the implication of this wealth of deeply personal information that is contained in DNA: it cannot truly be anonymized. In fact, Linda Avey, a co-founder of 23andMe, has acknowledged that "it's a fallacy to think that genomic data can be fully anonymized [9]."

4.6.3. Some participants express regret. Two participants expressed regret for having taken a DTC genetic test because they were unaware of the larger consequences of family inference when they tested. T009 and T020 framed their remorse in the context of law enforcement access to user genetic data. T009 remarked, "I wish I hadn't done that, but now that it's done, what's done is done. I know that I'm not going to get in trouble but I don't want to be linked DNA wise to my second cousins . . . what if they are a criminal, and they find me that way?" T020 comments, "When I first started taking the tests, back in 2008 and 2011, I was not concerned . . . But after I took the Ancestry test, it started to become more prevalent that they would be using it for crime scene investigation and crime detection, I sort of drew back and 'Hey, I don't necessarily want my information out there.' Not that I've done anything wrong, it's just that I feel like it's an invasion of privacy, it's an invasion of my family's privacy."

A third participant (T010) also expressed some regret, though to a lesser degree than the other two. Her issue was more centered around the marketing tactics of AncestryDNA rather than the consequences surrounding commercial collection of genetic information: she mentioned that some of their marketing techniques may "sensationalize slavery." N011 voiced a similar concern, though she has not taken any such test: "they just got backlash for their commercializing romanticizing slavery . . . I'm not interested in romanticizing that."

5. Discussion

We first synthesize four key findings from our results, then discuss associated recommendations and future work.

Privacy is a general concern but not a personal one. Our findings suggest that users are indeed aware of the privacy concerns surrounding DTC genetic testing. In fact, they organically bring up this issue without any prompting. This is fairly different from user behavior seen in other contexts, where privacy concerns often do not surface until researchers ask about them explicitly [54]–[56]. This is perhaps a hint that genetic data is somehow viewed differently than other kinds of user data.

On the other hand, despite the general consensus that privacy is an issue, few participants actually see it as personally relevant in a way that would preclude them from taking a genetic test. Participants' explanations for why privacy may not be an issue for them, personally, echo commonly repeated sentiments observed in contexts ranging from data sharing decisions in social media and smart homes to security-behavior decisions such as password choice or software updating. These explanations include the idea that no system can really be secure [57], their information is already out there [58], they are not important enough to be attacked [59], they have nothing to hide [60], and they have too much security fatigue to take further precautions [61]. Essentially, because they perceive their data as either unimportant or else completely beyond their ability to protect, the benefits of the genetic testing outweigh the marginal risk. This echoes the findings of previous work on genetic testing [37], [39].

Sharing need not be for users' direct benefit (but cannot be for anyone's detriment). We find that users, while generally opposed to the idea of sharing genetic data, acknowledge the fact that it is most likely happening. Faced with this reality, they offer more nuanced responses as to what they might consider to be appropriate or inappropriate uses of the data. Overwhelmingly, research in the public interest is deemed an appropriate use, whereas applications that would bring harm to the user or others are frowned upon. While sharing genetic data for scientific advancement was a sentiment found in previous literature, it was not explored in depth to uncover nuances in this position as done in the current work [37], [39], [41].

User desires are at odds with current practices, but also with average user behavior. Users claim they desire clear, accessible communication between the company and its customers. They insist on informed consent and transparency in all interactions, but also indicate that the current methods of doing so (generally via legal terms of service) are ineffective.

This problem is not unique to the genetic testing context, but rather can be observed in privacy and transparency notices more generally. There is a large body of work evaluating the problems inherent in the notice and consent paradigm. Schaub et al. discuss some the issues in current design choices of privacy notices, and offer recommendations for more effective models moving forward [62]. Reidenberg et al. interrogate the idea of interpretability of privacy notices, and find significant discrepancies between the intended meaning of the notice and the perceived meaning by users [63]. Furthermore, there were discrepancies in interpretation that were dependent on the user's level of privacy expertise. Cranor provides an overview of the long history of attempts to make privacy notices more useful and usable for end users, none of which have been particularly successful so far [64].

Phillips notes that users often conflate the existence of a privacy policy on a website with the assumption that their data cannot be sold or shared [23]. Of course, all of these issues only become relevant if users actually read privacy notices, which they generally do not [40], [65].

Users also indicate a desire to either opt in or opt out of specific sharing scenarios, exercising fine-grained control over how their data is used, echoing the desire for strong personal control over genetic data found in earlier literature [36]. Previous research, however, indicates that giving users fine-grained control overburdens them and is ineffective [66]. Moreover, the pay-to-play setup of many applications often makes it impossible for users to opt out of privacy-sacrificing features if they want to use the application for its intended purpose [67]. It therefore seems clear that although users want more transparency and control, common approaches for providing it are unlikely to work any better for genetic testing data than for other forms of data collection and sharing.

Implications of the uniquely identifying nature of DNA are not understood. Users often fail to realize that by giving their personal genetic data to commercial entities, they are also relinquishing familial genetic data. Additionally, genetic datasets are nearly impossible to effectively anonymize [1]–[7]. These two gaps in understanding may make the privacy problem multiplicative in nature, affecting multiple, non-consenting individuals each time a single, consenting user signs up to take one of these tests. This raises important ethical concerns, e.g., the potential use and release of familial genetic data without proper notice and consent. We discuss potential remedies below.

5.1. Recommendations

We suggest that DTC genetic testing companies employ default privacy settings specifying which cases the company can and cannot share user data. Our work indicates that medical research and other endeavors in the public interest are largely supported by users, and targeted marketing, weaponization, genetic engineering, and sharing with insurance providers were all viewed negatively. Defaults that align with the average user's desires are one mechanism to improve user protection and satisfaction. Regulation may be needed to enforce these requirements, or to explicitly disallow uses that are commonly considered inappropriate.

Additionally, we recommend greater efforts to promote awareness of the unique risks associated with genetic data (family inference, and the near-impossibility of anonymization) before people participate. This could include warnings issued from government agencies in the vein of [26] or [27], with an emphasis on privacy considerations in addition to the dubious nature of the advertised benefits and claims made by DTC genetic testing companies. Another approach could be concerted outreach and education efforts on the part of consumer advocacy groups. We note that the PGP specifically informs users of the familial risks associated with participation, which appears to have increased user awareness [37].

Finally, we suggest greater transparency regarding both regulations and company policies. Participants are clearly worried about the privacy, security, and potential mishandling of their data (see, e.g., §4.2.3). We observe — in alignment with prior work [40] — that this is partly due to lack of information about existing legal protections (e.g., HIPAA and GINA, see §4.2.3), and partly to confusion over company data-use policies (§4.5.1). This further exposes fundamental problems with notice and choice, particularly when company policies are frequently updated. Perhaps government or other third party intervention is necessary to ensure greater consumer awareness.

5.2. Future work

The four key findings we describe above can be treated as hypotheses for further in-depth investigation, possibly using surveys tailored to interrogate each one specifically. It would also be interesting to see how these ideas hold up (or do not) beyond the U.S., where cultural, political, and societal norms vary on specific and highly relevant axes (e.g., the minority experience, health care structure, and privacy laws). For example, prior research suggests that users believe clinical applications of genetic testing are best left to national health services (which do not exist in the U.S.) [33].

Another question worth exploring is why or how users seem to be more aware of privacy concerns in the context of genetic data than in other contexts (see §4.2.3). One possibility could be that the above-average education level of our sample caused an unrepresentatively high frequency of unprompted privacy concerns. It could also be an effect of extensive media coverage DTC genetic testing and associated issues. It is also possible that some knowledge of the nature of the information contained in one's DNA prompts users to think more about their privacy (though not necessarily enough to take steps to protect it).

We also find minor evidence suggesting that members of marginalized groups may be more likely to identify or even act on concerns related to institutionalized targeting on the basis of genetic composition (see §4.2.3). There is also relevant earlier work suggesting that a "social gradient" of privacy risk may affect user willingness to participate in a DTC genetic test [41]. Future work could explore these ideas further to see if they generalize across larger sample populations. Further, it would be interesting to see if the relevant groups of people vary based on the population distributions in different countries.

6. Conclusion

In this paper, we presented results from a semi-structured interview study of 24 people who had used or considered using a DTC genetic testing service. We find that participants do not fully grasp some of the potential implications of sharing genetic information, e.g., the ethical ramifications of giving up familial genetic data. Nonetheless, most participants recognize many potential privacy concerns; however, the combination of perceived benefits of genetic testing and more general privacy resignation tends to outweigh these risks. We find that participants are willing to share genetic data for public benefit, but express interest in transparency and fine-grained control over such sharing (which may be difficult to achieve in practice).

References

[1] M. Backes, P. Berrang, A. Hecksteden, M. Humbert, A. Keller, and T. Meyer, "Privacy in epigenetics: temporal linkability of microRNA expression profiles," in *Proceedings of the 25th USENIX Security Sympsium*, 2016, pp. 1223–1240, https://www.usenix.org/system/files/conference/usenixsecurity16/sec16_paper_backes-privacy.pdf.

[2] M. Backes, P. Berrang, M. Humbert, and P. Manoharan, "Membership privacy in microRNA-based studies," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 319–330, http://doi.acm.org/10.1145/2976749.2978355.

[3] P. Ney, L. Ceze, and T. Kohno, "Genotype extraction and false relative attacks: Security risks to third-party genetic genealogy services beyond identity inference," in *Network and Distributed System Security Symposium*, 2020, *In press.* https://dnasec.cs.washington.edu/genetic-genealogy/ney_ndss.pdf.

[4] Y. Erlich, T. Shor, I. Pe'er, and S. Carmi, "Identity inference of genomic data using long-range familial searches." *Science*, vol. 362, no. 6415, pp. 690–694, 2018, http://dx.doi.org/10.1126/science.aau4832.

[5] L. Sweeney, A. Abu, and J. Winn, "Identifying participants in the personal genome project by name," 04/24/2013 2013, http://dataprivacylab.org/projects/pgp/.

[6] Y. Erlich and A. Narayanan, "Routes for breaching and protecting genetic privacy," *Nat. Rev. Genetics*, vol. 15, no. 6, pp. 409–421, 2014, https://doi.org/10.1038/nrg3723.

[7] B. Malin and L. Sweeney, "How (not) to protect genomic data privacy in a distributed network: using trail re-identification to evaluate and design anonymity protection systems," *J. Biomedical Informatics*, vol. 37, no. 3, pp. 179–192, 2004, https://doi.org/10.1016/j.jbi.2004.04.005.

[8] M. Fox, "Drug giant glaxo teams up with DNA testing company 23andMe," *NBC News*, 2018, https://www.nbcnews.com/health/health-news/drug-giant-glaxo-teams-dna-testing-company-23andme-n894531?cid=sm_npd_nn_tw_ma.

[9] P. Pitts, "The privacy delusions of genetic testing," *Forbes*, 2017, https://www.forbes.com/sites/realspin/2017/02/15/the-privacy-delusions-of-genetic-testing/#3cbaacf1bba5.

[10] A. McDonell-Parry, "Report: Genealogy company allowing FBI to use private DNA database," *Rolling Stone*, 2019, https://www.rollingstone.com/culture/culture-news/dna-fbi-sharing-privacy-database-788304/.

[11] T. Romm and D. Harwell, "Ancestry, 23andMe and others say they will follow these rules when giving DNA data to businesses or police," *The Washington Post*, 2018, https://www.washingtonpost.com/technology/2018/07/31/ancestry-andme-others-say-they-will-follow-these-rules-when-giving-dna-data-businesses-or-police/.

[12] A. Regalado, "More than 26 million people have taken an at-home ancestry test," *MIT Technology Review*, 2019, https://www.technologyreview.com/s/612880/more-than-26-million-people-have-taken-an-at-home-ancestry-test/.

[13] E. Lander, "Initial impact of the sequencing of the human genome," *Nature*, vol. 470, pp. 187–197, 2011, https://doi.org/10.1038/nature09792.

[14] R. Iwasiow, A. Desbois, and H. Birnboim, "Long-term stability of DNA from saliva samples stored in the oragene self-collection kit," DNA Genotek, Ottowa, ON, Tech. Rep., 2011, https://www.dnagenotek.com/US/pdf/PD-WP-005.pdf.

[15] A. Psifidi, C. Dovas, G. Bramis, T. Lazou, C. Russel, G. Arsenos, and G. Banos, "Comparison of eleven methods for genomic DNA extraction suitable for large-scale whole-genome genotyping and long-term DNA banking using blood samples," *PLoS One*, vol. 10, no. 1, p. e0115960, 2015, https://doi.org/10.1371/journal.pone.0115960.

[16] Thermo Fisher, "Direct-to-consumer (DTC) genetic ancestry reports: Why genotyping is essential," *Consumer Genetics Blog*, https://www.thermofisher.com/blog/behindthebench/direct-to-consumer-dtc-genetic-ancestry-reports-why-genotyping-is-essential/.

[17] B. Weir, A. Anderson, and A. Hepler, "Genetic relatedness analysis: modern data and new challenges," *Nat. Rev. Genetics*, vol. 7, no. 10, pp. 771–80, 2006, https://doi.org/10.1038/nrg1960.

[18] E. Ayday, "Cryptographic solutions for genomic privacy," *Financial Cryptography and Data Security*, vol. 9604, pp. 328–341, 2016, https://doi.org/10.1007/978-3-662-53357-4_22.

[19] H. Shaban, "DNA testing service myheritage says 92 million customer email addresses were exposed," *The Washington Post*, 2018, https://www.washingtonpost.com/news/the-switch/wp/2018/06/05/ancestry-service-myheritage-says-92-million-customer-email-addresses-were-exposed/.

[20] M. Marks and T. Li, "DNA donors must demand stronger protection for genetic privacy," *Stat*, 2018, https://www.statnews.com/2018/05/30/dna-donors-genetic-privacy-nih/.

[21] K. Norrgard, "Protecting your genetic identity: GINA and HIPAA," *Nature Education*, vol. 1, no. 1, p. 21, 2008, https://www.nature.com/scitable/topicpage/protecting-your-genetic-identity-gina-and-hipaa-678/.

[22] S. Fendrick, "The role of privacy law in genetic research," *I/S: A Journal of Law and Policy for the Information Society*, vol. 4, no. 3, pp. 803–820, 2008, http://hdl.handle.net/1811/72811.

[23] A. Phillips, "Think before you click: ordering a genetic test online," *Scitech Lawyer*, vol. 11, no. 2, pp. 8–11, 2015, http://www.andelkamphillips.com/wp-content/uploads/2018/02/Think-before-you-click.pdf.

[24] ——, "Only a click away - DTC genetics for ancestry, health, love... and more: a view of the business and regulatory landscape," *Applied & Translational Genomics*, vol. 8, pp. 16–22, 2016, https://doi.org/10.1016/j.atg.2016.01.001.

[25] T. Caulfield and A. McGuire, "Direct-to-consumer genetic testing: perceptions, problems, and policy responses," *Annual Review of Medicine*, vol. 63, no. 1, pp. 23–33, 2012, https://doi.org/10.1146/annurev-med-062110-123753.

[26] Federal Trade Commission, "At-home genetic tests: a healthy dose of skepticism may be the best prescription," 2006, FTC Consumer Information. http://www.ftc.gov/bcp/edu/pubs/consumer/health/hea02.shtm.

[27] G. Kutz, "Direct-to-consumer genetic tests: misleading test results are further complicated by deceptive marketing and other questionable practices," 2010, Government Accountability Office. Testimony Before the Subcommittee on Oversight and Investigations, Committee on Energy and Commerce, US House of Representatives. http://www.gao.gov/new.items/d10847t.pdf.

[28] A. Pollack, "F.D.A. faults companies on unapproved genetic tests," *The New York Times*, 2010, https://www.nytimes.com/2010/06/12/health/12genome.html.

[29] 23andMe, "23andMe, Inc. provides update on FDA regulatory review," 2013, Press release. https://mediacenter.23andme.com/press-releases/23andme-inc-provides-update-on-fda-regulatory-review/.

[30] ——, "23andMe, Inc. granted first FDA authorization to market direct-to-consumer genetic health risk reports," 2017, Press release. https://mediacenter.23andme.com/press-releases/23andme-inc-granted-first-fda-authorization-market-direct-consumer-genetic-health-risk-reports/.

[31] M. Herper, "Ancestry launches consumer genetics tests for health, intensifying rivalry with 23andMe," *Stat*, 2019, https://www.statnews.com/2019/10/15/ancestry-health-launch/.

[32] J. Jouvenal, M. Berman, D. Harwell, and T. Jackman, "Data on a genealogy site led police to the 'Golden State Killer' suspect. Now others worry about a 'treasure trove of data'," *The Washington Post*, 2018, https://www.washingtonpost.com/news/post-nation/wp/2018/04/27/data-on-a-genealogy-site-led-police-to-the-golden-state-killer-suspect-now-others-worry-about-a-treasure-trove-of-data/.

[33] D. M. Levitt, "Let the consumer decide? the regulation of commercial genetic testing," *Journal of Medical Ethics*, vol. 27, no. 6, pp. 398–403, 2001, https://jme.bmj.com/content/27/6/398.

[34] L. Goldsmith, L. Jackson, A. O'Connor, and H. Skirton, "Direct-to-consumer genomic testing: systematic review of the literature on user perspectives," *Eur J Hum Genet*, vol. 20, p. 811–816, 2012, https://doi.org/10.1038/ejhg.2012.18.

[35] J. S. Roberts, M. C. Gornick, D. A. Carere, W. R. Uhlmann, M. T. Ruffin, and R. C. Green, "Direct-to-consumer genetic testing: User motivations, decision making, and perceived utility of results," *Public Health Genomics*, vol. 20, pp. 36–45, 2017, https://doi.org/10.1159/000455006.

[36] E. De Cristofaro, "Users' attitudes, perception, and concerns in the era of whole genome sequencing," *CoRR*, vol. abs/1306.4962, 2014, http://arxiv.org/abs/1306.4962.

[37] C. Cheung, M. J. Bietz, K. Patrick, and C. S. Bloss, "Privacy attitudes among early adopters of emerging health technologies," *PLoS ONE*, vol. 11, no. 11, p. e0166389, 2016, https://doi.org/10.1371/journal.pone.0166389.

[38] A. F. Westin, "Social and political dimensions of privacy," *Journal of social issues*, vol. 59, no. 2, pp. 431–453, 2003.

[39] E. Vayena, C. Ineichen, E. Stoupka, and E. Hafen, "Playing a part in research? university students' attitudes to direct-to-consumer genomics," *Public Health Genomics*, vol. 17, pp. 158–168, 2014, https://doi.org/10.1159/000360257.

[40] E. Christofides and K. O'Doherty, "Company disclosure and consumer perceptions of the privacy implications of direct-to-consumer genetic testing," *New Genetics and Society*, vol. 35, no. 2, pp. 101–123, 2016, https://doi.org/10.1080/14636778.2016.1162092.

[41] T. Haeusermann, M. Fadda, A. Blasimme, B. G. Tzovaras, and E. Vayena, "Genes wide open: Data sharing and the social gradient of genomic privacy," *AJOB Empirical Bioethics*, vol. 9, no. 4, pp. 207–221, 2018, https://doi.org/10.1080/23294515.2018.1550123.

[42] K. Baig, R. Mohamed, A. Theus, and S. Chiasson, ""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*, ser. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 1–13, https://doi.org/10.1145/3313831.3376800.

[43] K. Caine, S. Kohn, C. Lawrence, R. Hanania, E. M. Meslin, and W. M. Tierney, "Designing a patient-centered user interface for access decisions about ehr data: Implications from patient interviews," *Journal of General Internal Medicine*, vol. 30, no. 1, pp. 7–16, Jan 2015, https://doi.org/10.1007/s11606-014-3049-9.

[44] K. Caine and R. Hanania, "Patients want granular privacy control over health information in electronic medical records," *Journal of the American Medical Informatics Association*, vol. 20, no. 1, pp. 7–15, 2012, https://doi.org/10.1136/amiajnl-2012-001023.

[45] C. Campos-Castillo and D. L. Anthony, "The double-edged sword of electronic health records: implications for patient disclosure," *Journal of the American Medical Informatics Association*, vol. 22, no. e1, pp. e130–e140, 07 2014, https://doi.org/10.1136/amiajnl-2014-002804.

[46] T. Stablein, J. L. Hall, C. Pervis, and D. L. Anthony, "Negotiating stigma in health care: disclosure and the role of electronic health records," *Health Sociology Review*, vol. 24, no. 3, pp. 227–241, 2015, https://doi.org/10.1080/14461242.2015.1078218.

[47] N. Shen, T. Bernier, L. Sequeira, J. Strauss, M. P. Silver, A. Carter-Langford, and D. Wiljer, "Understanding the patient privacy perspective on health information exchange: A systematic review," *International Journal of Medical Informatics*, vol. 125, pp. 1 – 12, 2019, http://www.sciencedirect.com/science/article/pii/S1386505618303009.

[48] L. Bishop, B. Holmes, and C. Kelley, "National consumer health privacy survey 2005," California HealthCare Foundation; Forrester Research, Inc., 2005, https://www.chcf.org/wp-content/uploads/2017/12/PDF-ConsumerPrivacy2005ExecSum.pdf.

[49] G. Guest, A. Bunce, and L. Johnson, "How many interviews are enough?: An experiment with data saturation and variability," *Field Methods*, vol. 18, no. 1, pp. 59–82, 2006, https://doi.org/10.1177/1525822X05279903.

[50] K. Charmaz, *Constructing Grounded Theory: A Practical Guide through Qualitative Analysis*. Thousand Oaks, CA: Sage Publications, Inc., 2006.

[51] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977, http://www.jstor.org/stable/2529310.

[52] MarketWatch, "DTC genetic testing market 2019: Industry Size, Trends and Growth by 2024 - GMI Report," 2019, Press release. https://www.marketwatch.com/press-release/dtc-genetic-testing-market-2019-industry-size-trends-and-growth-by-2024---gmi-report-2019-07-29.

[53] ——, "U.S. DTC genetic testing market size 2019: Application and Future Forecast by 2025," 2019, Press release. https://www.marketwatch.com/press-release/us-dtc-genetic-testing-market-size-2019-application-and-future-forecast-by-2025-2019-11-20.

[54] A. De Luca, S. Das, M. Ortlieb, I. Ion, and B. Laurie, "Expert and non-expert attitudes towards (secure) instant messaging," in *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. Denver, CO: USENIX Association, Jun. 2016, pp. 147–157, https://www.usenix.org/conference/soups2016/technical-sessions/presentation/deluca.

[55] A. Alqhatani and H. Lipford, "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)*. Santa Clara, CA: USENIX Association, aug 2019, pp. 421–434, https://www.usenix.org/system/files/soups2019-alqhatani.pdf.

[56] A. Braunstein, L. Granka, and J. Staddon, "Indirect content privacy surveys: Measuring privacy without asking about it," in *Proceedings of the Seventh Symposium on Usable Privacy and Security*, ser. SOUPS '11. New York, NY, USA: ACM, 2011, pp. 15:1–15:14, http://doi.acm.org/10.1145/2078827.2078847.

[57] K. Fulton, R. Gelles, A. McKay, Y. Abdi, R. Roberts, and M. Mazurek, "The effect of entertainment media on mental models of computer security," in *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. Santa Clara, CA: USENIX Association, Aug. 2019, https://www.usenix.org/conference/soups2019/presentation/fulton.

[58] C. P. Hoffmann, C. Lutz, and G. Ranzini, "Privacy cynicism: A new approach to the privacy paradox," *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, vol. 10, no. 4, 2016, http://dx.doi.org/10.5817/CP2016-4-7.

[59] R. Wash, "Folk models of home computer security," in *Proceedings of the Sixth Symposium on Usable Privacy and Security*, ser. SOUPS '10, 2010, pp. 11:1–11:16, http://doi.acm.org/10.1145/1837110.1837125.

[60] D. Solove, "'I've got nothing to hide' and other misunderstandings of privacy," *San Diego Law Review*, vol. 44, pp. 745–772, 2007, https://scholarship.law.gwu.edu/cgi/viewcontent.cgi?article=1159&context=faculty_publications.

[61] B. Stanton, M. Theofanos, S. Prettyman, and S. Furman, "Security fatigue," *IT Professional*, vol. 18, no. 5, pp. 26–32, 2016, https://inside.mines.edu/UserFiles/File/ccit/security/NIST-Security_Fatigue.pdf.

[62] F. Schaub, R. Balebako, A. Durity, and L. Cranor, "A design space for effective privacy notices," in *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. Ottawa: USENIX Association, Jul. 2015, pp. 1–17, https://www.usenix.org/conference/soups2015/proceedings/presentation/schaub.

[63] J. Reidenberg, T. Breaux, L. Cranor, B. French, A. Grannis, J. Graves, F. Liu, A. McDonald, T. Norton, R. Ramanath, N. Russell, N. Sadeh, and F. Schaub, "Disagreeable privacy policies: Mismatches between meaning and users' understanding," *Berkeley Technology Law Journal*, vol. 30, no. 1, pp. 39–88, 2015, http://dx.doi.org/10.2139/ssrn.2418297.

[64] L. Cranor, "Necessary but not sufficient: Standardized mechanisms for privacy notice and choice," *J. on Telecomm. & High Tech. L.*, vol. 10, pp. 273–307, 2012, http://jthtl.org/content/articles/V10I2/JTHTLv10i2_Cranor.PDF.

[65] A. Felt, E. Ha, S. Egelman, A. Haney, E. Chin, and D. Wagner, "Android permissions: User attention, comprehension, and behavior," in *Proceedings of the Eighth Symposium on Usable Privacy and Security*, ser. SOUPS '12. New York, NY, USA: ACM, 2012, pp. 3:1–3:14, http://doi.acm.org/10.1145/2335356.2335360.

[66] W. Hartzog, "The case against idealising control," *European Data Protection Law Review (EDPL)*, vol. 4, no. 4, pp. 423–432, 2018, https://ssrn.com/abstract=3299762.

[67] R. Kang, L. Dabbish, N. Fruchter, and S. Kiesler, "My data just goes everywhere: User mental models of the internet and implications for privacy and security," in *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. Ottawa: USENIX Association, Jul. 2015, pp. 39–52, https://www.usenix.org/conference/soups2015/proceedings/presentation/kang.

Appendix A. Screening Survey

Potential participants completed the following screening survey (online, through Qualtrics) to indicate interest participating in our study. The first page of the questionnaire was a consent form, outlined in A.1. If consent was given, the potential participant was directed to the screening questions (see A.2).

A.1. Consent [ONLINE]

Project Title: User Attitudes on Commercial DNA Testing
Purpose of the Study: This research is being conducted by Dr. Michelle Mazurek at the University of Maryland, College Park. We are inviting you to participate in this research project because we would like to understand the motivations and overall perceptions of existing and prospective DNA toolkit users with respect to DNA and its testing technologies.
Procedures: You will be asked to complete a short <5 minute questionnaire (administered through Qualtrics) for the purposes of screening for eligibility in the interview study. Based on the given responses, you may be selected to participate in the interview study. An investigator will contact you by email at the address you provide at during this survey if this is the case.

If you are invited, the study will comprise a 45-min in-person interview plus a short <5min questionnaire (to collect basic demographic information). This interview will be audio recorded.
Potential Risks and Discomforts: There are minimal risks to participating in this research study. Certain questions may make you uncomfortable, but you may skip any question as necessary, or stop participating at any time. If you skip questions, it may make you ineligible to participate in this study.
Potential Benefits: There are no direct benefits to you from participating in this research. However, possible benefits include having a better understanding of what can be learned from your DNA. We hope that, in the future, others might benefit from this study through improved understanding of the needs and concerns of users of commercial DNA test kits. We hope that this this understanding will help to eventually build robust privacy policies to protect against misuse of collected genetic information.
Confidentiality: Any potential loss of confidentiality will be minimized by storing data in a password protected computer. Additionally, your name will be mapped to a code (identification key) and this code, rather than your name will be used on all data stored. Only the researchers will have access to the name-identification key pairings.

If we write a report or article about this research project, your identity will be protected to the maximum extent possible. Your information may be shared with representatives of University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law.
Compensation: If you are selected and agree to participate in the interview, you will receive $20. You will be responsible for any taxes assessed on the compensation. (Compensation will not be dispensed for simply completing the screening questionnaire.) If you pay to park on the UMD campus in order to participate in the interview, we will reimburse you for the cost of parking for the duration for the study.

Your name and address will be collected to receive compensation.
Right to Withdraw and Questions: Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify.

If you are faculty, staff, or a student at UMD, your grades, standing and/or employability will not be positively or negatively affected by your decision to participate in this study.

If you decide to stop taking part in the study, if you have questions, concerns, or complaints, or if you need to report an injury related to the research, please contact the investigator(s):

Debjani Saha
5108 Iribe Center
College Park, MD 20742
dna.testing.study@cs.umd.edu

Participant Rights: If you have questions about your rights as a research participant or wish to report a research-related injury, please contact:

University of Maryland College Park
Institutional Review Board Office
1204 Marie Mount Hall
College Park, Maryland, 20742
E-mail: irb@umd.edu
Telephone: 301-405-0678

This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.
Statement of Consent: By selecting the appropriate boxes below, you indicate that you are at least 18 years of age; you have read the above consent form; you agree to audio recording of your interview; and you voluntarily agree to participate in this research study.

Again, please ensure you have made a copy of the above consent form for your records.

- ☐ I am age 18 or older
- ☐ I have read this consent form
- ☐ I agree to be audio recorded
- ☐ I voluntarily agree to participate in this research study

A.2. Screening Questions

Have you used or considered using a commercial DNA test kit (23andMe, Ancestry, etc.)?

☐ Yes
☐ No

If yes, please indicate your reasons to do so (optional): [short answer]

Please provide your email address below so the we can contact you for the interview, if you are selected:

Your contact information will only be used to invite you to participate in the study. After the study is completed, all records of your contact information will be destroyed.

Appendix B. Interview

Screening survey respondents who met our selection criteria were invited to participate in the interview study. Written consent was obtained from these participants at the beginning of the interview, as described below in B.1. Once consent was obtained, we proceed to the semi-structured interview, outlined in B.2. After the interview, we collected some demographic information by means of a online survey administered through Qualtrics (see B.3).

B.1. Consent [WRITTEN]

Project Title: User Attitudes on Commercial DNA Testing
Purpose of the Study: This research is being conducted by Dr. Michelle Mazurek at the University of Maryland, College Park. We are inviting you to participate in this research project because we would like to understand the motivations and overall perceptions of existing and prospective DNA toolkit users with respect to DNA and its testing technologies.
Procedures: You will be asked to participate in a 45-min in-person interview. This interview will be audio recorded. Audio recording is required to participate.
Here are a few sample questions that we may ask:

1) What is your opinion regarding these sorts of tests [commercial DNA tests]? Explain
2) How do you think these companies (23andMe, etc.) handle and use the genetic data they collect?

At the end of the interview, you will be asked to complete a short ($<$ 5 minute) questionnaire to collect some basic demographic information.
Potential Risks and Discomforts: There are minimal risks to you for participating in this research study. Certain questions may make you uncomfortable, but you may skip any question as necessary or stop participating at any time, either during the interview itself or during the post-interview survey.

During the screening survey, participants will be asked to provide us with personal information regarding their participation in commercial DNA testing and their email address. Additionally, during the post-interview survey, participants will be asked to provide basic demographic information. We will do our utmost to protect this data - all digital data (survey data and interview recordings) will be collected and stored using coded identifiers that allow linking the same participant's responses over time, but do not connect to the participant's identity. The identity link will be maintained temporarily in order to facilitate selecting and scheduling interview participants, but will be destroyed thereafter. Furthermore, all data will be stored on secure, password protected University of Maryland and Qualtrics servers. Any paper data (signed consent forms, interviewer notes, etc.) will be kept in a locked cabinet in Dr. Mazurek's office. However, there is always a chance that participant's personal data may be compromised.
Potential Benefits: There are no direct benefits to you from participating in this research. However, possible benefits include having a better understanding of what can be learned from your DNA. We hope that, in the future, others might benefit from this study through improved understanding of the needs and concerns of users of commercial DNA test kits. We hope that this understanding will help to eventually build robust privacy policies to protect against misuse of collected genetic information.
Confidentiality: Any potential loss of confidentiality will be minimized by storing data in a password protected computer. Additionally, your name will be mapped to a code (identification key) and this code, rather than your name will be used on all data stored. Only the researchers will have access to the name-identification key pairings.

If we write a report or article about this research project, your identity will be protected to the maximum extent possible. Your information may be shared with representatives of University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law
Compensation: You will receive $20. You will be responsible for any taxes assessed on the compensation. If you pay to park on the UMD campus in order to participate in the interview, we will reimburse you for the cost of parking for the duration for the study.

Your name and address will be collected to receive compensation.
Right to Withdraw and Questions: Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify.

If you are faculty, staff, or a student at UMD, your grades, standing and/or employability will not be positively or negatively affected by your decision to participate in this study.

If you decide to stop taking part in the study, if you have questions, concerns, or complaints, or if you need to report an injury related to the research, please contact the investigator(s):

Debjani Saha
5108 Iribe Center
College Park, MD 20742

dna.testing.study@cs.umd.edu

Participant Rights: If you have questions about your rights as a research participant or wish to report a research-related injury, please contact:

University of Maryland College Park
Institutional Review Board Office
1204 Marie Mount Hall
College Park, Maryland, 20742
E-mail: irb@umd.edu
Telephone: 301-405-0678

This research has been reviewed according to University of Maryland, College Park IRB procedures for research involving human subjects.

Statement of Consent: Your signature indicates that you are at least 18 years of age; you have read this consent form or have had it read to you; your questions have been answered to your satisfaction and you voluntarily agree to participate in this research study. You will receive a copy of this consent form for your records.

If you agree to participate, please sign your name below.

Signature and Date:

NAME OF PARTICIPANT [Please Print]:

SIGNATURE OF PARTICIPANT:

DATE:

B.2. Interview Questions

Experience

On your screening survey, you noted that you [TOOK/ HEARD OF] [GENETIC TEST NAME].
If they took the test:
Can you tell me about your experience?
Otherwise:
What is your opinion regarding these sorts of tests? Explain.
Why do you think people usually choose to take these sorts of tests?
Why did you choose to participate/not participate?

Benefits and Drawbacks

What do you think are the benefits of participating in these sorts of tests?

How valuable do you think [BENEFIT] is?

Do you think there are any drawbacks or concerns to these tests?
Do you personally have any concerns about these sorts of tests? Why or why not?

What are your concerns?
Based on their concerns: How serious was your concern about [CONCERN]?

If they took the test:
According to your screening questionnaire, you did in fact take one of these tests. Walk me through your decision to do so (in spite of your concerns).
Otherwise:
Did the concerns you listed play a part in preventing you from taking the test? Elaborate.
[If the participant does not specifically mention it, ask if cost of testing played any role in their decision to take/not take a test.]

Knowledge

How do you think commercial genetic testing works?
[If they don't mention DNA: They work by taking a sample of your DNA and testing it.]
What do you know about DNA?
What do you think your DNA says about you?
Where do you get your DNA from?
Based on this information, what do you think can be inferred from your genetic makeup?
[If unsure, mention lineage, predictive genetic testing for disease markers, etc.]
Do you think anyone else should be involved in your decision to take a commercial DNA test? Why or why not?

Followup: Do you think your DNA relays information about more than just yourself?
[If they don't mention hereditary information/inference: Your DNA is passed down from your parents (half and half), who in turn get it from their parents, and so on.]

Privacy

What do you think are a company's (23andMe, Ancestry, etc.) responsibilities towards their customers upon data collection?
How do you think these companies handle and use the genetic data they collect?
Is there anything in particular you think these companies should or should not do?
Possible followup points:

How should data be stored?
How long do you think the data is kept?
Do they do their own in-house research and analysis (unrelated to the services they provide to customers)?
Do they sell/share it with third parties?
If so, how in what form is the data shared (anonymized, etc.)?
What sorts of third parties may be recipients?

For the things that you think these companies should not do: why do you feel this way?
Are there some uses of the collected genetic data that you feel are more appropriate than others? Why or why not?
What are your feelings towards...

The company (23andMe, Ancestry, etc.) having your genetic information? Explain.
Your doctors having/seeing your genetic information? Explain.
Your boss having/seeing your genetic information? Explain.
Your friends having/seeing your genetic information? Explain.
Your insurance provider having/seeing your genetic information? Explain.
Pharmaceutical and/or medical device companies having/seeing your genetic information? Explain.

B.3. Demographic Information [ONLINE]

Please specify the gender with which you most closely identify:
- ☐ Male
- ☐ Female
- ☐ Other
- ☐ Prefer not to answer

Please specify your age:
- ☐ 18-29
- ☐ 30-39
- ☐ 40-49
- ☐ 50-59
- ☐ 60-69
- ☐ Over 70

Please specify your ethnicity (you may select more than one):
- ☐ White
- ☐ Hispanic or Latinx
- ☐ Black or African American
- ☐ American Indian or Alaska Native
- ☐ Asian, Native Hawaiian, or Pacific Islander
- ☐ Other

Please specify below which country/state/province you live in:

Please specify the highest degree or level of school you have completed:
- ☐ Some high school credit, no diploma or equivalent
- ☐ High school graduate, diploma or the equivalent (for example: GED)
- ☐ Some college credit, no degree
- ☐ Trade/technical/vocational training
- ☐ Associate degree
- ☐ Bachelor's degree
- ☐ Master's degree
- ☐ Professional degree
- ☐ Doctorate degree

If you are currently a student or have completed a college degree, please specify your field(s) of study below (e.g. Biology, Computer Science, etc):

Please select the response option that best describes your current employment status:
- ☐ Working for payment or profit
- ☐ Unemployed
- ☐ Looking after home/family
- ☐ Student
- ☐ Retired
- ☐ Unable to work due to permanent sickness or disability
- ☐ Other:

Please specify the range which most closely matches your total, pre-tax, household income in 2018.
- ☐ $<$ \$29,999
- ☐ \$30,000-\$49,999
- ☐ \$50,000-\$74,999
- ☐ \$75,000-\$99,999
- ☐ \$100,000-\$124,999
- ☐ \$125,000-\$149,999
- ☐ \$150,000-\$199,999
- ☐ $>$ \$200,000

To your best estimation, how did your average income in the past five years compare to that in 2018?
- ☐ Higher
- ☐ Lower
- ☐ No change
- ☐ Don't know/variable income

We will maintain privacy of the information you have provided here. Your information will only be used for data analysis purposes.