

DEVELOPING PRIVACY ENHANCING TECHNOLOGY FOR DNA DATA
SHARING IN PUBLIC GENEALOGY PLATFORMS

by

Lipsarani Sahoo

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Computing and Information Systems

Charlotte

2023

Approved by:

Dr. Mohamed Shehab

Dr. Heather Lipford

Dr. Weichao Wang

Dr. Rich Lambert

ABSTRACT

LIPSARANI SAHOO. Developing Privacy Enhancing Technology For DNA Data Sharing In Public Genealogy Platforms. (Under the direction of DR. MOHAMED SHEHAB)

At-home DNA testing and sharing in public genealogy databases are becoming widespread. This will facilitate finding out ancestry, genetic relatives, biological parents, making new connections, advancing medicine, and determining predisposition to various diseases and health issues. While the biomedical community glorifies the uses of the genomics revolution, the expanded obtainability of such sensitive data has substantial implications for individual privacy as genes carry sensitive personal information about human traits and predispositions to any diseases. Furthermore, DNA data has identification capability (e.g., forensics) as well as reveals familial interconnections. However, commercial DNA testing is not vigorously governed by any laws and policies. The privacy implications of public DNA data sharing remain largely unexplored. This dissertation explores users' privacy concerns and proposes a method for communicating the risks to users to inform users when sharing their DNA data.

In the first study, we explored users' perceptions regarding DNA data. We asked about their views of at-home DNA testing and sharing, followed by their expected benefits and concerns. We also talked about public genealogy databases like GEDmatch. We focused on understanding the users' preferences and perceptions on the disclosure of their genetic information under the different types of platforms and entities. Our results show that users are mostly unaware and uncomprehending of the interconnected nature of genetic data. We noted users' general perceptions and focused on understanding their preferred privacy controls while sharing their DNA data, their desired settings, policies, and rules [1].

From this study, we identified the need to develop a privacy-enhancing technology such that the users can make an informed choice while sharing DNA data. We

also found that several policies and settings should be to preserve the privacy of sensitive data. With these findings in mind, the ultimate objective of this dissertation is to design and implement privacy risk communication methods that aid users in comprehending the risks and benefits associated with sharing DNA data, as well as enhancing transparency in access control. To evaluate the effectiveness of our developed risk communication approach, we deployed it within an existing platform, allowing us to assess users' decision-making processes and gain a deeper understanding of the nature of DNA data.

ACKNOWLEDGEMENTS

At this pivotal moment of completing my doctoral thesis, I am profoundly grateful for the invaluable support, guidance, and encouragement I have received from a multitude of individuals who have played significant roles throughout this transformative journey. First and foremost, I humbly convey my utmost gratitude to the almighty for granting me the strength and fortitude to embark on this challenging journey. In expressing my deepest appreciation, I am profoundly indebted to my beloved husband, Manas Sahoo, whose unwavering faith in my abilities and unwavering support have been the cornerstone of my accomplishments. I extend my heartfelt thanks to my cherished children, Sankalp and Misha, whose patience, understanding, and unconditional love have been a constant source of inspiration throughout this remarkable undertaking.

I am deeply grateful to my academic advisor, mentor Dr. Mohamed Shehab, whose expertise, guidance to my growth as a researcher have been instrumental in shaping the trajectory of my doctoral journey. My sincere appreciation also goes to Dr. Heather Lipford, my role model, whose guidance, insights, and support during the crucial stages of my dissertation were invaluable. Additionally, I would like to express my deep appreciation to Dr. Mary Lou Maher for her role as a teaching mentor, providing valuable guidance and wisdom throughout my academic journey. I extend my gratitude to my committee members, Dr. Richard Lambert and Dr. Weichao Wang, for their valuable contributions, insightful comments, and suggestions that have greatly enhanced the quality of my research.

I would like to extend my deepest gratitude to my dear friends for their support, encouragement, and camaraderie throughout this remarkable journey. I am deeply grateful for the tremendous support and encouragement I have received from my family members. Their belief in my abilities has been an invaluable source of motivation and inspiration. I am grateful to my colleagues Seetha, Sakib, Yousra, Elham, Jeba,

Jacqueline, and others for their collaboration, intellectual discussions, and support during my academic pursuits. I would also like to express my gratitude to Sandra Cruise for her assistance in navigating various challenges has been invaluable.

Finally, I would like to express my gratitude to UNC Charlotte for providing an enriching academic environment and for the incredible opportunities and experiences that have shaped me as a researcher and an individual. I am immensely proud to be a 49er.

TABLE OF CONTENTS

LIST OF TABLES	xii
LIST OF FIGURES	xiv
CHAPTER 1: INTRODUCTION	1
1.1. Commercial DNA testing & Sharing	1
1.1.1. What is DNA and how it is shared?	2
1.1.2. Direct To Consumer Companies, Public genealogy database: a Brief Introduction	2
1.1.3. Speciality of DNA Data	5
1.2. Potential Privacy Risks of sharing genetic data	6
1.2.1. Commercial DNA tests and police investigations	6
1.2.2. Informed Consent	8
1.2.3. Commercial DNA tests and Health Data	9
1.3. DNA Sharing: Current policies and Risk communication	10
1.4. Problem Statement and Proposed Contributions	11
CHAPTER 2: BACKGROUND	17
2.1. Privacy calculus theory	17
2.2. DNA data	19
2.2.1. Users' Perceptions of at-home DNA testing	19
2.2.2. DNA sharing Risks	20
2.2.3. User generated health data; Self disclosure	22
2.3. Users' Decision making and awareness	23
2.3.1. Content of risk communication	24

2.3.2.	Mechanisms to communicate personal / sensitive data disclosure to users	28
2.4.	Summary	33
CHAPTER 3:	Exploring Users' perceptions of at-home DNA testing and sharing of DNA data online [1]	35
3.1.	Introduction	35
3.2.	Methodology	36
3.2.1.	Recruitment & Demographics	36
3.2.2.	Procedure & Analysis	37
3.3.	Results	38
3.3.1.	Pre-Introduction of DNA testing: Non-Experienced Group	38
3.3.2.	Post-Introduction of DNA testing: Non-Experienced Group	39
3.3.3.	Users' experience: Experienced Group	41
3.3.4.	Post GEDmatch Demo	42
3.3.5.	DTC-GT VS GEDmatch	46
3.3.6.	Scenarios: DNA data Sharing	46
3.3.7.	Scenarios Effects	51
3.3.8.	Lack of Knowledge	53
3.3.9.	Race & Nation: DNA data sharing	53
3.3.10.	Future expectation DNA sharing/ Future Motivation	54
3.3.11.	Users' Suggestions	55

3.4. Discussion and Limitation	57
3.4.1. Privacy perceptions	57
3.4.2. Privacy trade-off	59
3.4.3. Attitude differences	60
3.4.4. Limitations	61
3.5. Summary	61
CHAPTER 4: Study 2 - Participatory design for privacy of online DNA data sharing	63
4.1. Introduction	63
4.2. Phase 1 - Need-Finding and Co-Design	65
4.2.1. Recruitment & Demographics	67
4.2.2. Procedure & Analysis	67
4.2.3. Results	68
4.2.4. Implication of Phase 1: Suggestions to initial designs	70
4.3. Phase 2 - Designs iterations phase	72
4.3.1. Methodology	72
4.3.2. Recruitment & Demographics	79
4.3.3. Procedure & Analysis	79
4.4. Design Versions 1 Feedback Results	80
4.5. Design Versions 2 and Final designs	83
4.6. Phase 3 - Comparing the designs	85
4.6.1. Methodology	86

	x
4.7. Evaluation	87
4.7.1. Demographics	87
4.7.2. Evaluation of risk communication	88
4.7.3. Feedback about the video:	96
4.7.4. Feedback about the infographics:	97
4.8. Discussion	98
4.9. Limitations	100
4.9.1. Conclusion	101
CHAPTER 5: Study 3 - The Effectiveness of risk communication for privacy of online sharing of genetic data	102
5.1. Introduction	102
5.2. Methodology	103
5.3. Analysis	105
5.4. Results	106
5.4.1. Understanding about at-home DNA testing	107
5.4.2. Interest in taking at-home DNA testing	107
5.4.3. Perceived Benefits of at-home DNA testing	109
5.4.4. Perceived Concerns of at-home DNA testing	110
5.4.5. Interest in sharing DNA data in GEDmatch	111
5.4.6. Perceived Benefits and risks of GED match	113
5.4.7. Intention of sharing their personal information in GEDmatch:	114
5.4.8. Privacy settings: GEDmatch	118

5.4.9. Understanding of potential access to the DNA data	121
5.4.10. Willingness to share DNA data	124
5.4.11. Understanding the issue of Informed consent	127
5.5. Limitations	128
5.6. Discussion	128
5.7. Conclusion	130
CHAPTER 6: Discussion and Conclusion	132
6.1. Contributions	137
6.2. Design Implications	139
6.3. Future Work	141
6.4. Conclusion	142
REFERENCES	144
CHAPTER 7: Appendix	153
7.1. APPENDIX A: Supplementary data for Chapter 3	153
7.1.1. Screening Survey	153
7.1.2. Interview Questions	153
7.1.3. Demographics of the participants	160
7.2. APPENDIX B: Supplementary data for Chapter 4	163
7.2.1. Video transcripts	163
7.3. APPENDIX C: Survey questions for Chapter 5	167

LIST OF TABLES

TABLE 3.1: Post-Introduction of DNA testing: Concerns discussed by NE participants	40
TABLE 3.2: Motivation behind taking the test	41
TABLE 3.3: Perceived benefits and concerns about sharing DNA data on GEDmatch	42
TABLE 3.4: Expected use of DNA data	44
TABLE 3.5: Concerns regarding law enforcement access to DNA data	47
TABLE 3.6: Privacy-enhancing suggestions by participants	56
TABLE 4.1: Participants' suggestions of the possible risks that must be included in the risk and benefit message	70
TABLE 4.2: Participants' suggestions of the possible benefits that must be included in the risk and benefit message	70
TABLE 4.3: Suggestions on Privacy enhancing settings for DNA sharing platforms	71
TABLE 4.4: Participants' designs suggestions	72
TABLE 4.5: Participants' feedback and suggestions on "Design Version 1"	82
TABLE 4.6: Participants' feedback and suggestions on "Design Version 2"	84
TABLE 4.7: Evaluating the informative message between those who watched a video message and info-graphic message including Mean, Median, and test statistic values of the Mann-Whitney with the reported p-value. The message refers to either Info-graphic or Video	91
TABLE 4.8: Evaluating the "Content Recall" between those who watched a video message and info-graphic message including percentages and test statistic values of the Mann-Whitney with the reported p-value.	92
TABLE 5.1: Interest in taking about at-home DNA testing	109
TABLE 5.2: Perceived Concerns of at-home DNA testing	111

TABLE 5.3: Interest in sharing DNA data in GEDmatch	113
TABLE 5.4: Intention of sharing their personal information in GEDmatch	115
TABLE 5.5: Privacy setting options chosen by participants	119
TABLE 5.6: Understanding of potential access to the DNA data	122
TABLE 5.7: Willingness to share DNA data	126
TABLE 7.1: Experienced Group	161
TABLE 7.2: Non-experienced Group	162

LIST OF FIGURES

FIGURE 1.1: One to many comparison result [2]	3
FIGURE 1.2: One to one comparison result [2]	4
FIGURE 4.1: Screenshots from the slides used in the interview	66
FIGURE 4.2: Personal Story Video	75
FIGURE 4.3: Conversational story video	75
FIGURE 4.4: Informational story video	76
FIGURE 4.5: Informational story video	77
FIGURE 4.6: Wizard flow	78
FIGURE 4.7: Participants' feedback On designs: Relatability, Engaging and Enjoyable and Easy to Understand	81
FIGURE 5.1: Study flow	105
FIGURE 5.2: Form field for Opt In/Opt Out policy with textual explanation	106

CHAPTER 1: INTRODUCTION

At-home Deoxyribonucleic acid (DNA) testing, or commercial DNA testing, has recently gained popularity. These tests are readily available and accessible. At-home DNA testing companies provide information about a person's ancestry, health predispositions, and traits. For example, an adoptee could be able to find their biological family. Moreover, someone could identify health predispositions to prompt them to reduce risks of some diseases.

Furthermore, this genetic information can be uploaded to different public online services like GEDmatch to find genetic relatives. GEDmatch is a free genealogy site that is publicly searchable and includes real names. This ability to publicly share DNA data raises new privacy issues and manage end-user privacy reward while leveraging these tests to contribute to the treatment of diseases; finding out suspected genetic conditions and genetic disorders will be imminent. Our primary goal in this dissertation is to help users understand the risks and benefits of genetic data sharing on such public platforms and develop and deploy effective privacy-enhancing technology or communicate that information.

1.1 Commercial DNA testing & Sharing

Commercial DNA testing is primarily used to find genetic relatives, ancestry, and health predispositions. The commercial tests analyze genes in the human body to predict health risk for conditions such as heart disease, determine disease carrier status and specify traits. These genetic tests are advertised directly to customers via the internet, television, or print advertisements. These test kits can be procured online or in stores. Consumers mail the company a DNA sample such as saliva and

receive their reports directly from the website or through written reports. These tests do not necessarily involve a healthcare provider or health insurance company in the process. The following subsections provide a brief explanations on DNA, commercial DNA testing, and DNA sharing platforms.


1.1.1 What is DNA and how it is shared?

DNA is a molecule that holds the entire biological instructions that an individual is made up of. Humans have 23 pairs or forty-six chromosomes. DNA is inherited from ancestors during reproduction. Both parents pass a set of twenty-two chromosomes and one sex chromosome to their offspring, which combine to make the whole genome. Therefore, genetic makeup is an equal blend of the mother's and father's genes in humans. The first twenty-two chromosome pairs are called autosomes, numbered chronologically from one to twenty-two, controlling the inheritance of an individual's characteristics. The last pair, heterochromosome, consists of two sex-deciding chromosomes marked either XX or XY. DNA Matching is the process of sequencing one individuals' DNA and comparing it to other individuals' DNA. When someone with a substantial part of DNA matching is found, that can imply that these two individuals share a common ancestor. For example, a person inherits 50% of DNA from each of their parents, 25% from each grandparent, 12.5% from each great-grandparent, etc. The DNA inheritance percentage decreases with each preceding generation.

1.1.2 Direct To Consumer Companies, Public genealogy database: a Brief Introduction

DNA profiling first started in the 1980s, and it has been highly thriving for testing in crime scenes, predisposition to diseases, and confirming paternity. A sample such as a blood specimen will be collected for testing. Then, genealogy research became popular in the United States early nineteenth century. It has been done out of curiosity

to discover one's ancestry, build a family tree, and reveal medical issues or genetic traits. Computerized innovation and the Internet have given speedy, simple, and convenient access to the tools to deliver those solutions. Commercial DNA or at-home DNA testing companies like 23andMe, and Ancestry.com fast emerged and became widespread. These Direct to Consumer companies (DTC) supply home testing kits. The consumer is required to mail a sample such as saliva or a piece of hair. Then the analysis is performed on autosomal DNA to look at specific locations of the person's genome to find ethnicity, ancestry, and health reports. Users can print or view a result that reveals information about their ancestry, health predisposition, wellness, and carrier status for any disorder. Additionally, the user can also download raw DNA data in zipped (.zip) text files that can be uploaded and analyzed by third-party tools, such as GEDmatch and Genetic Genie, that locate genetic matches. The number and kind of online services available to individuals are growing, and people are getting curious to take these tests. In 2021, the genetic genealogy testing market was comprised of over 26 million customers [3].


Tools for DNA & Genealogy Research

When you contact a potential match, you should include in your email your kit number and the kit number you think you match with. Many people manage multiple kits, and they have no other way of confirming the match you think you have.

It is generally considered to be bad form to send out emails to large numbers of names on your results list, without first reviewing those results in some detail. This falls into the category of 'Spam' email.

'Overlap' is the number of positions that exist in common between both kits, without regard to whether they match or not. The amount of overlap, along with the largest cM amount, is usually a good indication of the relative quality of the match.

Matches with low overlap have that field highlighted with a pink or red background, depending on the overlap value.

Matches with very low overlap are not shown.

[Here](#) is a link to a useful YouTube video on how to use One To Many.

Kit: 13532 (*BJ-m4) [Migration - V4 - M]

Kit	[1:1]	Name	Email	Largest Seg	Total cM	Gen	Overlap	Date Compared	Testing Company
9158	A	*AL	@gmail.com	214.5	3571.8	1.0	N/A	2018-06-30	
7649	A	andy	@yahoo.com	151.8	3570.9	1.0	149684	2020-03-01	
2216	A	*Phil Garnett	@gmail.com	151.9	3569.5	1.0	144089	2019-04-21	Ancestry

Figure 1.1: One to many comparison result [2]

GEDmatch is considered an open data personal genomics database. It is a free genealogy public site. GEDmatch offers features to upload raw genetic data results

GEDmatch[®] Autosomal One-to-one Comparison - V1.0

Software Version May 7 2021 01 21 15

[Here](#) is a link to a useful YouTube video on using the One to One DNA comparison tool.

Individual marker indications:

Base Pairs with Full Match	Green
Base Pairs with Half Match	Yellow
Match with Phased data	Purple
Base Pairs with No Match	Red

Validity of segments:

Significant	Blue
Moderate	Purple
Low	Pink
Insignificant	Orange
Large gap between adjacent SNPs	Light Orange
No Match	Black

Comparing Kit M123532 (*BJ-m4) [Migration - V4 - M] and Kit A349158 (*AL)

Segment threshold size will be adjusted dynamically between 200 and 400 SNPs

Minimum segment cM to be included in total = 7.0 cM

Mismatch-bunching Limit will be adjusted dynamically to 60 percent of the segment threshold size for any given segment.

Chr	B37 Start Pos'n	B37 End Pos'n	Centimorgans (cM)	SNPs
1	752,721	13,176,463	29.1	1,567
1	13,775,830	120,523,902	120	10,823
1	149,854,324	249,210,707	128.6	10,715

Chr 1



Figure 1.2: One to one comparison result [2]

from various genetic testing companies. Then GEDmatch identifies possible family members who have also uploaded their genetic data. By January 2021, the GEDmatch database had over 4 millions customers. Nelson et al. [4] found that GEDmatch was the most commonly used tool (84% of participants used GEDmatch), and participants in their study agreed that GEDmatch provided ancestry and relative information and helped them understand and interpret genetics. GEDmatch has different tools to analyze users' DNA files. For example, users can use the "One-to-Many Comparison Result" to search for relative matches within their GEDmatch database. It delivers a list of the 3000 closest matches with their names, email addresses, kit numbers, along with testing company (see Figure 1.1). As well, GEDmatch provides a "One-To-One

comparison tool" that can be used to analyze the DNA of two individuals on a one to one basis (see Figure 1.2). As these comparison results render names and email addresses, an interested person can contact their matches.

The GEDmatch forum is a platform for users to share information and network. "Search all GEDCOM" is another popular tool. In this tool, users can just put someone's first name and last name to get their details such as place of birth and death, father, mother and children. The feature of the pedigree chart and descendants can be used to look at the family tree [5].

1.1.3 Speciality of DNA Data

Genetic data is unique and an ultimate identifier of a living organism. As discussed above, DNA is inherited or shared in the family. DTC such as Ancestry.com and 23andme.com already provide genealogical services and health reports based on DNA testing. As a substantial part of a person's genome is shared with genetic relatives, other family members' predispositions to hereditary disorders could be implicated. Beyond the person who shared their own genetic data willingly, the question of informed consent is the most important. The genetic relatives have not shared their data but could be still identified and sensitive data like health data could still be inferred. This is utterly privacy-invasive. Another special feature of DNA relates to its proficiency in diagnosing health and behavior problems. Tests can indicate an increased probability for conditions such as Alzheimer's (the most typical form of dementia). This can have diagnostic significance as well as privacy ramifications. For instance, if a person's family learned about the person's increased chance of Alzheimer's, they might trust the person's judgment a little less, consciously or unconsciously. When considering privacy implications, these special apprehensions about sharing genomic data cannot be ignored. Hence, DNA data warrants special cautiousness.

1.2 Potential Privacy Risks of sharing genetic data

We envision various potential privacy risks of DNA data sharing, including informed consent. We discuss these in details below.

1.2.1 Commercial DNA tests and police investigations

Joseph James DeAngelo Jr. or the “Golden State Killer”(born November 8, 1945) is an American serial killer, serial rapist, burglar, and a former police officer who committed at least 13 murders, 51 rapes, and 120 burglaries across California between 1974 and 1986. After the decades-long investigation, on April 24, 2018, the State of California charged DeAngelo based upon DNA evidence; investigators found his fifth cousin on GEDmatch by creating a fake profile and using the crime site DNA sample. Then through forensic genetic genealogy or familial search, they captured the Golden State Killer. When law enforcement agencies had used GEDmatch in the Golden State Killer case, many people conveyed concern that the database was being used without the informed authorization of the users. The owners themselves had not been notified of the advancement of this use. Nevertheless, the GEDmatch site policy, which was introduced after this inflated debate in August 2017, was very broad. Although it did not expressly authorize access by law enforcement agencies, it did envision unanticipated usages: "While the results presented on this site are intended solely for genealogical research, we are unable to guarantee that users will not find other uses." On 28 April 2018, GEDmatch issued a statement on the website to alert participants about the use of the database by law enforcement agencies. Subsequently, on May 20, 2018, GEDmatch revised the Terms of Service and Privacy Policy to permit law enforcement to identify a perpetrator of a violent crime against another person; or identify remains of deceased individuals.

In January 2019, FamilyTreeDNA (FTDNA, another DNA sharing site) declared that they were cooperating with the Federal Bureau of Investigation(FBI). Now,

FTDNA authorizes the FBI to upload DNA profiles and create accounts with the exact level of access as regular users. All users, including existing ones, could choose to opt-out of the matching, but this would mean they would not obtain the benefits and services they had already paid for. It was later revealed that the FBI had already been accessing the FTDNA database for an undetermined time without the company's knowledge. Following public outrage, FTDNA introduced an opt-out from law enforcement matching in March 2019. Further, FTDNA has an international database, and non-US customers should be mandated to opt-in instead of opt-out. As with GEDmatch, there are also apprehensions about the participation of children and other genetic relatives in law enforcement matching or police inquiries. FTDNA's policy communicates that participants must be 13 years of age to participate in the database. Juveniles between the ages of 13 and 18 can only be tested with a parent or custodian's consent. But, it is not known how or if these guidelines are enforced, implemented, or accomplished. The other genetic genealogy testing firms like MyHeritage permit uploads from other testing companies, but law enforcement agencies have to obtain a court order or valid legal documentation to use their database. However, it is relatively feasible that the raw data files could be manipulated for upload, and the company would not realize or discover that they were processing law enforcement's manipulated files. Similarly, AncestryDNA and 23andMe do not accept transfers and do not permit law enforcement agencies to access their databases unless required by a valid legal process. However, It is not evident or transparent how the companies are able to monitor or implement their terms and conditions. It is possible likely that the guidelines could be breached without their knowledge, particularly provided that the FBI previously uploaded profiles to both GEDmatch and FTDNA before the changes in the terms and conditions which explicitly allowed such uploads.

1.2.2 Informed Consent

One issue is informed consent as DNA in a family is related to each other, so when a person choose to share their DNA data, they implicitly share the DNA data of others' in family without their consent. Many users also upload raw data files on behalf of other family members, and thus in some cases fully informed consent may not have been obtained. There is no established inspection on whether the DNA was uploaded with permission or not. GEDmatch's site policy forbids the service of the database by children under 13, but has no restrictions on minors aged between 13 and 18. Moray et al. [6] revealed that many genetic ancestry companies do not adequately address the issue of the testing of children and that fathers can potentially use DTC databases for confidential paternity testing. Thus, we can infer that there is a potential threat that children could be added to these databases and included in police investigations and matchings without their authorization or approval. Even if an individual has not taken a DNA test himself or has tested but has chosen not to share their results on GEDmatch, it could still be incorporated in an investigation because their sibling, cousin, or genetic relative has shared. It implies that people's decisions to share their own genetic information inadvertently reveal others across their family tree who may not be aware of, or interested in, their genetic relationships going public.

The genetic genealogy databases are multinational. With extended lineage members living around the globe, the decision of a person in one country to take a DNA test could mean that a genetic relative in another nation becomes involved in an investigation. For instance, in a recent case in Canada, immigration officers tested the DNA of a refugee by the name of Frank Goodwin at the genetic sharing company FTDNA. Two of his immediate relatives in the UK were in the company's database, and the officers reached them in an attempt to determine Goodwin's nationality. Therefore, concerns still stay about the lack of informed consent for participation in law enforcement matching. Moreover, as discussed in the previous section, the fam-

ily shares a significant portion of DNA, enabling family searching. DNA tests also reveal hereditary diseases. This could lead to exposure to the family's health history, conditions, or predispositions. This would lead to revealing other family members' health conditions on the internet or third parties without their consent.

1.2.3 Commercial DNA tests and Health Data

Today, numerous genetic tests are available to people at the clinical and consumer level. Most of the time, physicians use genetic testing to help ensure a genetic condition diagnosis in patients undergoing specific symptoms. As per the National Institutes of Health (NIH), genetic tests can be used to determine 2,000 hereditary conditions and diseases. For instance, these tests might find common hereditary diseases, including cystic fibrosis, familial hyperlipidemia, and muscular dystrophy. One of the advantages of these genetic tests is that they allow the detection of hereditary diseases at every stage of life, perhaps to prevent them. In some cases, if doctors can determine the disease before symptoms progress to an intense status, they can support patients in managing the illness. Nowadays, at-home genetic tests can provide information about whether consumers mutations associated with several types of hereditary cancer. This includes breast cancer too. As another example; one prominent cancer-related mutation that a DNA have check for is the BRCA 1/BRCA 2 gene mutation. If someone including has one of the BRCA genes mutations, the risk of developing breast cancer is very high. Currently, at-home DNA testing companies like 23andMe provide details about several medical illnesses, including celiac disease, Parkinson's disease, Late-onset Alzheimer's (a progressive brain condition that impacts memory), and many more. DTCs such as 23andMe also detect carriers for cystic fibrosis diseases. Thus, sharing DNA could lead to unintended disclosure of a variety of health data.

1.3 DNA Sharing: Current policies and Risk communication

Privacy policies concerning protecting genetic data in the United States depend on where the data is and what it is used for. The primary laws governing genomic data are the Genetic Information Nondiscrimination Act of 2008 (GINA) [7], and the Health Insurance Portability and Accountability Act of 1996 (HIPAA) [8]. This approach of irregular disconnected coverage is defective because it does not account for the complexity of DNA and its potential for being misused. According to HIPAA, if the individual provides the information to a primary care physician, it becomes "personal health data" and is governed by HIPAA [8]. Under HIPAA, genetic information cannot be disclosed to schools or employers. Still, law enforcement is authorized to access it for investigation purposes, and it may be admitted for civil or criminal trial. Because the genetic information is now part of one's health record, insurance companies have access to it. However, GINA prevents the insurance company from denying coverage or increasing premiums based on those genetic tests [7]. Regardless, not all health insurance companies fall within GINA's jurisdiction, and it does not apply to life insurance, disability insurance, or long-term care insurance [7]. Further, the genetic testing of parents is likely to disclose details about a child's genome. Despite the passage of GINA, existing regulations do not shield against all forms of discrimination. Hence a child with a potential harmful genetic alteration may still be mistreated or maltreated in schools, by health care providers, and perhaps by peers [9]. Further, no previous laws or regulations apply to DTC genetic testing or account for privacy implications of DTC sharing.

In 2018, Louisiana was the first state to pass legislation covering DTC testing kits [10]. The law demands any company selling such kits provide consumers with an easy-to-read notice informing individuals of how DNA is used, whether it can be used for other intents, whether it will be shared with third parties, and whether the consumer can request the information be destroyed, and whether the consumer loses ownership

of the information once it is disclosed [10]. But, due to a lack of uniformity between the federal and state rules in handling the significance of genetic privacy, even the existent protections are inoperable. For instance, Maryland and Washington D.C. expressly forbid familial searching by law enforcement, but there is nothing preventing law enforcement in other jurisdictions from running familial searches based on specimens from Maryland arrestees or convicts and thus implicating their relatives. Therefore, we can infer there is a huge gap in protecting people’s genetic privacy and little-to-no risk communication involved to make people aware of the benefits and risks of DNA data sharing.

1.4 Problem Statement and Proposed Contributions

We envision various potential privacy and security risks of at-home DNA testing and sharing. First of all, we posit that DTCs fail to safeguard consumers’ privacy adequately. Second, DTC genetic testing firms fall beyond the coverage of the Health Insurance Portability and Accountability Act (HIPAA), the primary privacy regulation for health data in the US. Without sufficient laws and supervision of the industry, the privacy guidelines of DTCs do not comprehensively declare to customers the risks of sharing their genetic data with DTC genetic testing companies and online platforms. These risks include inaccurate or undesirable health information reports, data breaches, and data misuse. DTC genetic tests that do not deliver health details are not deemed medical gadgets by the U.S. Food and Drug Administration (FDA) and thus are not inspected prior to entering the market. The FDA manages to examine DTC tests used for “high-risk medical purposes,” such as examinations for an individual’s genetic risk of diseases or conditions. The FDA has approved only four high-risk medical DTC tests for marketing until now. This absence of regulation for non-medical DTC tests has contributed to vast variation in DTC testing and analysis accurateness.

Moreover, no substantial rules and restrictions oversee law enforcement’s access to

these genetic data. Law enforcement has leaned on state databases of DNA to determine suspects for decades. Meanwhile, the advancement of DTC genetic testing and sharing in online platforms like GEDmatch has expanded the amount of genetic data that the government and state may access via third parties. Thus, the safeguarding of people's right to privacy is vague. The growth of DTC genetic testing companies and online genetic data sharing platforms has formed a demand for third parties that analyze and interpret genetic data for customers. These genetic interpretation services - such as "matching users to genetic families, marketing customized diet and wellness programs, and delivering health risk assessments" are predominantly unregulated, not overseen and managed by any established organizations or rules or laws. This raises privacy and safety concerns for individuals. At-home genetic testing can provide people with reports about their predispositions to diseases' health risks, which insurance companies could use to set premiums. Life insurance companies, disability insurance, long-term-care insurance, and small companies do not fall under the Genetic Information Nondiscrimination Act (GINA), which prohibits employers and health insurance companies from discriminating against a person based on their genes. Genetic data revelation could lead to the potential for insurance discrimination, prejudice, and biases that could deter people from participating in medical research. As a family inherits or shares genes, part of the health data of a genetic family could also be inferred or derived. This could cause problems for the family in case of insurance premiums or social acceptance. Also, in the case of law enforcement, using these data in investigations could unwillingly drag in innocent people or family members invading their privacy.

End-users need to understand the risks that come with at-home DNA testing and genetic data sharing online. They should be aware of both benefits and risks to make better informed, weighed decisions. Nevertheless, at-home DNA testers and public genealogy database users do not sufficiently comprehend DNA data's sensitivity and

interconnected nature. They barely understand when they choose to share their data, they implicitly disclose DNA data of their present, past, and even future relatives [11]. They are also not adequately knowledgeable that their data are being collected and shared with different stakeholders (law enforcement third-parties, commercial companies).

Currently, DTCs and public genealogy databases do not provide users with adequate awareness of the risks associated with DNA data sharing. As per my literature review, I did not find any study looking into finding a method to communicate the risks of genetic data sharing. Hence, the focus of this thesis is to explore users' knowledge and motivations for taking at-home DNA testing and sharing their data with public genealogy database. We also seek to identify their privacy concerns, awareness of the risks and benefits of taking a DNA test and sharing their DNA data. We aim to recognize aspects that affect their privacy concerns and preferences, and how the current systems accommodate users with their security and privacy needs.

As such, I am investigating multiple research questions:

- How do users comprehend and interpret the inter-connected nature of DNA data? What are the motivations, perceived benefits, and risks of taking an at-home DNA test?
- Is there any perception and attitude difference between those users who have already taken a DNA test vs. those who have not taken it?
- How do users perceive sharing DNA data online? Is there any difference in perception of sharing DNA data with testing companies vs. open databases?
- Are users knowledgeable of the current policies, rules, or laws of their respective testing companies that have shared their DNA and existing laws of the USA? What are their preferred settings, regulations, and laws for DNA data sharing online?

- What is the best or most effective approach to communicating the risks and benefits of DNA data sharing? How does the data sharing intention change after being informed of both risks and benefits?

I investigate these research questions through various user research methods that provide a thorough understanding of users' considerations in genetic data sharing and develop a strategy to communicate to the average population both perils and usefulness of taking DNA tests and sharing. The remaining chapters are organized as follows:

Chapter 2 will discuss literature exploring users' perceptions of at-home DNA testing. Subsequently, I will talk about studies that looked into DNA data sharing risks. Afterward, I will discuss current policies for genetic data privacy or protections and studies that have communicated risks about user-generated health data such as fitness trackers. Next, I will be discussing about studies that have applied different methods of communication for the online sharing of personal data to help users learn the current data practices and influence them to adopt secure behavior.

In chapter 3, I have conducted a qualitative study to explore users' perceptions, motivations, and opinions of at-home DNA testing and sharing DNA on public genealogy databases. We did a semi-structured interview study with 30 experienced participants who had taken a DNA test and 30 non-experienced participants who had not done a test. We asked them about their motivation, interest, perceived benefits, and risks of taking a DNA test. We demonstrated popular GEDmatch tools to understand their interest, motivations, and perceived benefits or risks of sharing their DNA data with public genealogy sites. Then, we discussed multiple scenarios with the participants to inform them about the potential benefits and risks involved when they take the test or share their data with private companies.

We found that users are generally unaware of the potential risks when sharing their DNA data. Users exhibited an extremely concerned attitude when they learned about

potential risks such as the possibility of involuntary surveillance by law enforcement in familial genealogical searches or access by insurance. Furthermore, we gathered their desired privacy setting and policies for DNA data sharing. To be specific, this initial study directed me to investigate “how or what is the effective way to communicate the risks and benefits to aware users to help them weigh the advantages and risks of sharing DNA data?”

In chapter 4, I turn to answer the question of “how or what is the effective way to communicate the risks and benefits to help users weigh the advantages and risks of sharing DNA data?” This study is comprised of three phases. The phase1 is the need-finding phase in which we aspired to understand users’ experiences and perceptions of at-home genetic testing and their perceived benefits and concerns associated with at-home DNA testing and collected users’ requirements to create or design a risk-benefit message. After phase1 analysis, we created 5 forms of risk communication methods. Essentially, Phase2 was an iteration stage where we sought to improve and update our initially abstract designs according to the user’s feedback. Finally in phase3 we tested these created design to determine which designs people prefer the most.

Consequently, in Chapter 5, I applied the developed risk communication method to a sample DNA sharing platform. This study utilized a between-subject design, where the treatment groups were exposed to the risk communication strategy, while the control group was not shown any information before deciding whether to upload their DNA to the site (simulated scenario). Subsequently, we collected their intentions to upload and their reasoning behind the decision.

In summary, the contributions of this research include:

- Provide an in-depth understanding of users’ awareness, perceptions, and concerns of at-home DNA testing and sharing their genomic data.
- Provide a detailed description of users’ perceptions and sharing intentions with different entities, such as with law enforcement research organizations.

- Provide lawmakers, DTCs, public genealogy databases, and researchers with design implications, guidelines, users' desired rules, and regulations control over their data that will contribute to the development of effective security and privacy mechanisms for desired user experiences.
- Provide an effective DNA sharing risk communication method to help users understand the benefits, risks, and implications of sharing DNA data online to facilitate their informed decision-making.

CHAPTER 2: BACKGROUND

In this section, we will begin by introducing privacy calculus theory and then review previous studies that have examined users' perceptions of Direct-to-Consumer Genetic Testing (DTC-GT) and public genealogy databases, focusing on DNA sharing risks. Additionally, we will explore research that has investigated user perceptions regarding user-generated health data. Following that, we will discuss various strategies employed in communicating risk and privacy issues, and provide a categorization of the most commonly used risk communication methods in recent studies.

2.1 Privacy calculus theory

Data privacy encompasses an individual's ability to have control over the sharing and communication of their personal information, including details such as their name, location, health information, genetic information, contact information, and online or real-world behavior. The privacy calculus theory suggests that people assess the perceived risks and benefits associated with privacy before deciding to disclose personal data. In today's world, individuals are increasingly adopting various smart devices, such as fitness trackers, to enhance their fitness levels or opting for at-home DNA tests to make informed lifestyle choices. While the use of these technologies is on the rise among the general population, these devices rely on sensitive user data to provide complete functionality and personalized features.

Among different technological domains, at-home DNA testing and sharing genetic data with private companies present the most significant privacy challenge. This is primarily because DNA data serves as the ultimate identifier of an individual. In my dissertation studies, I investigate whether the actual and perceived control over

collected data influences individuals' willingness to undergo at-home DNA testing and share their genetic data on public genealogy platforms.

Additionally, I measure actual behavior as a result of a risk-benefit trade-off within the framework of privacy calculus theory. Due to the sensitivity of the data that needs to be disclosed for the maximum benefits of DTC DNA testing and sharing DNA online, as well as other data collected for other purposes, such as by law enforcement or pharmaceutical companies, it is inherent that user privacy may be compromised. However, as the preceding statements show, at-home DNA testing could bring tremendous opportunities, especially in the area of healthcare (e.g., path-breaking DNA medicine), an adoptee can find biological family, or making better health choices. Consequently, the user must be aware of and weigh the benefits he or she expects from the online sharing of DNA data and the risks associated with a potential privacy violation.

This balancing process is illustrated by the privacy calculus theory [12]. The theory is based on the belief that people evaluate anticipated benefits and perceived risks to make a rational decision regarding disclosing their personal data. The application of privacy calculus developed from eCommerce [13] and was subsequently extended to other domains have such sensitive disclosure regarding people beyond the user such as group photo uploading on social networking sites [14, 15]. The online sharing of DNA data raises the potential threat to privacy not only to the individual who did it but genetic relatives of the individual. Therefore, this theory helps to explain the decision process and frame the challenge for users with DNA data sharing, where I have shown that in addition to online data sharing, advantages and disadvantages are weighed before users provide personal information when sharing DNA data just like it has been applied for IoT [16].

Numerous studies have examined perceived risks and anticipated benefits of information disclosure as determinants for the privacy calculus. For example; the results of

Kim et al. [17] demonstrated that both perceived benefits and perceived privacy risks have an effect on the willingness to provide personal information when using different IoT services. In their study, Princi and KrÄ mer [18] found that anticipated benefits of household IoT in private environments are a decisive element for the intent of their usage. Due to the great potential of IoT, specifically in healthcare, scientific interest is also advancing concerning eHealth. Although many studies substantiate the appropriateness of privacy calculus theory in the area of IoT in healthcare, social network [14, 15], there is a lack of research on the online sharing of DNA data. Through my studies, I aim to contribute to the existing literature by extending the application of Privacy calculus theory to the domain of online DNA data sharing. By expanding Privacy calculus theory in the context of online DNA data sharing, I seek to enhance our understanding of the decision-making process individuals undergo when faced with sharing their personal genetic information with external entities.

2.2 DNA data

2.2.1 Users' Perceptions of at-home DNA testing

This section summarizes recent studies that have investigated consumers' key concerns about DTC-GT data privacy and motivations behind sharing their genetic data in public genealogy databases. For instance, researchers [19, 20, 21] have investigated the motivations behind individuals opting for DNA testing, finding that users are primarily driven by interests such as obtaining health-related information, learning about their genetic makeup and possible disease risks, and discovering related family members. Curiosity, research purposes, and health improvement also emerged as significant motivations for users [22, 23].

Khan et al. [24] explored the appeal of genetic genealogy to individuals interested in ancestry and locating distant relatives, cautioning against potential exploitation by identity-tracking companies. However, Bollinger et al. [25] assessed user perceptions regarding government oversight and third-party access to genetic information. Their

findings indicated that a majority of participants opposed insurers, employers, and law enforcement agencies having access to DTC-GT genetic information. Similarly, a study in Germany by Schaper et al. [26] revealed that users expressed low trust in genetic testing companies.

Hendrick et al. [27] examined users' perceptions of trust in privacy regulations and maintenance, comparing general practitioners (GPs) to DTC-GT companies. The results showed that users significantly trusted the privacy rules and regulations provided by GPs more than those offered by DTC-GT companies.

In another study by Baig et al. [28], perceptions of users of at-home DNA testing companies were explored. The findings highlighted that users often dismiss privacy concerns regarding their own genetic data and their relatives. Similarly, Saha et al. [11] investigated user concerns and knowledge regarding at-home DNA tests, revealing users' difficulties in understanding the implications of sharing genetic information with business entities. Users generally possessed only a basic level of knowledge about DNA data. However, there is still a lack of understanding regarding the perceptions and attitudes of users who have already taken a DNA test towards different entities accessing their data, as well as any potential differences compared to individuals who have not undergone testing. Thus to ground my work, I extend these prior studies by exploring the perceptions of adopters and non-adopters of at-home DNA testing.

2.2.2 DNA sharing Risks

Genetic data are sensitive as it contains individually identifiable health information such as present, past, or expected health conditions, in addition to ancestry. Genetic data reveals not only sensitive information about the sharer but also related genetic individuals [29]. Individuals' genetic privacy is not covered by current legislation, and the guidance of HIPAA [30]. For instance, commercial genetic services, such as 23andMe and AncestryDNA do not fall under HIPAA guidelines.

Marchini et al. [31] conducted a study on genotype imputation, a technique that

completes genetic information from incomplete data. This technique enables geneticists to assess the evidence for association at genetic markers that are not directly genotyped (complete set of genetic material), thereby enhancing the accuracy of their evaluations. Additionally, Lee et al. [32] discovered that it is possible to infer predisposition for Alzheimer’s disease even when the specific gene associated with the disease is masked. These findings suggest that sensitive health information and an individual’s predisposition to certain conditions can still be identified even if the genome sequence is incomplete and partially masked. A study by Humbert et al. [33] confirmed the feasibility of genetic imputation by utilizing genetic datasets from OpenSNP.org (an Internet platform where genetic information is publicly available). Using Facebook searches, they were able to find relatives of the individuals that self-identified their genetic datasets [33]. Kaiser et al. [34] discussed how in Iceland, geneticists leveraged their large reference panel and genealogical information to infer genetic variants of an additional 200,000 living individuals who never donated their own DNA.

A study performed by Edge et al. [35] showed that the GEDmatch database is vulnerable to genotypes being revealed by artificial datasets and described various methods that point out the several possibilities to reveal users’ raw genetic data. He et al. [36] discussed the possibility of revealing substantial information about one’s possible genetic relatives. Many studies have proved re-identification of anonymized genetic data [35, 37, 38, 31]. Ney et al. [37] found that high-resolution images provided to Gedmatch users comparing the chromosomes of any two users can be potentially misused for reconstructing the target’s genotype. Furthermore, the 1-to-many tool in default Gedmatch reports 3000 of the closest genetic relatives with kit numbers, names, and email addresses. Thus, an adversary can iteratively search for all the kit numbers matching a known kit and get many kit numbers to use in 1-to-1 searches.

As previously discussed, research has highlighted numerous risks associated with

sharing DNA data, many of which users are likely unaware of. Therefore, it is vital to educate individuals about the potential risks and implications of sharing their genetic information.

2.2.3 User generated health data; Self disclosure

A personal health record (PHR) refers to a collection of health-related information, including medical conditions, medications, and self-care behaviors, which individuals document and maintain themselves. However, the disclosure of personal health information can have significant consequences, including privacy invasions, racial inequalities, and potential discrimination by employers and health insurance companies [39].

In a similar vein, when individuals choose to undergo at-home DNA testing and share their DNA data with Direct-to-Consumer (DTC) companies or public genealogy databases, it is a voluntary decision. Unlike mandatory requirements, consumers willingly disclose their genetic data, which often contains sensitive information about their health and ancestry. This act of sharing genetic information can be likened to other forms of self-disclosure, such as sharing data through self-monitoring health devices like Fitbit or participating in online health communities. The potential risks associated with revealing personal health information underscore the importance of considering the implications and understanding the privacy implications before engaging in these voluntary acts of data sharing.

Dahlstrom et al. [40] conducted a study to examine the users' attitudes towards their fitness trackers. The interviewees were presented with the data collection policies; they agreed to use the fitness tracker applications and third parties' potential use cases for such data. Their findings show a change in attitude, as the participants express more concern after being presented with Fitbit's Terms of Service, which indicates a poor knowledge of what terms are being accepted when they start using fitness trackers and an increase in concern when presented with the information. Similarly,

Bellekens et al. [41] found that users showed very inadequate understanding and a lack of knowledge regarding the technologies they chose to use.

Online health communities (OHCs) have developed as an alternative platform for users searching for health information and self-health care management [42] such as PatientsLikeMe, HealthBoard, and MedHelp. Zhang et al. [43] examined the determinants that influence the disclosure of personal information in the OHC context. The result shows that people with good health have fewer privacy concerns regarding health information disclosure behavior on OHCs. Similarly, Deng et al. [44] examine the effect of perceived health risk and health self-efficacy on health information-seeking behavior intention on mobile social media websites. They found that perceived health risk and self-efficacy significantly influence consumers' health information-seeking behavior intention. Specifically, health self-efficacy significantly moderates the relationship between perceived risk and behavior intention. In simpler terms, believing in one's ability to manage their own health has a significant impact on how perceived risks influence the intention to seek health information. This demonstrates that the perceived risk and privacy calculi play essential roles in individuals' decision-making. Thus I seek to understand how to better communicate risk to support user decision making in this new domain of DNA data sharing.

2.3 Users' Decision making and awareness

Human-Computer Interaction (HCI) researchers need to concentrate on portraying hidden privacy and security issues and need to communicate risks to enable people to remain safe while accomplishing their primary task. In the past decade, an increasing number of researches have focused on human factors and usability and privacy issues of various ubiquitous technologies. Researchers have used many strategies to communicate risk and privacy issues. This section represents the categorization of most used risk communication methods in recent studies.

2.3.1 Content of risk communication

Risk communication is the process of sharing information about potential risks and their associated uncertainties to individuals, groups, or communities who may be affected by them. My research work is closely related to the body of literature on risk communication, as I am exploring how to effectively communicate the risks associated with sharing DNA data online to users. In my work, I am incorporating the key elements of risk communication, including the nature and severity of the risks, the source of the risks, the likelihood of the risks, the level of uncertainty surrounding the risks, communication of risk-related decisions, and transparency. This section represents the categorization of contents of risk communication methods in recent studies.

2.3.1.1 Story-telling

Anecdotal stories play a crucial role in shaping mental models and influencing secure behaviors, particularly in relation to security and privacy threats. Recent research has emphasized the effectiveness of storytelling in facilitating informed security decision-making. Rader et al. [45] conducted a survey involving students ($n=301$) to investigate the factors impacting security perceptions and behavior. Their study was subsequently replicated by Pfeffer et al. [46] with a more diverse sample ($n=299$), further validating the positive influence of storytelling on learning and adopting secure behavior. The findings of these studies indicate that stories containing lessons have a direct impact on participants' behavior, while stories conveying serious threats influence their thinking and likelihood of retelling. Moreover, stories that evoke fear or anger affect both thinking and behavior. Interestingly, the context in which stories are shared also matters. Stories narrated in a work environment are more likely to drive behavior change compared to those shared in casual settings like at home or in a coffee shop. These insights shed light on the specific characteristics and contextual

factors that enhance the effectiveness of storytelling in shaping individuals' security behaviors.

Furthermore, the influence of stories extends beyond traditional research settings. Fennell et al. [47] conducted a user study investigating the impact of security stories on individuals' willingness to adopt two-factor authentication. They observed a significant increase in the adoption of this security measure when stories were used. Additionally, Fulton et al. [48] found that entertainment media, such as movies and series, shape users' mental models of security by allowing them to learn from the experiences depicted by actors. Thus, I will explore storytelling as one method to communicate to users regarding DNA data sharing.

2.3.1.2 Data visualization

Researchers have recognized the importance of providing users with easily understandable information regarding the risks associated with new technologies. Previous studies have highlighted the prevalence of usability issues and users' lack of awareness regarding potential consequences [49]. To address this, data visualization has emerged as a powerful tool for presenting information. By utilizing visual elements such as lines, points, areas, and graphics, data visualizations aim to create clear and concise representations that facilitate understanding of complex information [50]. Visual communication of risks has proven to be effective in revealing data patterns, facilitating comparisons, and capturing individuals' attention [51, 52].

For example; Sarma et al. [53] focused on creating informative risk signals for Android applications that are easily understood by both users and developers. By considering factors such as the permissions requested by an application, its category, and the permissions requested by other applications in the same category, the researchers aimed to provide users with an assessment of the risks associated with installing an application compared to its expected benefits. Similarly, Lipford et al. [54] addressed the challenge of managing privacy settings on Facebook by designing

a new interface that utilized visualization techniques. The interface allowed users to control the disclosure of personal information based on different audiences, such as search, network, friend, or self-search. The study revealed that many Facebook users struggled to comprehend the existing privacy settings, and the proposed visualization-based approach facilitated better understanding and control of personal information sharing.

Marett et al. [55] employed data visualization as a method of risk communication to investigate factors influencing users' willingness to share personal information on social networking sites. Their study involved a pre-intervention survey, an intervention phase where statistical data on privacy issues and recommended actions were provided, and a post-intervention survey to assess participants' motivations for sharing personal information online. The findings indicated that female participants were more mindful of the risks associated with public sharing of personal information and expressed a greater inclination towards adopting secure behavior. By leveraging data visualization techniques and lessons from these studies, I aim to enhance users' understanding of the risks involved in sharing DNA data online in my work.

2.3.1.3 Fear Appeal

Risk communication primarily relies on emotion-focused strategies, employing fear appeals and coping judgments to convey potential threats. Previous research has extensively investigated fear appeal approaches, which aim to communicate threats and enhance individuals' belief in their ability to cope with them [56, 57]. Such fear appeals typically consist of two main components: statements presenting an impending threat and suggestions for specific actions to mitigate the threat, including bolstering one's capabilities to follow the recommended course of action [57]. According to Witte [58], fear appeals encompass four essential elements within a communication: susceptibility to the threat, severity of the threat, precise instructions to counter the threat, and efficacy of the response. The first two elements aim to enhance under-

standing of the threat, while the latter two focus on strengthening perceptions of the response's effectiveness. Various theories, including the widely developed protection motivation theory (PMT) [59, 56], explain fear appeals by highlighting the role of threat and coping appraisals in motivating protective actions [57].

Several studies have examined the effectiveness of fear appeals in different contexts related to security and privacy risks. Vance et al. [60] conducted a field experiment involving 354 participants to observe the impact of interactive fear appeal text during password creation. They found that the group exposed to the interactive fear appeal text exhibited the highest increase in password strength compared to other groups. Boss et al. [61] investigated the influence of fear appeals on perceived fears, intentions, and performance related to backups and anti-malware software. Their studies showed that fear appeals had a significant impact on participants' perceived fears, intentions, and actual performance in terms of backups and software usage. Jansen et al. [62] explored the effects of strong and weak fear appeals on users' behavioral intentions and attitudes towards protective online information-sharing behaviors. Their findings indicated that using fear appeals with varying levels of intensity raised participants' security awareness and promoted security behavior. Albayram et al. [63] and Elham et al. [64] both investigated the effectiveness of fear appeal videos in changing users' security behavior, specifically focusing on enabling screen locks on smartphones. Both studies demonstrated the efficacy of fear appeal videos in positively influencing users' behavior and risk perception.

Although the research studies I have cited in this field primarily focus on other areas such as password protection and antimalware, the underlying theme of sharing sensitive information remains relevant to my research. By utilizing these elements of risk communication, my study aims to provide users with a clear understanding of the risks associated with sharing their DNA data online, ultimately leading to informed decision-making regarding the sharing of such sensitive information.

2.3.2 Mechanisms to communicate personal / sensitive data disclosure to users

Personal data refers to any information that is linked to an identified or identifiable individual, such as health data, biometric data, and genetic data [65]. Sensitive data, on the other hand, encompasses information that should not be accessed by unauthorized parties. It may include personally identifiable information (PII), such as Social Security numbers, financial details, or login credentials [65]. In order to communicate the collection of personal data and potential security and privacy risks to users, researchers have employed various methods, including alerts, nudges, awareness mechanisms to deliver communications. The focus of my research is to explore different risk communication mechanisms specifically related to the disclosure of personal or sensitive data, particularly DNA data online. While existing studies have examined risk communication in other domains, my research aims to contribute to the literature by investigating various themes associated with sharing of DNA data and which methods may be most useful in this domain.

2.3.2.1 Alerts

Communicating the risks associated with sharing DNA data online is crucial for promoting users' understanding and adoption of secure behaviors. Previous studies have shown that users often have a relaxed attitude towards data privacy if they have not experienced any negative consequences in the past [66]. Additionally, users tend to neglect security warnings and dialogs, despite their importance in alerting users to potential risks [67]. To address this issue, researchers have explored various strategies to effectively communicate risks and promote user awareness.

One mechanism is "Alerts" to provide users with real and informative information about a decision they need to make. One such decision is granting permission for access to information. Egelman et al. [68] suggested design modifications to the Facebook connect dialog, such as displaying the real information about the public profile

permission. While these modifications were noticed by users, the low expectations for privacy hindered the effectiveness of the added information. Passive eye-tracking techniques were used to examine the readability of dialog designs, providing insights into users' attention and gaze fixations [68].

Another approach is to provide feedback and decision-making support to users in disclosing their personal information. Patil et al. [69] developed an app called "Locasa" that facilitated decision-making in location sharing. They found that a significant number of participants exhibited inconsistencies between their disclosure settings and contextual choices, with approximately 65% of responses resulting in mismatches. Immediate feedback on location disclosure was found to provoke feelings of oversharing, but when participants were in control of making the decision, oversharing was significantly reduced [69].

Furthermore, alternative interaction design solutions have been proposed to enhance the communication of data flow and disclosure while granting permissions. Lindegren et al. [70] evaluated different design options, including swiping, Drag and Drop (DAD), and checkboxes, for selecting personal information on mobile apps' permission dialogues. Their study demonstrated that while checkboxes were faster, swiping and Drag and Drop engaged users more effectively [70].

In conclusion, effective risk communication regarding DNA data sharing involves utilizing personalized examples, and informative design modifications to draw users' attention to data privacy and promote awareness of secure practices. These strategies aim to bridge the gap between users' perception of risks and the actual implications of their behaviors, ultimately leading to improved data protection and user decision-making.

2.3.2.2 Nudges

Numerous studies have explored use of nudges to effectively communicate the risks of personal data sharing. Nudges refer to predictable strategies that steer individuals

towards more desirable options without limiting their freedom of choice [71]. In the realm of online security and privacy, researchers have employed nudges in diverse scenarios, such as mobile app installation, password creation, and social network post sharing [72, 73, 74, 75].

Behavioral economics and human-computer interaction literature suggest that social nudges, which inform users of public opinions, may be particularly effective in deterring risky behavior [76, 73]. To prevent sensitive data disclosure, previous studies have drawn insights from behavioral decision research and soft paternalism, designing mechanisms that nudge users to consider the content and context of their data disclosures before sharing [72, 77]. For instance, Choe et al. [72] designed a visual privacy rating system for mobile apps to nudge users away from privacy-invasive applications. They found that positively framed visual representations influenced participants' perception of an app's trustworthiness. Similarly, Masaki et al. [77] utilized negative framing nudges to discourage risky sharing behavior on social networks, showing that participants were more likely to avoid potentially risky choices when presented with negative frame nudges. On the other hand, Besmer et al. [78] found that social navigation cues had minimal effects on users' privacy settings on Facebook, suggesting that only a small subset of users who customize their settings may be influenced by strong negative social cues. Wang et al. [74] developed privacy nudges for Facebook sharing behavior, including picture nudges, timer nudges, and sentiment nudges, which were successful in encouraging users to reconsider their posts and be more cautious.

In addition to privacy concerns, researchers have also focused on promoting secure behavior, such as adopting strong passwords and login strategies. Ur et al. [73] developed a data-driven password meter that provided accurate strength measurements and detailed feedback to users, motivating them to generate more secure yet memorable passwords. Frik et al. [79] investigated nudging users towards adopt-

ing two-factor authentication (2FA) for improved computer security. Their study demonstrated that allowing users to delay or schedule security actions significantly increased their willingness to engage in the proposed security behavior. Furthermore, Almuhiemedi et al. [80] conducted a field study on mobile privacy managers and privacy nudges, showing that participants benefited from access to permission managers that provided nudges about the frequency of app access to sensitive data. As a result, participants reassessed their permissions and further restricted some of them.

In summary, to develop effective DNA risk communication, it is crucial to explore various risk communication mechanisms, including nudges, that encourage privacy and safety-conscious behavior. Social nudges, framed visual representations, and feedback mechanisms have been shown to be effective in promoting privacy-aware behavior. Likewise, data-driven password meters, nudges for 2FA adoption, and privacy managers with nudges have successfully enhanced secure behavior among users. By leveraging insights from behavioral economics and human-computer interaction, future efforts can be directed towards developing more effective approaches for DNA risk communication.

2.3.2.3 Awareness

Behaviors related to security and risk are often challenging to control, and individuals may lack adequate training or knowledge about the associated risks. However, ensuring the security of information and data is crucial. This lack of awareness about appropriate security behaviors, such as identifying phishing emails, avoiding malicious websites, or creating strong passwords, underscores the need for awareness campaigns across different domains. These campaigns play a significant role in influencing users' security behaviors and can be delivered through various methods, including message distributions (e.g., booklets, emails, text messages, and posters), live presentations (meetings, face-to-face training courses, and conferences), or media platforms (videos, games, and online websites) [81].

To address the risks of data breaches and social engineering attacks, recommendations include implementing educational training programs for employees to foster an information security culture and raise awareness about attackers' techniques [82]. Emphasizing the importance of information security awareness throughout the organization is also crucial to prevent potential leaks of classified data [83]. Innovative learning platforms, such as gamification, have been utilized to enhance employees' awareness of information security, helping them understand the principles exploited by attackers and develop resistance strategies [84].

Studies have examined the effectiveness of different security awareness delivery methods. For example, video-based delivery has been found to be the most preferred and effective method for raising awareness about phishing attacks [81]. Security education, training, and awareness programs have also proven effective in increasing employees' awareness regarding information security, security policies, and potential threats [85].

In the context of password security, hands-on exercises and traditional lecture approaches have been compared, with mixed results. While a study suggested that hands-on exercises might not yield significantly different outcomes compared to lecture-based approaches, factors such as sample size and participants' prior experience can influence the results [86].

In educational settings, information security awareness programs have been implemented to educate students, faculty, and staff. These programs involve a combination of in-person and web-based training, as well as various communication channels to disseminate information about risks and security practices. The programs have shown positive outcomes, leading to increased reporting of virus infections and phishing emails [87].

Understanding employees' compliance with information security policies is another important aspect. Various theoretical behavior constructs have been used to assess

employees' behavior and intention to comply. Social influence from administration and normative beliefs have been found to influence employees' intention to comply with information security policies, while rewards did not significantly impact actual compliance behavior [88].

To promote compliance with information security policies, a multi-step approach involving measuring employees' security awareness, conducting e-learning-based awareness campaigns, and reevaluating awareness levels has been proposed. This approach has demonstrated a positive influence on employees' behavioral intentions to comply with information security policies [89].

DNA sharing risks and benefits are complex, and may need to be considered by users long before sharing DNA data. Thus communication may resemble other kinds of awareness campaigns. For example, video may be a good medium, along with considering a multi-step approach,

2.4 Summary

This chapter provides an overview of the current state-of-the-art in my research area, focusing on several key aspects. Firstly, we examined the growing trend of at-home DNA testing and its implications. This includes the motivations behind individuals opting for these tests and their perceptions regarding the sharing of DNA data. Understanding these factors is crucial in comprehending users' attitudes and behaviors related to online DNA data sharing.

Next, we delved into the risks associated with genetic data sharing. By exploring potential privacy and security concerns, we highlighted the importance of effective risk communication in this domain. Recognizing the sensitive nature of genetic data, it is vital to educate users about the potential risks involved in sharing such information online.

In the latter part of this chapter, we explored various communication mechanisms and contents employed in different fields and technologies to inform and educate

users, as well as improve their awareness regarding data practices. By examining these communication strategies, contents we gain valuable insights that can guide the design of communication methods and messages specific to online DNA data sharing. The goal is to raise awareness among individuals about the potential risks and benefits associated with sharing their DNA data online.

By synthesizing the findings from these areas of research, we are better equipped to develop effective approaches for communicating the risks and benefits of online DNA data sharing. This understanding will enable us to create informative and persuasive communication methods and messages that address users' concerns, promote informed decision-making, and ultimately enhance privacy protection in the context of genetic data sharing.

CHAPTER 3: Exploring Users' perceptions of at-home DNA testing and sharing of DNA data online [1]

3.1 Introduction

Considering the rapidly increasing popularity of at-home DNA testing and public genealogy databases, it is essential to understand users' perceptions of the benefits and risks of commercial genetic testing and sharing their genetic data in the open genealogy databases. Therefore, we conducted a semi-structured user study to investigate the following research questions.

- RQ1: Do users comprehend and interpret the interconnected nature of DNA data?
- RQ2: What are the motivations, perceived benefits, and risks of taking an at-home DNA test?
- RQ3: Is there any perception and attitude difference between those users who have already taken a DNA test vs. those who have not taken it?
- RQ4: How do they perceive sharing DNA data online? Is there any difference in perception of sharing DNA data with testing companies vs. open databases?
- RQ5: Are users knowledgeable of the current policies, rules, or laws of their respective testing companies that have shared their DNA and existing laws of the USA?
- RQ6: What are their preferred settings, regulations, and laws for DNA data sharing online?

To investigate our research questions, we conducted and analyzed 60 interviews. We demonstrated popular tools of GEDmatch and discussed about their perception of the platform. We also discussed about sharing their DNA data with different entities such as law enforcement, research organizations, and insurance companies. We also assessed their knowledge of the current policies or laws for genetic privacy. We found that users' have insufficient understanding of the nature of DNA data and the risks of sharing DNA data in open databases. Users are not informed of the privacy policies present and assume DNA data is not sensitive, leading to less concern for privacy. We also noted changes of point of view and increased privacy concerns subsequently nudged through the scenarios such as potential access by insurance. We render a discussion of the implications of our findings.

3.2 Methodology

This chapter focuses on studying users' perception of DNA data sharing; we designed the interview study. To compare views, we interviewed people with and without experience of at-home DNA testing. All interviews were done through online video conferences using Zoom. We divided the participants into two groups;

- **Experienced group (EG):** already have done at-home DNA testing with DTCs.
- **Non-experienced group (NE):** have not done at-home DNA testing.

3.2.1 Recruitment & Demographics

We recruited 60 participants, 30 for each group, based on the initial screening survey attached to the email. The screening survey asked whether they have done DNA testing or not. Participants were initially recruited by sending out emails via our university mailing lists. We complemented recruiting with snowball sampling to gain a more diverse representation, where initial participants suggested new interviewees.

We did not mention privacy perceptions in our email, instead stating: “ The purpose of this study is to explore the understanding, awareness, impression of DNA testing and sharing for ancestry and family finding user " to more closely study participant behavior. Interviews occurred between May 2021 and July 2021. All participants were compensated with a \$10 gift card. The university’s Institutional Review Board approved the study.

Among 30 participants of the experienced group, 2 were males, and 28 were females. Participants’ age ranged from 18 to 65 years. Among 30 participants of the non-experienced group, 8 were males, and 22 were females. Participants’ age ranged from 18 to 57 years. Participants had different fields of education or occupations such as religious study, IT analyst (cf. Appendix B).

3.2.2 Procedure & Analysis

In the recruitment email, we included a screening survey (cf. Appendix A.1) asking about participants’ experience on at-home DNA testing and the purpose of the testing. The survey helped us to divide the participants into experienced group and non-experienced group. Having both experienced and non-experienced participants enabled us to investigate if there is any difference or effect of experience on the users’ views about DNA data sharing. All interview questions can be found in appendix section (cf. Appendix A.2). First, we started asking the experienced group about the motivation behind their test and their feelings after the test. On the other hand, we asked the non-experienced participants whether they know at-home DNA testing or not. Subsequently, we gave a description of at-home DNA testing and asked them whether they would like to take the test if they get it as a gift, followed by the reason behind their decision. Then we introduced the GEDmatch site to both the groups and showed them a video explaining few popular tools such as one-to-many autosomal DNA comparison, one-to-one autosomal DNA comparison GEDmatch forum, and GEDCOM. After showing the video, we asked the participants if they would be

interested in sharing their result on these platforms. If a participant showed interest in sharing their DNA data, we asked why they want to share; if they did not show interest, we asked why they do not want to share. Then we discussed the opt-in and opt-out policy of the GEDmatch. After the opt-in and opt-out policy discussion, we asked them if they would choose opt-in and opt-out, followed by the rationale behind their decision. Subsequently, we discussed the “Golden State Killer Case”, their opinion on Law enforcement using the database, their feelings if family shares their own DNA data as their family shares part of their DNA data.

Next, we asked them about their interest in sharing the DNA data for research, medicine purpose, and the motivation behind their choice followed by their feelings family shares their own DNA data as their family shares part of their DNA data. Besides, we also talked about their views on sharing DNA data informing their hereditary diseases and their interest to share or not. After this, we asked about their opinion or concerns about insurance getting access to DNA data. Finally, we discussed their future expectation of DNA data sharing and design suggestions for DNA data sharing platforms

All interviews were transcribed for analysis and adopted an inductive approach. Both coders used the QDA miner software [90]. The data was coded independently by two researchers. We then reviewed, refined, and updated the two sets of coded data to settle disagreements. Thus, we did not administer Cohen’s Kappa (inter-rater agreement). While discussing the results, we enumerate the participants from E1 to E30 for the Experienced group, NE1 to NE30 for the Non-experienced group.

3.3 Results

3.3.1 Pre-Introduction of DNA testing: Non-Experienced Group

Before introducing at-home DNA testing, NE members were asked about their familiarity with commercial DNA testing and their expected gains and concerns.

Foreknowledge about DNA testing and procedure: We asked the participants

- "Do they know about at-home DNA testing?" All the participants knew about at-home DNA testing. Regarding the procedure of the testing, 29 out of 30 participants have different levels of knowledge. 17/30 participants talked about collecting saliva or swab and sending it to the labs. Ten participants talked about DNA sequencing, and two participants (biology background) talked about the procedure in detail. 20/30 participants watched commercials or ADs, some of their family members have done the testings. We can conclude that the participants are familiar with at-home DNA testing.

Expected benefits & Concerns: We asked participants about their expected benefits and concerns of taking at-home DNA testing. All our participants found the at-home DNA test easy to perform and found it an easy and fast way to find out about their ethnicity, heritage, and ancestral line. Most (25) of the participants talked about the advantages of knowing the health predispositions. They said knowing their health predispositions can help them to be prepared for their future. For instance, (NE29) said: *"I've seen in TV; they show your heritage, which is interesting. If you are carrying any gene like cancer gene, they tell you that maybe regarding sleep paralysis, like just little things which are super important."* A few participants (5) said knowing health dispositions would increase anxiety if the issue has no remedy or treatment. Two of these four participants questioned data ownership and privacy. The other two talked about the accuracy of these tests, and one said these tests could reveal some of the unwanted things (truth). In this context, (NE4) told: *"The concerns were ownership aspects, technically someone else owned my DNA is weird. Also, chance that criminal investigations, they can track family members through your DNA, which is concerning."*

3.3.2 Post-Introduction of DNA testing: Non-Experienced Group

After introducing at-home DNA testing, members of the NE group were asked about their interests, motives, and concerns about a test.

Interest and motivations: In the NE group, 21 participants showed interest in taking a DNA test with DTC-GT after introducing DNA testing. Three out of these 21 mentioned that they would be interested in taking the test if they get it free. Another two members talked about the hassle of mailing the samples though they wanted to do the testing. 15 participants of these interested participants would like to take the test to know about their health and predisposition to any health conditions. Five participants wanted to take the test out of curiosity or to explore ancestry. One participant mentioned that they would be interested in taking the test here in the USA but would have never shown interest in taking it in their home country (Asian country). They mentioned that governmental oversight, discrimination, and tracking is the primary cause behind this.

No Interest and concerns: Nine participants did not show interest in taking the test even if they received it as a gift or free of cost and 28 out of 30 participants had some concerns. The reasons were mentioned in the table 3.1

Table 3.1: Post-Introduction of DNA testing: Concerns discussed by NE participants

Concerns	Frequency/ (N=30 (NE))
Selling data to third parties	29
Unknown misuse of the data	27
Frequent policy changes	25
Lack of trust in companies	25
Data breaches	23
Ownership of the data	14
Racism	9
Government Surveillance	6
Tracking for criminal activities	3
Hustle of mailing the sample	2

A few (6) participants said they would not like to take the test because they feel they can be monitored more by the government. (NE23) said; *“If a company has this data about you, then they can make certain guesses about you, You may be discriminated*

against because of your color, ethnicity, gender, heritage. Interestingly, you might not even know you come from that area, but your DNA kit says you come from that area."

3.3.3 Users' experience: Experienced Group

In the screening survey and beginning of the interview, we asked EG participants motives behind taking a DNA test and benefits and concerns after the test.

Background: We asked the participants about the reason behind taking at-home DNA testing. Participants did the testing to find family or explore personal identity, ancestry research or finding biological family. Results summarized in table 3.2.

Table 3.2: Motivation behind taking the test

Motivations	Frequency (N=30 (EG))
Explore personal identity / Finding biological family	22
Ancestry Research	19
Satisfy curiosity	6
Participate in genetic research	2

19/30 users did not share their testing results anywhere else. Eight participants shared the ethnicity analysis outcomes in social networking sites. Three participants have shared with GEDmatch for finding connections.

Concerns and benefits after taking test: 26 participants did not express any concerns after doing the test. Only four participants talked about selling the data to third parties, curiosity about ownership of data, and wondering what happens to sample after a user gets result though they never tried to find out about it. All participants perceived the test beneficial. Almost all (28) participants responded that they gained insights into their ancestry, new connections & confirming existing family relations. Six members informed that they obtained knowledge of family health-related

Table 3.3: Perceived benefits and concerns about sharing DNA data on GEDmatch

Willing to share in GEDmatch	EG (N=30)	NE (N=30)
Yes	25	9
No	5	21
Perceived Benefits	EG	NE
Family finding / Finding connections	23	16
Exploring ancestry	19	12
Building family tree	16	8
Allows raw DNA data from any source	12	6
Participate in genetic research	4	1
Perceived Concerns	EG	NE
Presence of email ID, names, locations of birth, and deaths in the results	27	28
Targeted tracking or stalking	13	18
Identity theft	7	13
Could reveal health conditions	0	2
Doubt the reliability and accuracy of GEDmatch tools	1	2
Reidentification	1	0

problems and issues. (E15) commented: *"I m adopted and I had very little information on my biological family, family health history. so I did for health information, and the DNA relatives to find connections."* Overall, insight into the family was the primary driving factor.

3.3.4 Post GEDmatch Demo

This section discusses participants' views on GEDmatch and their perspectives after discussing GEDmatch tools.

Benefits and Concerns: We asked the participants if they took the DNA test - "whether they would like to share in the GEDmatch after the demo." All participants of the NE group said the GEDmatch is interesting; however, only nine participants of the NE group would like to share their DNA data in the GEDmatch contrast to most (25) of the EG members. Most (58) participants of both groups liked the matching tools of the GEDmatch. They liked the idea of family finding or finding connections, exploring their ancestry, building family trees, and the ability of this platform to let users upload raw DNA from various sites.

Though most participants admired the tools, many participants of both groups expressed concerns. We saw they were worried about the presence of email ID, names,

locations of birth, and deaths in the results of GEDmatch tools. Participants specifically pointed out that the detailed results can be obtained by just typing someones' first name and last name through the GEDcom tool. They considered these personal details and have the potential to be misused in many ways like targeted tracking or stalking. 5/60 participants asked the interviewer- if the detailed chromosome matching can be interpreted, which may disclose some sensitive details, for example, health conditions about the matches. Five participants also spoke about identity theft. One participant indicated that the data could never be disconnected from the donor. Three participants doubted the reliability and accuracy of these tools. Table 3.3 represent detail overview of participants responses.

Incomprehension is prominent in participants of both groups. These participants mentioned they are happy to share their DNA sequence but not email IDs or names. They believed name and email addresses are more sensitive data than DNA sequences. **Low privacy expectancy** is another uttered statement in participants. A few participants said that almost all the information about a target could be obtained on the Internet. So, they will not be bothered about sharing their DNA data in a GEDmatch. For example, (E10) said: *"Someone could use these details to steal your identity, but as far as like the DNA information, it will just connect you with other people, I would be okay with that part of it. I would not put much personal information on there such as location, I'd create a separate email address."* Some African American participants mentioned they would be more comfortable sharing their details in the African ancestry company as they might get more connections and would not be racially prejudiced. Overall, both the groups expressed interests and concerns about sharing their information in the GEDmatch. However, the EG group showed a relaxed attitude concerning sharing their DNA data in the GEDmatch.

Expected use of DNA data: After asking about expected benefits and concerns, we asked the participants - "what do you think will happen to your data if you upload

in GEDmatch?" to know their expected data usage by GEDmatch. Overall, there are apparent tensions in participants of the NE group. While almost all (28) NE participants mentioned the probabilities of using the DNA data in unapproved or unknown ways, most EG (26) members articulated the good uses. Participants of both groups discussed hacking or data breaches, selling data to third parties, mining the data for targeted commercials or ads. Members of both groups mentioned that about governmental access and oversight of these databases. Details are in table 3.4 below

Table 3.4: Expected use of DNA data

Expected use of DNA data	EG	NE
Probabilities of using the DNA data in unapproved or unknown ways	4	28
Hacking/data breaches	5	21
Governmental oversight	6	12
Selling data to third parties	5	10
Targeted commercials	23	5

A few (6) NE participants and the majority (16) of Experienced participants said these databases must have been used for criminology, genealogical research, and health research. A few participants of both groups said they do not know what would happen with it. A couple of NE participants said DNA data could be manipulated against some races, framing in a police case. Referring to this (*NE10*) said: *"They might not be selling the information now, but suppose in five years they were not profitable, they are going to sell."* Overall, the experienced group talked about good uses more frequently, while the NE group was more concerned.

Expected data access and data handling: When we asked the participants - "who do you think to have access to their data on GEDmatch if they upload?" most (29) NE members said anyone with the internet could have access to data on GED-

match. On the other hand, the majority (17) of experienced participants believed only website users could have access. The next question was - "Rate your concern about the data handling in GEDmatch." The majority (34/60) of the participants are little to moderately concerned. We can deduce that a majority of the participants in both groups were **uncomprehending** of the sensitivity of DNA data. They do not adequately comprehend the risks of sharing genetic data. Their concerns regarding their email ID, name, location are way more than their DNA data. (NE27) commented: *"I am absolutely fine with sharing my DNA data if they would not give access to my sensitive data like my name, email ID, location, kind of the things that can identify me."* Few users of the NE group and most of the experienced group conceptualized that only users who uploaded their DNA can only gather data about anyone. Most participants showed little understanding of DNA data and had an attitude. Some NE members had a misunderstanding that, by this DNA test, they would share only part of DNA, not the entire DNA sequence. This certainly demonstrates a huge gap in people in general about the characteristics of DNA data.

Opt-in or Opt-out: We talked about the Opt-in and Opt-out policy of the GEDmatch and asked participants - if they upload their DNA data in the GEDmatch, would they Opt-in or Opt-out? Twenty NE participants fifteen EG members declared the wish to opt-in. The reverberated reasons are "Nothing to Hide," "Have not committed any crime," "Law-abiding," "Helpful for law enforcement," and "Helpful for law-enforcement to identify me if something happens to me." The common reiterated comment was, "I have nothing to hide." Fourteen EG and Seven NE members stated they would opt out. The reasons were mentioned as - "Minority and vulnerable race", "Do not want to deal with law enforcement," "Can put my relatives and me in trouble," "Misinterpreted or false allegation," "Framing or abuse," "Mistrust law enforcement," "Want to keep my data in my control." A few said that it does not matter what they prefer; if law enforcement wants to get the data, they will even if

opt-out.

3.3.5 DTC-GT VS GEDmatch

Participants of both the groups exhibited higher trust in DTC-GT than GEDmatch. We asked participants who either have already taken the test or are interested in taking one but do not want to share their data with GEDmatch. There was a considerable gap of trust between GEDmatch and DTC-GT. Most participants said they trust and feel their data is safe and protected in DTC-GTs, but not in GEDmatch. They mentioned that their familiarity with DTC-GT through Ads or by other channels is way higher than GEDmatch. They assume DTC-GTs are paid companies, so they would keep their data protected. Also, these companies are tied with terms and conditions that will enforce laws, rules, and regulations and prevent sharing any personal information of the consumers, whereas GEDmatch is an open database. (E8) said: *"DTCs are tied to agreements and laws. They can not share my data without my permission. I don't want my history to be pulled up into the system without my permission the way you can do in GEDmatch."* We questioned if they would be interested in sharing their data after knowing that DNA data can reveal hereditary diseases or the probability of future health conditions. Many participants showed reservedness to share their DNA data on any online platform though they would still be interested in taking the DNA test with a DTC-GT. The rationale behind this was that they have a higher degree of trust with DTC-GT and mentioned that DTC-GTs are verified companies. Whereas, they perceived free DNA data sharing online platforms like GEDmatch can be breached or hacked or can be accessible to insurance with ease. This manifests that users have higher confidence in DTC-GT than GEDmatch.

3.3.6 Scenarios: DNA data Sharing

We investigate users' feelings about sharing their data with different entities in public databases like GEDmatch. To obtain deeper understandings, we asked the

Table 3.5: Concerns regarding law enforcement access to DNA data

Concerns	EG	NE
Creating a fake account or putting false information is Perjury, immoral, unethical	28	26
Obtaining users' data without their consent is overstepping boundaries and a breach of users' privacy	27	27
Could be dragged into investigation	23	27
Unwillingly involves innocent relatives	24	26
Opt-in and opt-out are nearly meaningless	23	23
Could be used to manipulate or frame	23	23
Could lead to racial disparity and targeted accusation of minority/ religion	16	23
Wrongfully convicts or falsely alleges	16	22
Could be a slippery slope	13	11
Violation of Health data privacy as DNA can substantially reveal health information	10	9

same set of questions after discussing four scenarios. The scenarios are based either on an actual incident, fact, or future potential use or misuse of DNA data. Our questions are based on the fact that when someone shares their DNA, it affects the person who has given consent and affects those who have not given consent and vice-versa.

Subpoenas and sharing: Law Enforcement: We asked participants their **opinions on law enforcement** using at-home DNA databases for solving cold cases after discussing "*Golden State Killer Case*" where police tracked down the criminal by using GEDmatch database. We received mixed responses from most of the participants. While participants expressed, these databases are a great tool to capture notorious criminals and serve justice, at the same time, there are apparent tensions and concerns. The concerns are presented in table 3.5

The next question we asked was about their **opinion on implied DNA data**

sharing. That means when a genetic relative shares DNA, it will share part of their DNA too. We found that almost all participants did not realize that. Most were concerned. Few of them changed their mind about taking a DNA test. Matter of consent, victimize of others' action, helplessness theme emerged, elaborately presented in the "Scenario effect" section. Following, we explore how they feel about being **involuntary surveilled** by law enforcement. The majority of the participants would be uncomfortable to very uncomfortable. They would feel upset, harassed, privacy violations if dragged to criminal investigations. Some stated they could not do anything about this as law enforcement is in a higher power. A few said they have nothing to hide and want criminals to get off the street; thus, they are comfortable. When the participants were asked if they would like to share their DNA data voluntarily with law enforcement, the majority of the participants felt comfortable repeating the phrase "Nothing to hide." Some of the participants were unwilling to share mentioning reasons like "framing," "Wrongful conviction or mishandling of data," "Racial discrimination," "Could drag family and me into unwanted matters," "Lack of trust," "privilege to do anything if they own the data."

Moving forward, we asked about their **opinion on other family members sharing their DNA data** with law enforcement or where law enforcement can have access. Most of the participants were neutral about it, mentioning they do not control others' decisions. Few of the participants were in both extreme ends. They felt very comfortable because they have "Nothing to hide," and the uniqueness or specifics of their DNA is not being shared. On the other hand, few of them showed concerns as they believe they can be tracked or surveilled by law enforcement. Finally, we asked participants - "what would families think if they share their data?" The majority of the participants said their family would be neutral as they are not knowledgeable of the specifics of DNA data sharing; they are law-abiding people; hence they have nothing to cover. Few participants said their families would be annoyed if they

could be traced back. Overall the perceptions of both groups were similar about law enforcement access though the Experienced group was more comfortable .

Perceptions about sharing data in health research: Seventeen members of NE group showed willingness to share their DNA data anonymously for research and the medical field, in contrast to almost all (27) EG participants. Ten EG participants have already approved DTC-GTs their data to be used in research. The prime urge was to advance medicine and the health field. They stated higher trust in research and medical organizations as they believe those are tied to government regulations and policies. Eight NE participants expressed the desire to know the credibility of the research institute and their research. On the other hand, twelve NE participants showed reservedness or denied sharing their DNA data for research and medical institutions. Their concerns revolved around detrimental researches around racial biases, non-transparency of data practice, and storage. A few (5) raised concerns about the probability of insurance companies and employers looking at the data. (NE24) said: *"I am familiar with the case of Henrietta Lacks, where basically, I am paying to have something done. And then with the potential of some company making a ton of money off of my biological product."* Then, we asked participants' opinions on families sharing their data for research or medicine purposes. Like above, most of the participants were comfortable; some felt neutral since they can not constrain others' choices and a few were worried for the same reasons discussed before. Hereafter, we explored participants' perception of - what their family will think if they share their DNA data. The majority of them perceived their family would appreciate it as they are helping advance science. Six NE participants told their family would be very uncomfortable because of the history of racial prejudice and might fear insurance access. On the whole, the experienced group was more comfortable.

Hereditary diseases: We explicitly asked the participants- "would they like to share their DNA if it reveals hereditary diseases and probability of some health is-

sues." 17 NE participants showed interest in getting a DNA test (if possible with a medical institution, not with DTCs) to know about their predispositions and hereditary health issues. They perceived that by learning this, they could be prepared and would take precautionary steps. In contrast, seven NE participants mentioned it would make them very anxious and can lead to emotional turmoil. Nevertheless, 21 NE participants did not show interest in sharing their health data or DNA in open databases like GEDmatch. Their concerns were insurance access, employer access, or disclose family's sensitive health data out. We also noted an interesting trend. Specific race participants (we never asked them about the race, they self-revealed) explicitly stated they need to undergo DNA testing to know their health predispositions before marrying or having children. Participants reacted and perceived similarly about family sharing their DNA data revealing health conditions and their family's reaction to them sharing their data. In this context (NE21) told: *"I don't want my aunt to suddenly call me up and say, cancer is in our family, you have a high likelihood you're going to have cancer, I don't want to know that. If I want to make that decision to try and find that out. I don't necessarily want someone just randomly to tell me that. I want to prepare myself for it."*

DNA data access by Insurance company: Almost all participants of both groups (29 NE + 29 EG) strongly rejected the idea of sharing their data online if insurance has access to it. The foremost causes noted are; genetic marks about pre-existing conditions or predispositions or family history that can make many people uninsurable, premiums could be raised, and probabilities of denial to claims. This can give insurance companies more power, ultimately leading to political lobbying or policy-making in their favor rather than on the common people's side. (E25) Said: *"I feel like by them having even more access to family history, predispositions is almost finding a way not to insure the person."* Only a couple of participants felt neutral or slightly comfortable. They perceived themselves as healthy and that having access

to their DNA would not create any problems. Twenty EG participants and 15 NE participants said if insurance gets access to or uses their DNA data, they would take legal action against the insurance firm and company. Most participants of both groups responded similarly (very uncomfortable) when we asked about their family members sharing their data. Eight NE members would be neutral as they feel they do not have power over others' decisions. Also, 28 NE participants of them told their families would be uncomfortable only if they knew that the family's DNA is connected to reaching others. 18/30 EG participants spontaneously asked the interviewer about the policy and rules to know if insurance has access to or could access their genetic data. Overall, this scenario led to changes in decisions in participants and surely put stress on thinking and digging into the privacy policies of genealogical databases.

3.3.7 Scenarios Effects

This section articulates the impacts of scenarios on participants' points of view and attitude.

Helplessness & Resignation: Surprisingly, 12 participants from the NE group showed a resignation attitude and expressed the desire to get the test because many family members have now taken the test, which ultimately shared or publicized part of their data. They displayed helplessness or low efficacy or control over their data. In this context, (NE4) said: *"All my family has done it. It's almost like the cat has been left out of the bag. So it doesn't matter anymore."* The helpless feelings were prevailing among the NE group. All participants from the NE group felt they do not have power over others' decisions. (NE3 commented: *"I can advise [to take the test or not] if I have been asked, but it's ultimately their decision."*

Fear & attitude change: After discussing all the use cases or possible scenarios, 28 participants expressed they would be more cautious about privacy policies if they plan to do a DNA test with DTC-GT or share their data. Interestingly, some NE members who showed high interest in taking the test initially now changed their

decision. All most all (29) of the participants in both groups expressed worried or changed their decision after hypothetical insurance company scenario. (NE21) spoke: *"If I had done the test before the interview, I would not have read the policy, but now if needed to take a test, my first step is extensive research on companies."* Twenty-eight NE participants responded similarly (very uncomfortable) when we asked about their family sharing their data. All participants said business or profit company should not have access to DNA data.

Regret & Realization: We asked about the users' interest in sharing or giving access to their DNA data with law enforcement from the GEDmatch database. Most participants (27 EG + 29 NE) expressed that they never knew or realized that law enforcement could access these databases. We found that most (29 EG + 24 NE) of the participants did not understand when their DNA data is not only revealing their DNA data but also sharing part of their relative's data. After the scenario, when they realized their data could be used to trace them or their relatives, 24 of the EG participants regretted taking an at-home DNA test. Also, NE participants who were earlier interested in taking a DNA test now changed their decision and expressed no interest in taking the test. (E6) said: *"I wish I would have read their terms, I feel terrible about myself."* Similarly (NE11) told: *"Now, I realize I would be taking someone's right to privacy away by doing it myself"*

Defensive & low expectation of privacy: Seventeen NE participants and 26 EG participants expressed low expectations of the privacy of their data and especially law enforcement surveillance. (NE29) told: *"I think whatever you put online, law enforcement has access to it, your text messages, your phone call history. I'm not going to fight against it even if I do not like it. Because there's nothing, I can do right now in my hands to change that."* Thirteen of the EG participants showed a defensive attitude after knowing law enforcement could access the data. (E11) mentioned: *"if somebody wants it [DNA], they can just follow you to McDonald or Starbucks to get*

it."

Consent & Victimize: Twenty-four NE participants acclaimed the taking a DNA test or sharing DNA data need to be done with the consent of family members. They said sharing DNA data without the family's consent is a breach of privacy. (NE15) said: *"It is a matter of consent; you should always discuss with your immediate family first, it's unfair to share without their consent."* Few other participants perceive they can be a victim of others' decisions. (NE14) told: *"It feels like i can be a victim of some others' ignorance."*

3.3.8 Lack of Knowledge

We also asked the participants if they knew any laws, rules, or policies protecting DNA data stored in public genealogy or DTC-GTs databases. None of the participants had any knowledge about it, but most (53) participants assumed there should be some laws and regulations or should fall under HIPPA policy. We also asked the participants if they read terms and conditions when they share their data. 49 participants said they never read terms and conditions as they are very lengthy, uses convoluted legal languages, and find it boring. Four of the EG participants said that as they want the service or access it immediately, they do not have time to read it. 21 experienced participants did not read through the privacy policies, terms and conditions before taking the test or signing in the testing company's database to look at the dashboard. (E11) said: *"No, I did not read them [policies] properly and do not remember much of it."* Overall, both the groups were having lack of awareness about any privacy policies or data access.

3.3.9 Race & Nation: DNA data sharing

(We did not ask the participants their race, they deliberately mentioned)

Fourteen participants of race and nation expressed their concerns about both testing and sharing. Twelve of these participants are from the NE group. A couple of

EG participants acknowledged this also. The access of the DNA data by law enforcement can trace or put people in surveillance. The concerns were around distrust in law enforcement and government in general. They mentioned that framing, racial discrimination, could be potentially a problem. (NE1) told: *"I am a minority; we have seen a lot of issues with African Americans and the law enforcement, like the recent Floyd case; a lot of things went wrong there. Our people are more vulnerable; history is evident that we are treated unfairly by any authority, even in the hospital. In education also, they think black means not being creative and manipulate science."* A few races felt unsafe about the government accessing the database; they feared the government could target them for deportation. Further, they added as DNA testing result talk about ethnicity, which can make some particular races uninsurable as there are prejudices of health conditions tied to race. Some participants expressed concerns about research organizations obtaining their data. They believed DNA data could be manipulated to target some particular race. Unlike conventional means, when someone shares their DNA data, that affects them and affects others in their family as it inherently shares some DNA data of their family. This was one of the concerns among few particular races. Also, we found that people of a particular nation perceived DNA testing and sharing perceptions can put them in trouble and did not show interest in getting the test. Some races were very interested in getting the test for their benefits mentioning they are privileged. (E20) said: *"As a white person, I am privileged, I would not be targeted; I understand why a black person would resist."* This demonstrates that particular races and nationalism can have strong privacy perceptions against DNA testings and sharing.

3.3.10 Future expectation DNA sharing/ Future Motivation

We asked participants about the possible future motivation to take a DNA test and share DNA data. Nearly half of the participants of the NE group said they would not share their DNA data even if their families and friends were giving it. The fu-

ture motivations mentioned by participants could be success stories of family finding, reconnecting to missed families, to find out family history and biological parents if adopted. In the health field, to know about own and family's health conditions in need of emergency, path-breaking DNA medicines could be likely urged to take the test. Exploring and preserving data about heritage for future generations could be another interest. (NE7) mentioned: *"If needed before having children, I would do DNA testing to make sure that I wasn't a carrier for any genetic diseases. That is important to me; I think here benefits outweighed any possible risks."* Most of the participants of both groups expected at-home DNA testing and sharing to grow more popular as users are curious to explore and connect. Furthermore, as it is the era of globalization, to preserve heritage or lineage and ethnicity, discovering biological parents if adopted these at-home DNA testing and sharing can be predicted to be widespread in the future.

3.3.11 Users' Suggestions

We asked the participants – their suggestions of setting, preference changes or policies changes that would make these DNA testing and sharing platforms more privacy-preserving. This section lists all the users' suggestions that we gathered in Table 3.6 after discussing all the scenarios. Almost all users emphasized "No access to any business (59)", referring that any business organization should not access DNA data. Healthcare doctors should only offer DNA testing services as they believe their data is safe with healthcare than any other business company.

Table 3.6: Privacy-enhancing suggestions by participants

Suggestions	EG	NE	Total (N = 60)}
Any business should not have access to genetic data.	30	29	59
Communicating information about who, when, and how data is being used.	30	28	58
A fine print or small understandable version of the policy.	29	28	57
Different levels of functionalities corresponding to the level of data sharing.	28	26	54
DNA data should be kept private unless both parties accept each other requests to reveal information.	29	22	51
Explicitly tick marks or checkboxes for all data sharing practices.	25	25	50
Not one time opt-in or opt-out. Every access should have opt-in or opt-out options.	20	16	36
DNA data in public genealogy sites should only be used for genealogy.	20	14	34
Genetic data needs to be anonymized without any identifier tracing back.	19	13	32
Secure encryption should be enforced on public genealogy databases to prevent data breaches.	17	15	32
DNA data should be censored to be uploaded to any commercial website.	12	14	26
Genetic data should be under HIPAA, and make it a protected class.	12	12	24
DNA data should be owned and regulated by the government.	5	19	24
In order to upload DNA, family members should be asked and need to give consent.	0	10	10

3.4 Discussion and Limitation

We explored users' perceptions of at-home DNA testing and sharing in public genealogy databases and different entities. Our motivation is that DNA data sharing in open databases are increasingly becoming popular [91]. Acknowledging this, we explored the end-users views, opinions, and assumptions concerning the opportunities and risks of at-home DNA testing and sharing DNA data, their future motivations to take the test and share. We found that people are unaware of the sensitive nature of DNA data. They also wrongly perceive that by doing the at-home DNA testing, they do not give away entire DNA rather they just share small part of their DNA. There is a huge attitude difference between experienced and non-experienced participants. All the implications have been discussed below.

3.4.1 Privacy perceptions

Almost all participants were unaware of the related nature of DNA. People are not mostly comprehending the interconnection of DNA between the family. They were apprehensive when they learned about this phenomenon. We explicitly mentioned it while discussing the "Golden State Killer Case." They felt it is a breach of privacy and merely unethical as someone could be traced without consent. They considered themselves helpless as they could be a victim of others' decisions. DNA data discloses not only ancestry; but also health information such as hereditary diseases, which can be tied to other family members, raising a question of privacy of health data. Like earlier studies [11, 28], we found that users' are ignorant of the jeopardies of DNA data sharing. Though most have some privacy concerns, it is not adequate to stop them from taking a DNA test. However, they would prefer to test with a health organization for an essential and inevitable purpose, not curiosity. Also, they showed interest in learning the privacy policies, T&Cs, or different settings like opt-in and opt-out options, which can guide informed decision-making and current data-sharing

practices. Nevertheless, learning about GEDmatch databases and potential data sharing with different entities (law enforcement, insurance) can hinder sharing of genetic data in public platforms. Hence, offering such testing commercially available for end-users with profit companies necessitates attention, caution, and proper rules regulations must be developed. Another misconception regarding DNA data was that part or piece of DNA data is revealed by giving away saliva or spit, not the entire DNA sequence. Hence, people need to be aware that even if they give away just their saliva, they are giving away their entire DNA sequence. Learning this bothered lots of experienced participants too. Therefore, educating people is very important when they share such sensitive data. Moreover, most people talked about the need for consent while using someone's DNA from DTC-GT or uploaded in open databases. The majority agreed to give consent to law enforcement using the data for violent crime but demanded need to be asked beforehand. Unethical tracking or involuntary surveillance and being dragged into law enforcement cases was highly condemned and spoken about. There are prevailing worries regarding the chances of false convictions, framing, or discrimination against few races. People were pleased to share their data for health and medicine research with credible organizations with all information about how their data is being used and handled. There were prominent privacy perceptions after the insurance scenario leading to decision change.

Further, the need for laws, rules, regulations was raised among users' opinions. They stated that if the DNA testings are commercially available to the end-users, there should be appropriate rules, legislation like HIPPA laws, or government oversight. Also, restraints should be enforced to share and sell people's genetic data with any third parties so that the users' private health information should not be accessible or used without their consent. Race and nationalism are some of the most influential factors in people's privacy perceptions. The participants were kept on pointing out the history of discrimination, biases, prejudices against the race; they mentioned the

"Floyd case" and prejudices in the health sector such as "high pain tolerance" to explain their issues. The African Americans were way more concerned about taking the test and sharing their data with law enforcement, health research organizations, or insurance companies. Similarly, Hispanics were worried regarding the law enforcement access and government access of the data, followed by explaining the forceful deportation of the immigrants. Some Asians were quite worried about if the government or officials could obtain their data in their own country. All these perceptions were obtained from populations educated. We conclude that there need for appropriate risk communication [64, 92, 93] to the user to facilitate informed decision making.

3.4.2 Privacy trade-off

From the analysis, we found that people are pretty concerned regarding their privacy. For example, users were frequently against At-home DNA testing if it would be done in a profit company (not in a health organization), accessed by other entities, especially by insurance, and minimal regulations or accountability of their data practices. Additionally, their concerns revolved around surveillance or usage of their data without their consent. However, if they have an urgent need, for instance, to learn about their health conditions before planning for children, to find family if adopted, they consider at-home DNA testing could be quick and useful and do not debate concerning privacy. Hence, we can infer that if users perceive the benefits are adequate, they would be interested in taking a test. Furthermore, users are highly concerned about various entities such as law enforcement, research organization, insurances, or third parties obtaining the DNA data. They also talked about the possibilities of selling the data to third parties. Still, they acknowledged that these databases could help solve cold cases or advance health research. Besides, people were concerned about being traced, getting dragged into law enforcement investigations, or trouble their family or genetic relatives. Still, they recognized and appreciated law enforcement catching heinous criminals. Some stated that these databases are an excellent tool

for law enforcement, even if that would put a relative behind the bar. They suggested that there should be appropriate regulations and procedures to use these databases. They also mentioned there should be no discrimination against any race or person by their DNA. Users' also recognized utilizing DNA data for advancing health fields. Most users mentioned that the research should be done transparently and by credible research organizations eliminating any biases or prejudice. This indicates there was an evident tension in users regarding usage and privacy. For instance, At-Home DNA testing should be done by health organizations or used by law enforcement to solve violent crimes or credible research organizations with consent and transparency. However, not misuse it like framing, biases, tracking people involuntarily or by insurances.

3.4.3 Attitude differences

We found differences between the experienced group and the non-experienced groups' opinions regarding DNA testing and sharing. In general, there is a high echo of privacy concerns of DNA data in the non-experienced group. Most NE participants disputed the usage of DNA data by law enforcement or insurance and showed little interest in sharing their data with GEDmatch. They felt that taking consent or discussing with family before taking the test is essential than the experienced group. Additionally, the feelings of a victim of others' choices were eminent in the NE group. On the other hand, the experienced group exhibited a relaxed stance toward this though the regrets and anxiety tones were very apparent after discussing the scenarios. The perceived benefits are higher in the experienced group. There were commonly "low expectations of privacy" among users of the experienced group. The rationale behind these contradictions between the experienced and non-experienced groups can be the low efficacy of data control in the experienced participants as they have already shared their data. This might also be a reason behind their defensive attitude. Most of the experienced members regretted their decision when they re-

alized the interconnected nature of DNA. There was a huge lack of understanding about DNA data before we explained our first scenario. Hardly, there was any knowledge about policies, terms and conditions, or data practices of the testing or sharing companies. We believe that our approach to explore people’s perceptions of at-home DNA testing successfully noted the gap between those who have already experienced and who have not taken a test. Consequently, it is apparent that there was a great difference between the two groups about testing and sharing DNA data in public genealogy databases. This gives insight into understanding people’s privacy decision-making, suggesting that there should be adequate and apparent information to assist people in making informed decisions.

3.4.4 Limitations

We investigate the users’ perception of at-home DNA testing and sharing by qualitative approach. A common hurdle is the sample size in such investigations. Participants are typically recruited till reaching data saturation[94]. However, we recognize that this result does not provide for quantitative comparisons. Though we did snowball sampling to get diverse participants, it would have been better to collect data from the average population as the university population is higher educated than the average population. Additionally, we discussed few functionalities of the GEDmatch and few scenarios. Discussing more scenarios and GEDmatch tools might point to added insights.

3.5 Summary

This chapter discusses the critical investigation points in privacy research for at-home DNA testing and sharing. We discussed potential scenarios and investigated users’ perceptions about at-home DNA testing and sharing in public genealogy databases and with various entities like law enforcement, research companies, and insurance. We assessed users’ perceptions with and without experience of at-home DNA testing. We

used a video GEDmatch demo to provide users a better understanding of public genealogy databases to gather their valuable opinion. We discussed scenarios inspired by the current use cases and potential threats to elicit users' judgments. We found that people, in general, need more awareness about the nature and risks of sharing DNA data. Most importantly, DNA data sharing needs further research on possible knowledge delivery methods to inform people of privacy implications. To sum up, in the following part of our research, we want to investigate what and how users want to be briefed about the risks and benefits of DNA data sharing on online platforms.

CHAPTER 4: Study 2 - Participatory design for privacy of online DNA data sharing

4.1 Introduction

In our previous study, we explored users' perception and awareness regarding DNA data. We collected users' views about commercial DNA tests, followed by their expected benefits and concerns. We also studied the users' perception towards third-party tools or public genealogy databases like GEDmatch. In addition, we focused on understanding the users' preferences and perceptions on the disclosure of their genetic information with different types of platforms and entities. We found that users are mostly unaware and lack understanding of the interconnected nature of genetic data. Also, users are not aware of the potential risks of sharing DNA data online. These findings demonstrate a need to create an effective risk communication method to enable users make an informed choice while sharing their DNA data. In order to address this need, we designed a study to develop and design an effective method of risk communication to help users understand both the benefits and risks before they decide to share their genetic data.

Our study is a multi-phase participatory design study in which the participants are involved in the design process to help ensure the results meet the needs and expectations of the stakeholders. The study consists of three main phases or sub-studies, with the goal of understanding and gathering requirements and content to effectively design risk communication messaging, and so involve the user participating in creating and developing these methods. The different phases are described below:

- **Phase 1 - Need-Finding and Co-Design:**

In this phase of the study, we seek to understand what people generally think about the idea of genetic data sharing and what factors they consider or would like to consider when they choose to share their genetic data online with private companies. We also aim to determine how they are presented with the risks and benefits involved and how they would prefer these risks + benefits to be communicated and presented to them. To determine the messaging content, we aim to understand which risks and benefits they seek to understand and how they trade off privacy in response to earning the benefits. Additionally, we gathered their preference on how much time they would want to spend learning about these risks and benefits and at what stage of the site registration this messaging can be presented. Finally, participants were asked to draw on paper their envisioned design. Participants suggested video messaging, info-graphics, and interactive surveys/wizards are as good messaging methods. That is why we conceived five different messaging methods (3 videos, one info-graphics, and one wizard) described below:

- Personal story video - A cartoon character speaking about her own experience of at-home DNA testing highlighting the benefits she got and risks she discovered.
- Conversational story video - Two cartoon characters chatting about the risks and benefits of at-home DNA testing.
- Informational story video - All the information about risks and benefits is explained on a whiteboard.
- Info-graphics - All the information about risks and benefits is explained in a one-page paper.
- Wizards - All the information about risks and benefits are presented according to the user's response.

- **Phase 2 - Designs iterations:**

In this phase of our study, we collaborated with users to iterate, improve, and enhance the designs we created according to users' responses in the first phase. In order to do so, we showed the users all five designs in order to collect their feedback on the designs. Participants were asked questions about relatability, enjoyability, ease of understanding, engagement, and several other questions. Upon completion of this phase, based on the collected feedback, we enhanced our designs to incorporate user-provided input to create the final design versions and eliminate the two most incoherent methods.

- **Phase 3 - Comparing the Designs:**

In the third phase of the study, we tested the final three designs against each other to finalize the most effective design to communicate the risks and benefits of genetic data sharing. This phase focused mainly on comparing users' understanding of the message, user recall of the information, and genetic data sharing intention after encountering the message.

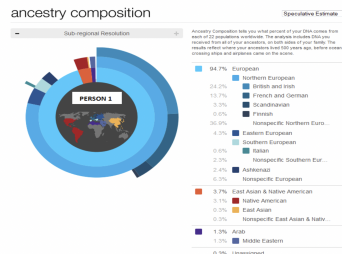
4.2 Phase 1 - Need-Finding and Co-Design

Our goal is to gather feedback from the study, aiming for participants' input on: (1) What information to emphasize to highlight the risks and benefits? (2) Gathering participants' feedback on how to deliver these contents. (3) How long should these communications be?

To understand and gather user's requirement and content of the risk communication message, we conducted phase1 of our study. We aimed to understand what and how users would like to get informed of the benefits and risks involved when they decide to share their DNA data online. To reach this goal, we designed a semi-structured interview study. To get an in-depth point of view, we introduced participants to the possible benefits and risks of sharing DNA data online by showing them slides focusing

Benefits - Ancestry

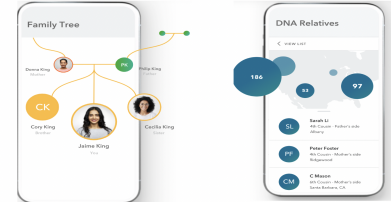
- **Ancestry composition**
- **Your ancestry background**



(a) Slide 1

Benefits - DNA relatives finding

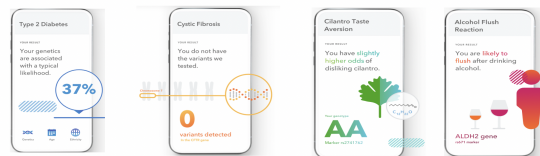
- **Discover people who share your DNA**
- **Contact your relatives**
- **Create a family tree**



(b) Slide 2

Benefits - Health reports

- **Health Predispositions**
- **Carrier Status**
- **well-being and lifestyle choices**
- **Family health history**
- Discover how your body **processes certain medications**.
- **Traits**



(c) Slide 3

Risks - Reveal sensitive info

- Family shares a large portion of DNA, so you implicitly **share your present, past, future family's DNA**.
- Your and yours' **family's health information is shared**.
- Your **data shared with the Profit Research organization**.
- **Anyone** with the internet could access your information.
- **Employer** could access.
- **Insurance** could access.
- **Law enforcement** access.
- Could be **hacked**

(d) Slide 4

Risks - Implications

- Could lead to **Genetic discriminations**
- Could be **denied of services** (example; Education)
- Could lead to **Racial profiling**
- Could control **drug market**
- Could lead to **denial of insurance or insurance premium higher**
- You or your family could be dragged into **police investigations**
- Reveal genetic **data of all family**

(e) Slide 5

Figure 4.1: Screenshots from the slides used in the interview

on both risks and benefits. Figure 4.1 presents the screenshots of the slides. The informational slides enabled us to introduce these risks and benefits to the participants and to provide them with deeper understanding of at-home DNA testing and sharing platforms without requiring them to have previous experiences with such platforms.

The slides content and scenarios were inspired by our literature reviews and current news related to DNA testing. We created topics that summarized the risks and benefits of sharing DNA data. In addition, the slides emphasized the privacy challenges involved with sharing DNA data and the possible risks of identification and other privacy risks.

4.2.1 Recruitment & Demographics

We recruited a total of 10 participants through our University’s mailing list. The study was carried out in March 2022. We complemented recruiting with snowball sampling to get a more diverse representative, where initial participants proposed other interviewees. Among the 10 participants, 5 were male and 5 were female. The participants were aged from 20 to 55 years. Participants had various fields such as graphics designer, computer science doctoral students, health informatics graduate students, etc. Each participant was compensated with a \$10 Amazon gift card. We recruited participants until we felt we had an adequately diverse sample and then found we attained saturation (i.e., no new information attained) during analysis.

4.2.2 Procedure & Analysis

The researchers reached interested participants via email to schedule a video interview. We used zoom video conferencing for all the interviews. The interview was semi-structured, with a set of preliminary questions that were adopted based on participant’s responses. The interviews were recorded via the zoom recording feature. Interviews lasted, on average, 54 minutes. Right after participants joined the zoom interview, we discussed the consent form with them, followed by introducing the research topic with the study purpose. The interview steps are summarized below:

- Step 1 - Collected demographics such as age, gender, and field of profession.
- Step 2 - Discussed general details of at-home DNA testing and asked if they would be interested to take the test and reason behind the response.
- Step 3 - Showed them the slides and asked multiple questions about genetic testing risks and benefits to understand the user’s perception.

The slides explained the benefits of taking at-home DNA testing, such as digging deeper into ancestry, finding genetic relatives around the world, learning about

health predispositions, and knowing more about traits to make better lifestyle and well-being choices. For the possible or potential risks part, we discussed that when someone shares their DNA, they share part of their family's DNA. So their family's sensitive information, such as family or hereditary health revelation, law enforcement access to these databases, genetic discrimination, probable insurance access, racial profiling, and access by third parties.

- Step 4 - Asked participants to perform a drawing task to elicit conceptual ideas on the design and the risk communication messaging.

Participants were asked to verbally explain the method of message delivery and the reason for choosing that message delivery technique. Then, they were asked to draw how they would like to see the message and how that information should flow and to explain their ideas verbally during the drawing activity. We asked participants to draw the designs with pen and paper so that they could effectively explain their ideas and reflections and put forward their initial thoughts. The interview are recorded.

Though we did this drawing exercise virtually, participants described their drawings as they were constructing them, similar to an in-person interview, and extensively talked about it once they sent the picture to us. All participants sent pictures of their drawings via email during the interview. All interviews were transcribed manually for analysis. The interviewer collected and summarized all the preferred content and design ideas.

4.2.3 Results

We first questioned participants about their experiences, perceived usefulness, and perceived threats of at-home genetic testing and sharing with private companies. Below is the synopsis of the participants' answers.

Experience: All participants were largely aware of at-home DNA testing and

were interested in taking one. 90% participants talked about the testing procedure, such as sending saliva or swab to companies for DNA reports. We can infer that the participants are familiar with at-home DNA testing.

Perceived Benefits: We asked participants about their expected benefits and concerns of taking at-home DNA testing. We found that 80% participants wanted to take the test to know more details about their ancestry, ethnicity, and heritage, find DNA relatives and learn about their health predispositions. Participants assumed learning about their health predispositions and traits could help them to change lifestyles and be better prepared for their future.

Perceived Risks: 50% participants expressed some concerns, such as a lack of trust in private companies. They said private companies could sell their data or use it in unapproved ways without any oversight. They also talked about other security issues like data breaches or hacking.

Privacy trade-off: Though 50% participants expressed concerns regarding hacking and data breaches, they were still curious about taking the test to learn about possible health issues and the subject of helping someone in the family. This implies the privacy trade-off point for users is learning about potential health issues for better preparation for the future and supporting family members.

Next, we showed participants all the slides illustrating points about the advantages and menaces of sharing DNA data. Then, we asked participants which pieces of information we should emphasize and should be highlighted to assist people in making informed decisions. Tables 4.1, 4.2 summarize the topics participants suggested that must be emphasized or spoken aloud to users.

We collected participants' suggestions on methods or settings for DNA data sharing platforms to enhance data sharing privacy. The below table 4.3 represents the summary of our findings.

Finally, we collected participants' feedback on methods to deliver these risks and

Table 4.1: Participants' suggestions of the possible risks that must be included in the risk and benefit message

Risks that must be discussed	Percentage of participants
Third party access	100%
Hacking or data breach	100%
Reveal health information about family	100%
Law enforcement access	80%
Employer or insurance access	80%
Genetic discrimination	80%
Involuntary surveillance or Dragged into police investigation	70%

Table 4.2: Participants' suggestions of the possible benefits that must be included in the risk and benefit message

Benefits that must be discussed	Percentage of participants
Health predispositions	90%
Traits	70%
Know your Ancestry	60%
Wellbeing & lifestyle	60%
Family finding	40%
Genetic medicine	20%
Participate in research	20%

benefits contents. **Story-telling** emerged as a key approach. People described that story-telling videos and info-graphics could be very pleasing, helpful, understandable, or effective ways to pass or convey the message. We compiled all information about what contents, which way, and how long the risk communication strategies should be comprised of. Our findings are summarized in the Table 4.4:

4.2.4 Implication of Phase 1: Suggestions to initial designs

Based on the findings summarized in Table 4.4, we created the following designs:

Designs

- **Videos**

- Personal story video - A cartoon character speaking about her own expe-

Table 4.3: Suggestions on Privacy enhancing settings for DNA sharing platforms

Suggestions	% of Participants
Permission of all parties involved before relatives match	100%
Option to allow or not allow to share data with any entities	100%
Option to completely delete data and information anytime	100%
Notify when someone matches more than certain limits	100%
Company ask authorization before someone can see / access contact details	100%
Notifying user when someone else accesses info	100%
Option to set a certain limit of match for access of contact details	100%
Should be asked access each time law enforcement asks for and inform details about the case and reasons for access	100%
Notify about policy updates of the company	100%
Option to auto-delete data after a certain period of time	90%
Ask each time for any research use and inform details of the research and rights of the participant	90%

rience of at-home DNA testing highlighting the benefits she got and risks she discovered.

- Conversational story video - Two cartoon characters chatting about the risks and benefits of at-home DNA testing.
- Informational video - All the information about risks and benefits is explained on a whiteboard.

- **Info-graphics** - All the information about risks and benefits is explained in a one-page paper.

Table 4.4: Participants' designs suggestions

Designs suggestions			% of participants
How do they want info to be communicated?	Watching video	With personal examples or stories	70%
		With data & information	50%
	info-graphics		70%
	Interactive survey/wizards		20%
	Reading policy		10%
How much time are they willing to spend?	To watch a video	1-2 mins	30%
		2-3 mins	60%
		3-4 mins	10%
	Length of info-graphics	One standard page (2-3 min read)	50%
		2 standard pages (2-3 min read)	50%

- **Wizards / Interactive survey** - Wizard provides information based on the user's choices.

All these videos are within 2-3 mins' length. We also designed one info-graphic page and one wizard or interactive survey highlighting all the benefits and risks participants suggested in table 4.1, 4.2 and aiming to enhance them by gathering users' feedback in phase 2 of our study.

4.3 Phase 2 - Designs iterations phase

4.3.1 Methodology

We utilized a semi-structured interview to enhance our initial designs, along with collecting users' thoughts and points of view on the risk communication message.

Initial Designs (Version 1): We created five initial designs based on the design suggestions from users that we collected in phase 1. We added all the risks and benefits points the users would like to be highlighted in the message. All the video transcripts are in (cf. Appendix A.). Below, we describe the different designs:

- **Personal story video**

- Characters / Persona and Setting - A female cartoon is sitting in her living room (informal setting). We used machine-generated voice-over.
- Main idea - A woman gives illustrations of her own experiences. First, she discusses her experience with at-home DNA testing. Then, she discusses all the benefits, such as finding her DNA relatives, etc. Finally, she discusses all the risks she realized after sharing her DNA. Figure 4.2 presents a screenshot of the video frame.
- Reasoning - The reason behind choosing this method of story-telling is that 70% of participants suggested that informing people in a casual environment through personal examples would be a helpful way to make people understand the message.

- **Conversational story video**

- Characters / Persona and Setting - Two cartoon characters (one male and one female) discussing at-home DNA testing while shopping and dining in a market (informal or casual setting). We used machine-generated voice-over for both characters.
- Main idea - At the start, in a shopping store, the female is bringing the at-home DNA testing kit to the male, saying the kits are at a discounted price and explaining all the benefits of the test. The guy gets excited as his adopted aunt could find her family. Then the guy asks the girl to dig

deeper about the tests. While casually dining, they discuss and browse about commercial DNA tests, where they get to know the possible risks of taking the test. Figure 4.3 video Frame presents a screenshot of the video frame.

- Reasoning - As 70% participants suggested that informing people in a casual environment through scenarios or conversation with personal examples would be a helpful way to make people understand the message.

- **Informational video**

- Characters / Persona and Setting - Formal setting. Handwriting on whiteboard and explaining. We used machine-generated voice-over for both characters.
- Main idea - First, the video presents the data about the number of people who have already taken the test discussing all the benefits of taking at-home DNA testing. Then explains the possible risks with current genetic protection laws and gives statistics about law enforcement's usage of the public genealogy databases until now. Figure 4.4 presents a screenshot of video the frame.
- Reasoning - We designed an informational video as 50% participants pointed out that they would like to see data and information about risks that we mentioned in slides, such as; how law enforcement accesses the DNA data, how many times genealogy databases have been used, etc. Therefore, this informational video is data-driven; this video doesn't involve any personal examples or casual way of conversation.



Figure 4.2: Personal Story Video

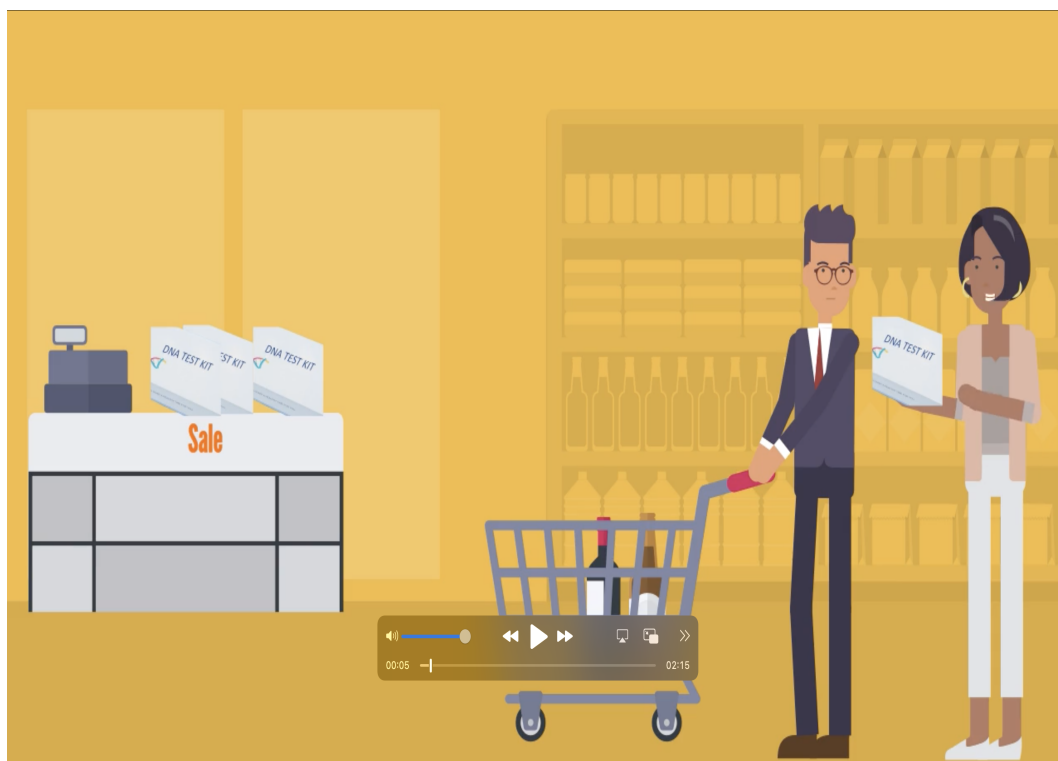


Figure 4.3: Conversational story video



Figure 4.4: Informational story video

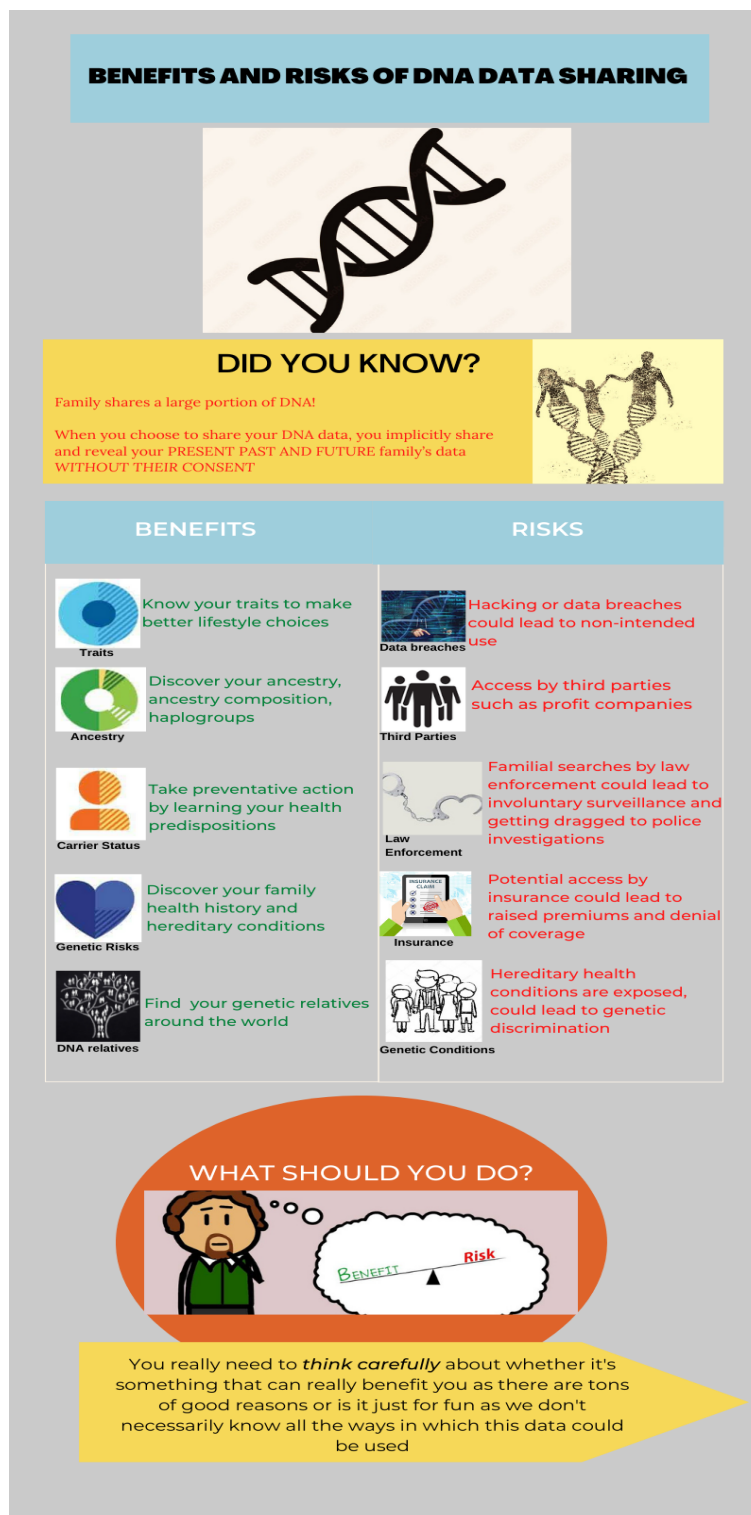


Figure 4.5: Informational story video

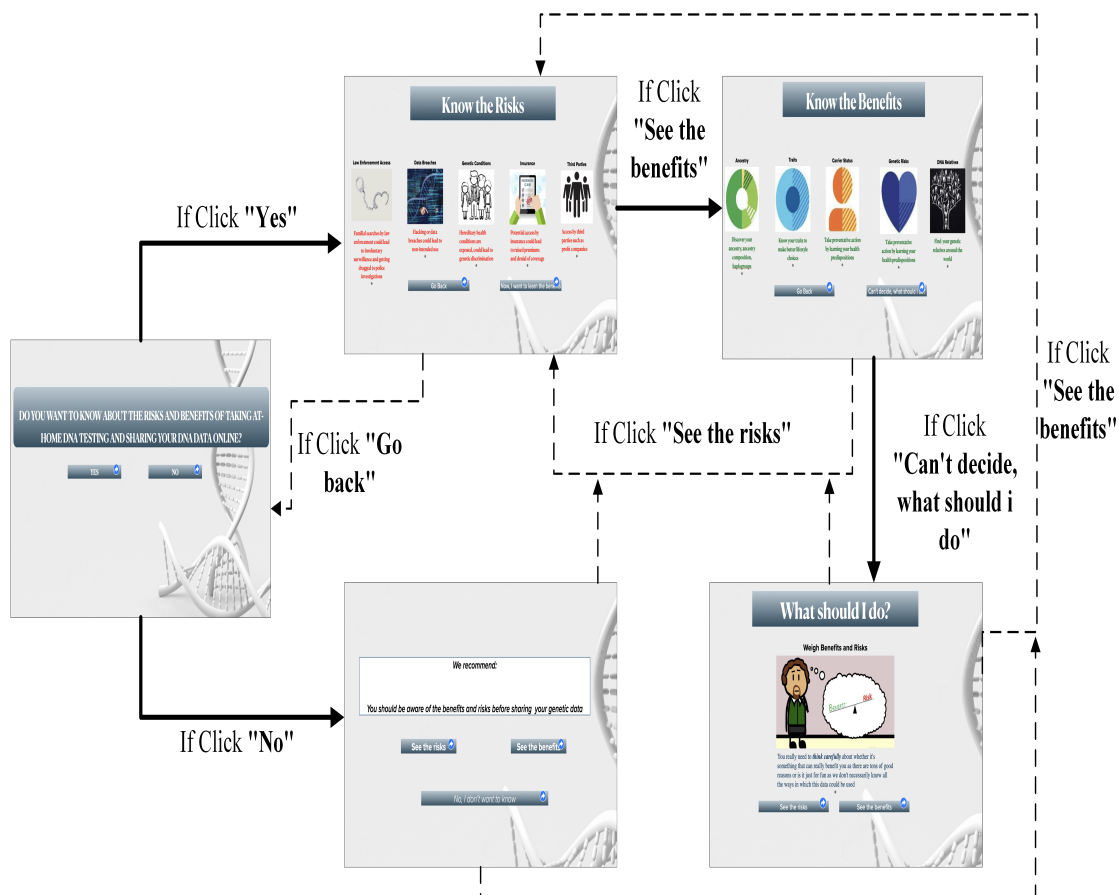


Figure 4.6: Wizard flow

- **Info-graphics** - The title of the info-graphics is "Benefits and risks of DNA data sharing". It is one page. All benefits and risks are listed side by side. User can decide their pace of learning the risks and benefits by reading themselves. The reason behind designing an info-graphic is that 70% participants suggested that info-graphics is an excellent method to outline the risks and benefits. Figure 4.5 present the info-graphics.
- **Wizards / Interactive survey** - Figure 4.6 presents the screenshots of wizards. First, the user is asked if they want to know about the benefits and risks. According to their response, it takes them to the window where it shows them the benefits and risks or recommends them first to know the benefits and risks. Thus, the Wizard provides information based on the user's choice. The user has absolute control over what information they wants to know. The reason behind designing wizards is that 20% of participants in phase 1 suggested that wizards / interactive surveys could be an effective method to communicate the risks and benefits.

4.3.2 Recruitment & Demographics

All participants were recruited through the university research mailing pool. All the data was collected in September 2022. Among the seven participants, three were male, and four were female. The participants were aged from 20 to 50 years. Each participant was compensated with a \$10 Amazon gift card. Initially, we interviewed seven participants and showed them the version 1 designs. We compiled their feedback to incorporate into the initial designs.

4.3.3 Procedure & Analysis

The researchers contacted interested participants via email to schedule a zoom video interview. The interview was semi-structured, with a set of basic questions about the design. The questions mainly focused on the design's positive and negative aspects

and the message’s understandability, which varied depending on the participants’ responses. The interviews were recorded via the zoom recording feature. Interviews lasted on average an hour. The study was approved by our university Institutional Review Board (IRB). We started the interview by asking general feelings about the topic, followed by all-around questions about the overall look and feel of the designs one by one and how they relate with the personas in the videos (if present), ease of understanding of the message, and enjoyability. The designs were shown in random order for each participant to eliminate the ordering bias. We asked them to talk about the best and worst parts of each design and to provide their suggestions to improve them. Participants were asked to illustrate their ideas verbally to enhance the designs. We then focused on participants’ perceptions of the message and asked them to explain or summarize to the interviewer what they understood from each message. Next, we asked participants about the helpfulness of the message in making an informed decision concerning sharing genetic data. We then asked them their suggestions on how to make the designs and messages more helpful for an average user to understand the topic and enhance their interest in learning about it. Finally, we collected participants’ demographic information at the end of the interview. The interviewer collected the participants’ feedback and revised the designs accordingly after the first seven interviews. Then, the three most incoherent designs were eliminated, and modifications were made to the remaining designs leading to version 2 designs. Then, the version 2 designs will be demonstrated to another 7 participants to gather their feedback, followed by developing the final design versions.

4.4 Design Versions 1 Feedback Results

In this section, we summarize all the feedback gathered from participants on design versions 1. The “informational video” was the most preferred, and the “Conversational story video” design was the least liked. The input on the personal story video is in table 4.5. Four out of 7 participants said they have earlier exposure to personal story

videos where a person tells about their experience with a product. They gave an example of weight loss products, hair products, etc. They also mentioned that these stories only talked about the benefits of the product but not the risks or side effects of it. They particularly liked the casual setting for story-telling and personal examples illustrated by the persona. All participants said the design and message were easy to understand. They disliked the robotic voice used in the video and asked to add a human voice and character rather than an animated character and robotic voice.

Designs: Relatability, Engaging and Enjoyable and Easy to Understand

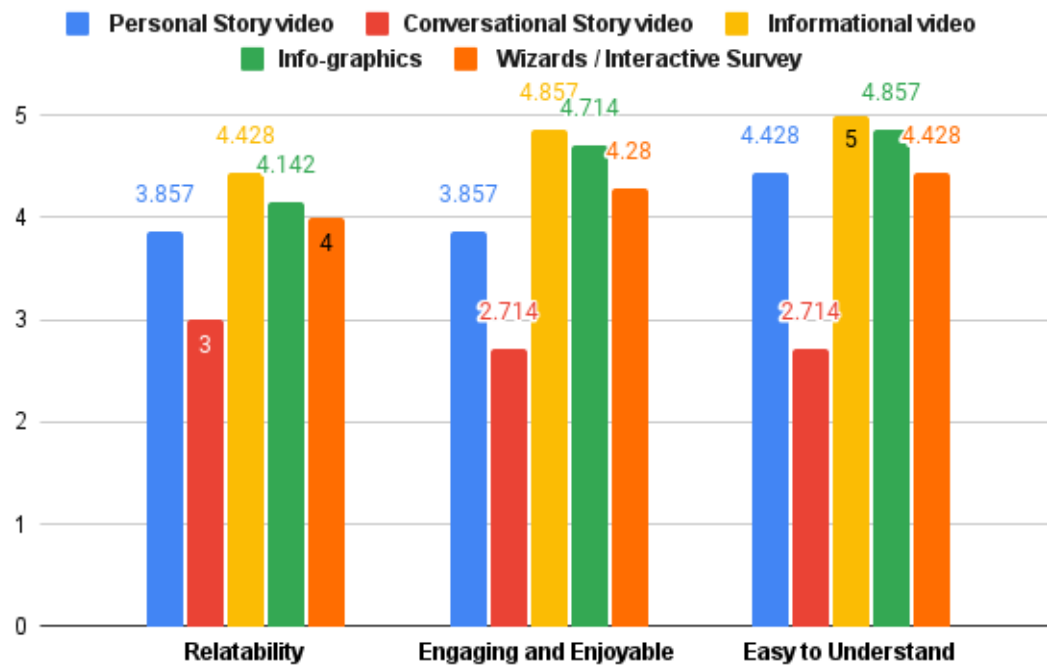


Figure 4.7: Participants' feedback On designs: Relatability, Engaging and Enjoyable and Easy to Understand

The conversational story video is the most disliked design. Three out of 7 participants did not find this story-telling easy to understand and engaging. Therefore, we will eliminate this design and will not develop it further. No participant has any earlier exposure to this type of messaging. Participants (6) liked the informational

Table 4.5: Participants’ feedback and suggestions on “Design Version 1”

Liked (N=7)	Disliked (N=7)	Enhancements (N=7)
Personal Story video		
Casual environment (5)	Robotic voice (5)	Add data as an example (5)
Personal examples (5)	Monotonous presentation (4)	Pop-up data and figures while explaining (4)
Talked about both benefits and risks (4)	No data or evidence (4)	Add more movements (2)
Not authoritative figure (3)	White color character (2)	Natural / human voice (2) Replace the character (2)
Conversational Story video		
Every day or causal environment (3)	No data or evidence (3)	Add data as an example (4)
Personal example (1)	Robotic voice (3)	Pop-up data and figures while explaining (3)
Talked about both benefits and risks (1)	Could not connect (3)	Use real humans (2)
One male and one female (1)	Expressions (2)	Natural / human voice (2)
Informational video		
Presentation (6)	Authoritarian voice (2)	Change the voice (4)
Data and evidence (6)	Lack of suggestion on “what to do next?” (1)	Use female voice (2)
Talked about both benefits and risks (4)		Use characters (2)
Transitions (4)		Add suggestion for people (1)
Info-graphics		
Brief and coherent (5)	Colors and graphics (2)	Add bright colors (2)
Good colors (5)	Last section is too long (2)	Shorten some text (2)
Talked about both benefits and risks side by side (5)		Add some data and figures (1)
Not overwhelming (3)		
Wizards / Interactive Surveys		
Brief and coherent (6)	Lack of data and evidence (3)	
Good colors (5)	No audio (2)	Add data and figures (3)
Presented both benefits and risks side by side (5)		Add suggestions on what people should do (2)
Easy to keep with the pace (5)		Add audio (2)
Not overwhelming (3)		

video the most. They enjoyed the presentation of it, and all participants found this video easy to understand and engaging. Two participants have earlier exposure to this type of messaging in the context of environmental pollution and marketing of some products. Two participants complained about the male voice used in the video and found the voice authoritative. Participants liked the briefness of the info-graphics and found it easy to understand and engage with. They suggested adding some bright colors, data, and figures to make it more visually appealing. Four participants have earlier exposure to info-graphics in the context of medicine and vaccines. Most (6) liked the wizard/interactive survey method of presenting the risks and benefits and engaging and enjoyable. Some (2) participants recommended audio should be added along with text and should include data and figures. All participants have earlier exposure to this type of wizard or interactive survey in the context of fitting guides to buy clothes, shoes, glasses, etc. Detailed feedback is in the table ???. Figure 4.7 presents participants' responses to relatability, understandability, engagement, and enjoyment. Over all, the informational video was the most preferred followed by info-graphics. The conversational story video is the least preferred followed by the personal story video.

4.5 Design Versions 2 and Final designs

The informational video and info-graphics were the most liked designs by participants. That is why, we decided to focus on improving the design of infographics and informational videos related to DNA and DNA testing based on feedback from 7 participants.

To improve the design of the infographics, we made several changes, including updating the fonts and colors to make them more visually appealing as it was commented by participants. We also shortened the text and removed the last section to make the information more concise and easier to understand.

For the informational video, participants requested a female voiceover to remove the

Table 4.6: Participants' feedback and suggestions on "Design Version 2"

ENHANCEMENT FEEDBACK (n=7)	
Info-graphics	Informational Video
Add more context on DNA and DNA testing (7)	Used a human voice for better engagement (7)
Add examples of risks and benefits, such as traits and ancestry (5)	Slow down the animation to improve comprehension (7)
Improve color scheme (4)	Balanced the content to avoid bias towards risks or benefits (3)
Replace/remove technical terms like "haplogroup" with simpler language (3)	Remove the use of hands in the video (3)

authoritarian voice feeling, so we used a machine-generated female voice to provide a consistent and engaging narration. We also made small improvements in the video for the overall user experience of the video and to increase engagement with the content.

We again collected feedback from another 7 participants on the design versions 2 of infographics and informational video. The interview included the same questions to allow participants to provide detailed feedback on what they liked and disliked about the designs. Table 4.6 is the feedback we received from them.

Final designs: The feedback from the participants was taken into consideration, and the final version of the infographic has been updated to provide a more clear and informative representation of DNA testing and its associated benefits and risks.

The infographic starts by providing general information about DNA testing, including the fact that over 26 million people have taken a commercial DNA test, and genes are passed down through families. It also highlights the fact that sharing your DNA data can reveal information about your present, past, and future family members' DNA data without their consent. Additionally, it notes that over 70% of people with European descent can be identified as having a family member who has taken a DNA test.

To address the feedback regarding the need for more context, the infographic now

includes more information about at-home DNA testing and what it entails. This is done to ensure that readers who are not familiar with DNA testing can better understand the concept and context behind the infographic.

In response to the feedback requesting examples, the infographic now provides specific examples of the benefits and risks associated with DNA testing. For example, it notes that DNA testing can provide information about traits and wellness, such as identifying lifestyle factors that can be modified, such as avoiding alcohol if you have alcohol reflux. The benefits are highlighted in green, while the risks are highlighted in red.

To address the feedback on color, the background of the infographic has been changed to a white and light blue color scheme, which makes the text and images easier to read and understand. Additionally, technical words such as haplogroup have been removed and replaced with simpler, more understandable language.

Overall, the updated version of the infographic incorporates the feedback from the participants to provide a clearer and more informative representation of DNA testing and its associated benefits and risks. It is hoped that this version will better educate readers on the subject of DNA testing and encourage them to make informed decisions about their DNA data.

4.6 Phase 3 - Comparing the designs

The designs generated in the previous phases have been evaluated as part of the research. During Phase 2, our focus was on evaluating and comparing each of the designs in terms of ease of use, engagement, relatability, and recall. A total of three designs were generated and considered for comparison.

To further investigate the effectiveness of these designs, a between-subject study was conducted as the next step. The study aimed to determine which design proved to be the most effective, understandable, and helpful in aiding users' comprehension of the risks and benefits associated with sharing DNA data. Additionally, we sought

to gather insights on how the designs could be improved.

Our research questions were formulated as follows:

- RQ1: Among the designs, which design(s) did the users find the easiest to understand and recall most effectively?
- RQ2: What potential areas for improvement can be identified for the designs to enhance their effectiveness and user experience?

By addressing these research questions, we aimed to gain valuable insights into the comparative effectiveness of the designs and gather suggestions for enhancing their overall quality and user experience.

4.6.1 Methodology

To comprehensively evaluate the designs, we conducted a between-subject study involving 55 participants assigned to each individual design. Our participant pool was sourced from social networks and the university research pool, ensuring a diverse sample. Utilizing the Qualtrics survey platform, we administered the study, allowing for efficient data collection and analysis.

At the outset of the survey, participants were queried about their familiarity with at-home DNA testing and public genealogy databases. This initial assessment aimed to gauge their prior knowledge and understanding of the subject matter. Following this, we provided a concise explanation of at-home DNA tests and public genealogy databases to ensure that all participants had a foundational understanding of these concepts.

To facilitate participants' comprehension and provide context, we created website prototypes resembling actual at-home DNA testing websites and public genealogy databases. These prototypes served as visual aids, enabling participants to better grasp the intricacies and gain a realistic perspective of the platforms. After examining the prototypes, participants returned to the survey and were presented with either

the infographics or informational video design, thereby being assigned to a specific group.

Next, participants were asked a series of questions to evaluate their understanding and interpretations of the communication within the assigned design. This allowed us to gain insights into how participants comprehended and derived meaning from the presented information.

Following the assessment of the designs, participants completed a comprehensive questionnaire. The questionnaire encompassed various aspects, including ease of understanding, content recall, and intention to share DNA data. By probing these areas, we aimed to obtain quantitative data regarding participants' subjective experiences and perceptions related to the designs.

Finally, participants were given an opportunity to provide feedback on the designs, enabling them to share their thoughts and suggestions for potential improvements. This valuable feedback served as a crucial step in refining and enhancing the overall effectiveness and user experience of the designs.

4.7 Evaluation

In this section, we will evaluate the demographics of both group and both the risk communication methods.

4.7.1 Demographics

We conducted a Chi-Square test to assess the demographic characteristics of our sample, and our findings revealed no significant differences between the two groups, each group with 55 participants. Out of all the participants, 40.6% were identified as male, while 55.4% were female. The analysis did not indicate any significant disparity in terms of gender distribution between the groups ($\tilde{\chi}^2 = 2.3$, $p = .5$).

In terms of age distribution, the majority of our participants (79.2%) fell within the 18-29 age range. Additionally, 11.9% were between 30-39 years old, and 5.9% were

aged 40-49. Our analysis did not yield any significant differences in age distribution between the groups ($\tilde{\chi}^2 = 9.5$, $p = .05$).

Moreover, only 16.8% of the participants identified themselves as Hispanic, Latino, or Spanish. Among the remaining participants, the majority (52.5%) identified as White/Caucasian, 22.8% as Southeast and Southwest Asian, 7.9% as Black/African American, 3% as East/central Asian, 3% as Middle Eastern/North African, and 3% as Native American/Alaska Native. The Chi-Square test did not reveal any significant differences in ethnic or racial composition between the groups ($\tilde{\chi}^2 = 6.5$, $p = .3$).

Regarding the highest level of education attained, 38.6% of the participants had completed some college/associates' degree/technical degree, 25.7% held a graduate or professional degree, and 18.8% had obtained a high school degree or equivalent. The Chi-Square test did not find any significant differences in educational attainment between the groups ($\tilde{\chi}^2 = 2.1$, $p = .7$).

4.7.2 Evaluation of risk communication

The results of the study are presented in the form of participant responses to various statements related to the info-graphic and video designs. The responses were measured on three criterias: a) Easy to understand and use b) Content recall c) Sharing intention.

Easy to Understand and Use: To evaluate the participants' understandability, usability we adopted SUS scale questions [95]. The responses were measured on a Likert scale, with higher scores indicating stronger agreement with the statement. The mean (median) scores for each statement and the associated statistical test results are provided.

Participants' perception of the complexity of the message was measured using the statement "I found the [message] unnecessarily complex." The info-graphic design received an average score of 2.1 (with a median score of 2), while the video design received an average score of 1.9 (with a median score of 2). Statistical analysis using

the U-test revealed no significant difference between the two designs ($p = .49$). These results suggest that participants did not perceive a significant disparity in complexity between the info-graphic and video designs. Participants' perception of the ease of understanding the message was evaluated using the statement "I thought the [message] was easy to understand." The info-graphic design received a mean score of 4.1 (with a median score of 4), while the video design received a mean score of 4.2 (with a median score of 4). The results of the U-test indicated no significant difference in perceived understanding between the two designs ($p = .35$). This suggests that participants did not perceive a substantial distinction in the ease of understanding between the info-graphic and video designs.

Participants' perception of the need for technical support to understand the message was assessed using the statement "I think that I would need a technical person's support to understand the [message]." Both the info-graphic and video designs received low mean scores, with 1.8 (with a median score of 2) for the info-graphic design and 1.7 (with a median score of 2) for the video design. The results of the U-test indicated no significant difference in the perceived need for technical support between the two designs ($p = .95$). This suggests that participants did not consider either design to require extensive technical assistance for comprehension.

Participants were asked to rate the integration of risk and benefit messages in the design using the statement "I found the risk and benefit messages in the [message] were well integrated." For the info-graphic design, the mean score was 4.0 (with a median score of 4), while for the video design, the mean score was 4.1 (with a median score of 4). The U-test results showed no significant difference in the perceived integration between the two designs ($p = .59$). This indicates that participants did not perceive a substantial disparity in the degree of integration of risk and benefit messages in the info-graphic and video designs. The statement "I thought there was too much inconsistency in the [message]" received mean scores of 2.1 (median of 2) for both the

info-graphic and video designs. The U-test results suggest no significant difference in perceived inconsistency between the two designs ($p = .37$).

Participants' perception of how quickly most people would understand the message was captured by the statement "I imagine most people would understand the [message] very quickly." Both designs received mean scores of 4.0 (median of 4) for this statement. The U-test results indicate no significant difference in participants' expectations of quick understanding between the two designs ($p = .75$). The statement "I found the [message] very cumbersome to understand" received a mean score of 2.2 (median of 2) for the info-graphic design and 1.9 (median of 2) for the video design. The U-test results suggest no significant difference in perceived level of cumbersome understanding between the two designs ($p = .21$).

Participants' confidence in understanding the message was measured by the statement "I felt very confident understanding the [message]." Both designs received mean scores of 4.1 (median of 4) for this statement. The U-test results indicate no significant difference in participants' confidence levels between the two designs ($p = .553$).

For the statement "I needed to learn a lot of things before I could understand the [message]," the mean score was 2.0 (median of 2) for the info-graphic design and 1.9 (median of 2) for the video design. The U-test results suggest no significant difference in participants' perception of the need to learn additional things before understanding the message between the two designs.

The participants' ratings on the understandability and usability of the message (video or info-graphics) are presented in Table 4.7. Utilizing the Mann-Whitney test, we investigated potential differences between individuals who watched a video message versus those who viewed an info-graphic message. However, our analysis did not uncover any significant distinctions between the two groups.

Table 4.7: Evaluating the informative message between those who watched a video message and info-graphic message including Mean, Median, and test statistic values of the Mann-Whitney with the reported p-value. The message refers to either Info-graphic or Video

Statements	Info-graphic Mean (Median)	Video Mean (Median)	Test Statistic U-test (p)
1) I found the [message] unnecessarily complex.	2.1 (2)	1.9 (2)	1180 (p=.49)
2) I thought the [message] was easy to understand.	4.1 (4)	4.2 (4)	1385 (p=.35)
3) I think that I would need a technical person's support to understand the [message].	1.8 (2)	1.7 (2)	1280 (p=.95)
4) I found the risk and benefit messages in the [message] were well integrated.	4.0 (4)	4.1 (4)	1339 (p=.59)
5) I thought there was too much inconsistency in the [message].	2.1 (2)	2.1 (2)	1389 (p=.37)
6) I imagine most people would understand the [message] very quickly.	4.0 (4)	4.1 (4)	1313 (p=.75)
7) I found the [message] very cumbersome to understand.	2.2 (2)	1.9 (2)	1101 (p=.21)
8) I felt very confident understanding the [message].	4.1 (4)	4.2 (4)	1350 (p=.553)
9) I needed to learn a lot of things before I could understand the [message].	2.0 (2)	1.9 (2)	1215 (p=.68)

Content Recall - Informational Video Vs Infographics: In this section we will compare the content recall of the two risk communication methods to understand which method facilitate effective information retention. Table 4.8 presents details about the percentages of contents recall by participants.

Recall of communication about Trait and Wellness - A comparison of data sets from the video and infographic responses regarding the "Trait and Wellness" content reveals intriguing patterns. When analyzing the feedback, a large number of respondents from both categories couldn't recall any content, with approximately 24% from the video group and 37% from the infographic group answering "none" when asked to recount the information.

Among those who could remember, the information recall appears to be more detailed and specific in the video group compared to infographic group. In the infographic group, many responses were vague, with only a handful, approximately 11%, remembering key aspects like genetic diseases, alcohol-related traits, and lifestyle factors. In contrast, nearly 33% of the informational video group provided more detailed recollections, mentioning inherited health risks, genetic diseases like sickle cell anemia, and the impact of traits on lifestyle. By conducting Mann-Whitney U-test, we found significant difference between video group and infographic group with p value of 0.013.

Additionally, in the infographic group, responses were largely generic, with phrases

Table 4.8: Evaluating the "Content Recall" between those who watched a video message and info-graphic message including percentages and test statistic values of the Mann-Whitney with the reported p-value.

Topic	No Recall		All details Recall		Test statistics U-test (p)
	Video Group (%) (N=55)	Infographic Group (%) (N=55)	Video Group (%) (N=55)	Infographic Group (%) (N=55)	
Trait and Wellness	23.63%	36.36%	32.72%	10.9%	1131 (p =.013)
Law Enforcement}	20%	38.18%	47.27%	21.81%	1042.5 (p =.003)
DNA Relative Finding	29.09%	43.63%	49.09%	25.45%	1161 (p =.025)
Data Breaches	20%	16.36%	67.27%	56.36%	Not Significant

such as “something about alcohol,” “traits you may have,” and “lifestyle factors.” This suggests that while the infographics might have engaged the viewers, it didn’t necessarily facilitate effective information retention. Contrastingly, in the video group, many respondents not only recalled the primary topic of “Trait and Wellness” but also provided specific details like the role of DNA testing in identifying genetic diseases and traits linked to alcohol consumption. These responses demonstrate that video can lead to enhanced recall and understanding of complex topics.

Recall of communication about Law enforcement - By analyzing the responses related to “Law enforcement” content from both the video and infographic feedback, we found that 19% in the video group and 38% in the infographic group is lacking retention of information. Among respondents who recalled content, the video group appears to have a more comprehensive understanding of the topic. Approximately 48% of this group could recite specifics such as law enforcement using DNA databases to solve cold cases, law enforcement’s potential use of DNA data for surveillance, the threat of involuntary surveillance, and the possible discrimination due to DNA databases. On the contrary, only about 21% of respondents from the infographic group mentioned specifics. By conducting Mann-Whitney U-test, we found significant difference between video group and infographic group with p value of 0.003.

The participants’ memory of the “Law enforcement” topic shows apparent contrast between the video and infographic groups. In the video group, the responses encompassed various aspects, like using DNA databases to solve cold cases and the potential privacy implications of sharing DNA data. The responses demonstrated an understanding of the broad spectrum of legal and ethical issues associated with DNA testing and law enforcement. On the other hand, the infographic group responses focused mainly on the potential for involuntary surveillance, with less emphasis on other aspects of law enforcement’s use of DNA data. This could imply that the infographic

was less successful in conveying the complete range of associated issues compared to the video. Overall, we can say video seemed to facilitate better overall understanding and retention of the topic, suggesting it might be a more effective medium for presenting complex issues.

Recall of communication about DNA relative finding - When comparing responses about "DNA relative finding" content, the video and infographic groups show apparent differences in recall ability. About 29% of the video group participants could not recall any content, compared to a considerably more 43% in the infographic group, suggesting that the video was more effective at conveying and reinforcing this specific concept. Among those who did recall content, about 49% of the video group could recollect exact aspects, such as finding relatives based on DNA matches.

In contrast, the infographic group responses were less precise and mixed. Only about 26% of this group could remember specific aspects. The video group demonstrated a broad understanding of DNA relative finding, encompassing various aspects such as the potential to find global DNA matches and the potential privacy implications related to sharing DNA data. The infographic group, on the other hand, mainly focused on the basic notion of finding relatives through shared DNA, with fewer responses mentioning the potential for worldwide connections or the detailed implications of DNA testing. By conducting Mann-Whitney U-test, we found significant difference between video group and infographic group with p value of 0.025.

Thus, the video seems to have been more effective in conveying a broad and detailed understanding of "DNA relative finding."

Recall of communication about Data breaches - In our review of participant responses about "Data breaches," we found that 19% of those exposed to the video and 15% of those exposed to the infographic reported no recollection of the content, suggesting that the infographic may be slightly more effective in retention. Among those who did recall the topic, 67% from the video group detailed aspects

of data breaches such as hacking risks, DNA data misuse, and potential third-party interference. This suggests a more wide learning of the issue.

In contrast, only about 56% of the infographic group could remember specific details about data breaches, especially focusing on risks of hacking and the likely exposure of personal data. The video group showed a greater awareness of potential repercussions of data breaches, such as the potential misuse of DNA data by insurance companies. They comprehended that these breaches could disclose sensitive genetic predispositions with negative impacts. The infographic group, while acknowledging the risk of data stealing via hacking, did not expound as much on potential consequences of such breaches. By conducting Mann-Whitney U-test, we did not find significant difference between video group and infographic group.

In conclusion, although both groups recognized the risk of data breaches in DNA testing, the video group indicated a slightly deeper understanding of potential consequences.

Sharing Intention: Furthermore, participants were asked to provide their likelihood of taking an at-home DNA test after viewing the message. Our examination revealed no noteworthy differences between the groups ($U = 1019$, $p = .07$). Consequently, we found that 32.5% of participants who viewed the info-graphics message and 37.5% of those who watched the video expressed an not likelihood to take an at-home DNA test.

In addition to the above question, participants were also asked about their inclination to share their DNA in the Open genealogy database after viewing the message. Similar to the previous analysis, no significant differences emerged between the two groups ($U = 1107$, $p = .15$). Notably, we discovered that 64.2% of participants who viewed the info-graphics message and 77.1% of those who watched the video responded "No" to this question, indicating their unwillingness to share their DNA in the Open genealogy database.

4.7.3 Feedback about the video:

Here we lists all the positive and negative feedback about the video risk communication method, aimed at understanding viewer preferences and identifying the aspects that made the video an effective risk communication method.

About 85% of the participant, appreciated the use of visuals such as illustrations, graphics, and animations. These respondents found that the use of visual aids, paired with auditory explanations, greatly enhanced their understanding of the material and helped maintain their engagement throughout the video. The clarity and simplicity of the explanation emerged as another highly valued aspect. Approximately 67% of the participants liked how complex terms and ideas were explained in a simplified, easy-to-understand manner. This was particularly appreciated in the sections where the risks and repercussions of the subject matter were discussed. About 46% of participants found the overall organization and structure of the video appealing. They liked the logical flow of information, the balance between both the risks and benefits information, and the clarity of points made. A subset of this group specifically appreciated the concise and equal attention given to the pros and cons, aiding in the balanced understanding of the topic. Approximately 46% of respondents acknowledged the clear audio and the articulate voiceover. Participants found that the quality of the narration contributed to their understanding and overall learning experience. Around 25% of the participants particularly appreciated the historical context provided in the video, especially regarding sickle cell disease among African Americans. They found this perspective valuable for understanding the broader social and racial implications of the subject matter. Additionally, 23% of the respondents applauded the video for being informative and concise, effectively delivering substantial information within a brief runtime which kept the viewers engaged, despite the complexity of the topic.

We also delved into areas that viewers found less desirable in the video by asking the question, "What aspects did you not like in the video?" About 14% of the respon-

dents felt that the video was too long and sometimes repetitive, suggesting a need for more concise and streamlined content. Approximately 11% of viewers were not satisfied with the video's audio quality and the narrator's delivery. They commented on the varying volume levels, strange noises, and the narrator's tone of voice and speaking style, which some described as disinterested or lacking confidence. Also, 9% of participants felt that the visuals, particularly the animations, were too simple or not engaging enough. They suggested that more professional, dynamic, and engaging visuals could help to maintain viewer interest and facilitate comprehension. Lastly, 7% of viewers found the information such as medical terminologies, and acronyms were not adequately explained.

Interestingly, a substantial proportion of respondents (59%) mentioned they liked every aspect of the video, which could be interpreted as a generally positive reception. In conclusion, while the video was generally well-received, improvements can be made in terms of audio quality, visual presentation, and the simplification of complex information. Attention to these details can enhance the effectiveness and audience reception of the video.

4.7.4 Feedback about the infographics:

We also aimed to analyze participant feedback on the effectiveness and perceived shortcomings of an infographic as a tool for risk communication. Respondents' feedback was categorized into positive and negative sentiments, elucidating distinct aspects such as aesthetic appeal, content comprehension, and structural organization. In terms of positive feedback, a significant majority (63%) admired the design of the infographic, commending its color scheme, font, layout, and the integration of graphical elements. These features contributed to the aesthetic allure of the infographic and facilitated better comprehension of the information. The simplicity of the infographic emerged as another liked attribute, appreciated by 59% of the participants. This appreciation was linked to the infographic's ability to convey information with clarity

and ease, indicating its effectiveness in communication. Moreover, 45% of respondents valued the organization of information in the infographic, especially appreciating the clear segregation of sections and the use of bullet points, which augmented readability. A subset of these respondents specifically commended the structure of presenting pros and cons, reinforcing the balanced presentation of information. Impressively, the content of the infographic resonated with 79% of respondents. They found it informative, even learning new information, underlining the educational efficacy of the infographic. Some participants particularly valued the incorporation of examples and statistics, accentuating that concrete details promote understanding.

On the other hand, feedback also highlighted areas for improvement. A segment of respondents (22%) found issues with the infographic's design, particularly criticizing the font type, size, and aesthetics of images, which they felt detracted from its professional appeal and readability. Additionally, 19% of participants found the layout and organization too complex and busy, indicating a need for a more streamlined presentation. Finally, 16% of respondents perceived the information as too basic or brief, suggesting a demand for more comprehensive, detailed content.

Interestingly, despite these criticisms, a sizable proportion (36%) of respondents expressed no objections to the infographic, suggesting a general level of satisfaction with its current state. Overall, while the infographic demonstrated efficacy as a risk communication tool, feedback suggests room for enhancement, particularly in design aspects, structural simplicity, and depth of content. This would serve to improve the infographic's overall effectiveness and appeal.

4.8 Discussion

The third phase of our research aimed to assess the effectiveness of two distinct forms of risk communication infographic and video in shaping participants' comprehension, recall, and DNA data sharing intention concerning at-home DNA tests and participation in open genealogy databases. It is noteworthy to mention that our anal-

ysis found no significant differences between the two groups in demographic variables such as age, gender, race, or level of education. This similarity in demographic characteristics bolsters the reliability of the comparative evaluations derived from the two groups' responses.

In terms of risk communication, our analyses unveiled no considerable differences in the participants' perceptions of complexity, ease of understanding, usability, necessity for technical support, comprehension of risk and benefit messages across the two modalities. This indicates that the infographic and video formats were comparably effective in their comprehensibility and perceived complexity. While both forms of communication demonstrated similar degrees of comprehension and perceived complexity, participants exhibited noticeable differences in the level of detail recalled. Generally, the video format outperformed the infographic in terms of content recall across various topics, such as "Trait and Wellness", "Law enforcement", and "DNA relative finding". This observation may suggest that the engaging and multi-sensory nature of video presentations could aid in boosting information retention, especially when the subject matter is elaborate or intricate.

Upon querying about their likelihood of undergoing an at-home DNA test after exposure to the information, participants' responses across both modalities showed no significant differences. However, a larger proportion of participants who viewed the video message expressed a reluctance to undergo the test compared to those who viewed the infographic message. Similarly, when asked about their willingness to share their DNA in an open genealogy database, a higher percentage of video viewers expressed unwillingness compared to infographic viewers, even though the difference was not statistically significant. These findings suggest that while both modalities were similarly effective in communicating risk information and influencing participant understanding, the video format might exert a slightly stronger influence in discouraging participants from engaging in at-home DNA testing or sharing their

DNA in open genealogy databases. This effect could be due to the video’s more robust ability to visually and audibly deliver complex information, weave narrative storytelling, and invoke emotional responses [96].

Furthermore, it’s important to highlight that a significant proportion of participants expressed reluctance to partake in the suggested actions (undergoing an at-home DNA test or sharing their DNA in open genealogy databases) after viewing the messages. This could potentially reflect the effectiveness of risk communication in both formats, as they successfully highlighted potential concerns and influenced individual decision-making processes.

4.9 Limitations

This study, despite its valuable insights, was not without limitations. One key restriction pertains to the sample size and population. The population involved in the study primarily consisted of students and university affiliates, and the total number of participants was relatively small. This significantly limits the generalizability of the study results to broader, more diverse populations. Our sample was primarily young, educated, and likely more technologically adept, potentially influencing their comprehension and reception of the risk communication delivered via infographic and video designs. Previous research has suggested that age, educational attainment, and digital literacy may affect individuals’ comprehension and interpretation of health risk information [97]. Consequently, our findings may not be directly applicable to older, less educated, or less technologically adept populations. Moreover, the limited sample size may have affected the statistical power of the study, potentially influencing our ability to detect significant differences in participant responses across the two designs. This limitation may partially explain the lack of statistically significant differences observed in participant perceptions and action inclinations in relation to the infographic and video designs. Despite these limitations, the study provides important preliminary insights into the comparable effectiveness of infographic and video designs

in communicating risk information. Further research in this area, incorporating larger and more diverse populations, will be instrumental in refining our understanding of the most effective modes of risk communication for various audiences.

4.9.1 Conclusion

In our study, we found that both the informational video and the infographic did not really change how well people understood the topic based on the SUS scale or their interest in sharing DNA data or doing at-home DNA tests. Nonetheless, we noticed that a few more participants in the video group (37.5%) compared to the infographic group (32.5%) were not interested in doing at-home DNA tests. When we asked if they wanted to share their DNA data in the Open genealogy database after they saw the video or infographic, the responses were comparable between the two groups. However, more people from the video group (77.1%) than the infographic group (64.2%) said 'No,' meaning they did not want to share their DNA. This shows us that the informational video may have been better at making people think twice about sharing their DNA than the infographic.

Additionally, the group that watched the informational video remembered substantially more about the pros and cons of the online DNA data sharing than the group that looked at the infographic.

Future research should consider exploring the specific elements of video and infographic designs that most significantly impact participant understanding and action inclinations. This could provide further insights into optimizing risk communication strategies in different contexts. Furthermore, exploring potential moderators, such as prior knowledge or attitudes towards the topic, may provide a deeper understanding of how different individuals respond to various modes of risk communication.

CHAPTER 5: Study 3 - The Effectiveness of risk communication for privacy of online sharing of genetic data

5.1 Introduction

The findings from Study 1 highlighted that people are often unaware and confused about the nature of DNA data and existing policies and laws. However, we found that users are concerned about their genetic data after learning about the potential risks of sharing genetic data. To mitigate this problem, in study 2, we designed multiple methods of communication to inform users of the benefits and risks of sharing DNA data with private companies. We also collaborated with users to develop an effective risk communication methodology by finding or exploring users' requirement and preferences, followed by enhancing the designs by collaborating with users, and finally testing the designs with users to finalize the preferred method of risk communication in our second study.

Based on the findings from the studies, we have several primary research challenges we will now address. The first challenge is to deploy a risk communication method to facilitate users' decision-making regarding sharing DNA data with third parties or in public genealogy databases. The second challenge is to develop a privacy-preserving platform for users to control their data privacy even after sharing on public genealogy databases. The challenge here is usability vs. privacy, that is not to take away the benefits that users can gain by sharing their DNA, such as finding biological family, participating in medical research to advance health fields, the same time. It is a challenge to enable users to preserve, maintain and control the privacy of their DNA data. Therefore, my study focused on the addressing the following research questions.

- **RQ1:** How does the deployment of the developed risk communication technique, the informational video, influence behavior in public genealogy databases?
- **RQ2:** What is the users' response to this method and how well do they understand the risks and benefits of sharing DNA data when exposed to the informational video?
- **RQ3:** How do spatial and temporal factors influence the effectiveness of this risk communication?

To do this, we performed study with participants who were not previously involved in our research. These participants were then divided randomly into one of three categories: two experimental groups and a control group. Initial discussions explored their understanding of public genealogy databases and concerns about DNA data sharing. Following this, Group 1 watched a video detailing the advantages and disadvantages of DNA data sharing and then decided their data sharing preferences using a privacy policy form. Similarly, Group 2 watched brief videos explaining the risks and benefits of DNA data sharing before making their decisions. The control group, on the other hand, only reviewed the privacy policy form before deciding their preferences, without any prior exposure to the benefits or risks. We observed that our interventions had a substantial effect on the attitudes of the test groups towards DNA data sharing, while the attitudes of the control group remained stable.

5.2 Methodology

We aim to evaluate the effectiveness of our risk communication messages in informing users about the advantages and threats of sharing DNA data. We conducted online interviews and surveys with 63 participants who were 18 or older and had not participated in our previous studies. We recruited them through the university, flyers, and Craigslist. We divided them into three groups of 21: Group 1, Group 2, and a Control Group. We used Signupgenius to schedule online meetings with them

on Zoom. Before the meeting, we sent them a consent form and asked them to agree to participate and to allow us to record the session.

First, we asked them what they understood about public genealogy databases. Then, we showed them some tools and features of these databases, such as one-to-many matches, family trees, and one-to-one matches. We also asked them about their interest and concerns in sharing their DNA data. Then they completed a survey about their intention to share their DNA data with different entities.

Next, we randomly assigned them to one of the three groups. Group 1 watched a video that explained the benefits and risks of sharing DNA data for each tool and feature. This video was based on the messages we developed with users in our previous study. It had pictures and graphics to illustrate the points. After watching the video, they saw a privacy policy form with the same options as GEDmatch, such as opting in or out of law enforcement searches. They had to choose their settings and then decide whether to share their DNA data or not. Group 2 saw the same privacy policy form as Group 1, but with brief and precise messages about the benefits and risks of each option. They also had to choose their settings and then decide whether to share their DNA data or not. The Control Group did not see any messages about the benefits and risks of sharing DNA data. They only saw the privacy policy form with textual explanations of each option. They also had to choose their settings and then decide whether to share their DNA data or not. Figure 5.2 shows the form filed for Opt In and Opt out policy.

Figure 5.1 shows the steps of the study.

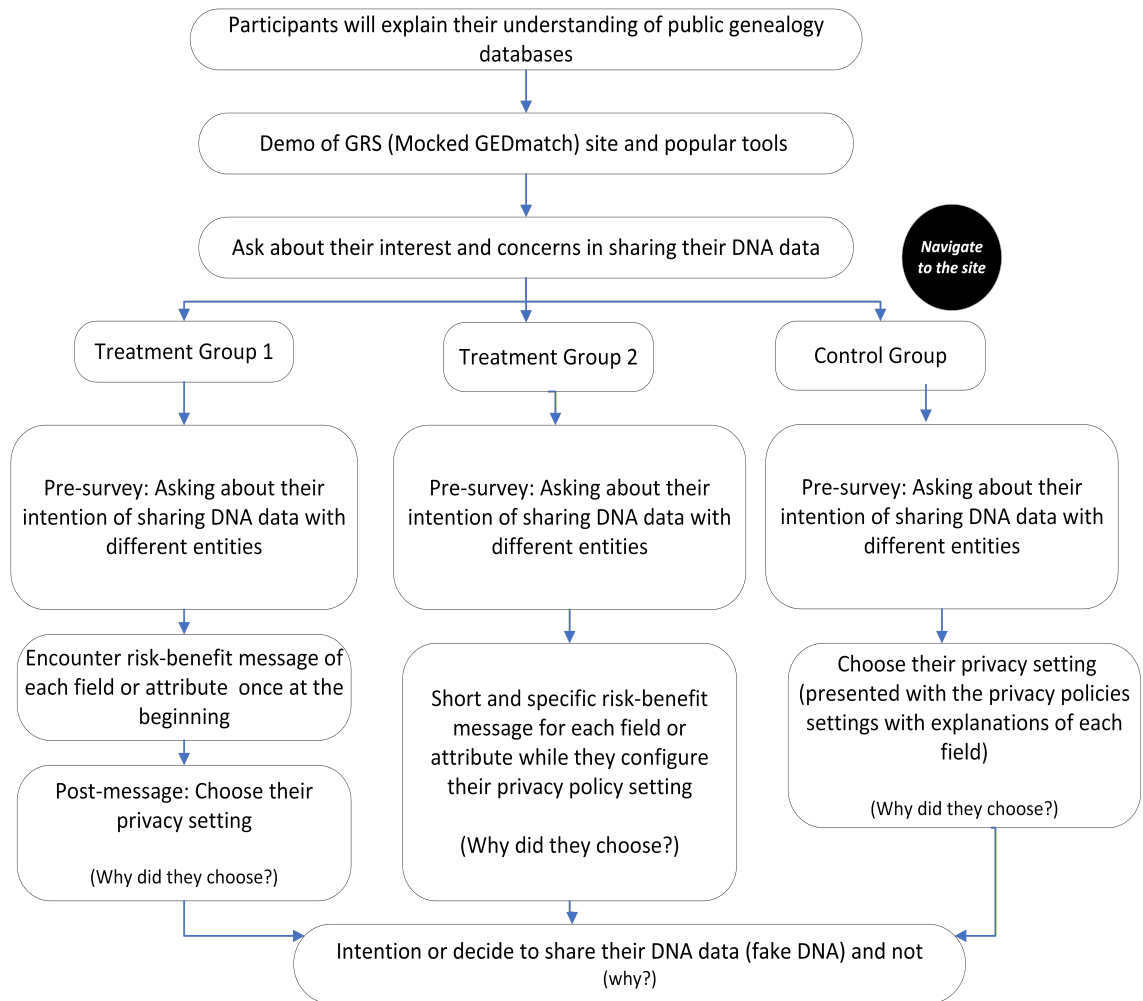


Figure 5.1: Study flow

5.3 Analysis

We used a between-subjects design for this experiment. We placed participants either in the control or one of the test groups. We collected demographics and general privacy concerns. We also collected their sharing intentions, and awareness through presurvey and post-message to understand the effectiveness or impact of the message. Additionally, we collected and compared each group's reasoning for their decision regarding privacy policy configuration and changes in decisions (pre-survey Vs. post-message) to comprehend the impact of the message on their decision-making.

Privacy Options

Select one of the following privacy options for this kit. (required) ⓘ

☒ Opt In

Most Popular

We will compare your DNA kit to all other kits in the GEDmatch database to find your matching genetic relatives. Kits in the database include those submitted by users undertaking personal genetic genealogy research, adoptee searches, users (including law enforcement) attempting to identify unidentified human remains, and law enforcement attempting to identify perpetrators of violent crimes. Your kit **WILL** be compared with kits submitted by law enforcement to identify perpetrators of violent crimes. The operators of GEDmatch encourage everybody to select this option.

☐ Opt Out

We will compare your DNA kit to all other kits in the GEDmatch database to find your matching genetic relatives. Kits in the database include those submitted by users undertaking personal genetic genealogy research, adoptee searches, and users (including law enforcement) attempting to identify unidentified human remains. Your kit **WILL NOT** be compared with kits submitted by law enforcement to identify perpetrators of violent crimes.

☐ Research

This kit will not be shown in match result reports generated for other kits. Genealogy and Genetic Genealogy require the sharing of information. This option is provided primarily for artificially created research kits, but may be used for regular uploads if you have specific reasons for doing so.

☐ Private

This kit will not be available for any matching. The kit will be in the database, and it will be batch processed, but no comparison results will be shown unless this privacy setting is changed by you later.

Figure 5.2: Form field for Opt In/Opt Out policy with textual explanation

5.4 Results

Demographics: Regarding the demographics of our sample, 30.6% of our participants were male, and 69.4% were female. Therefore, 33.3% of participants in the Group 1, 28.6% in Group 2, and 30% in Control were male, whereas 66.7% of participants in the Group 1, 71.4% in Group 2, and 70% in Control were female. By performing a Chi-Square test, we did not find significant differences in terms of our participants' age ($\tilde{\chi}^2 = .118$, $p = .9$). Most of our participants (46.8%) were between the ages 18-29. Also, the majority of our participants (93.5%) in all groups were not Hispanic, Latino, or Spanish ($\tilde{\chi}^2 = 2.03$, $p = .36$). Also, 52.4% of participants in all groups were White or Caucasian, 19% were South, Southeast, or Southwest Asian, 14.3% were Black or African American, and 9.5% were East or Central Asian ($\tilde{\chi}^2 = 5.8$, $p = .83$). The current level of participants' education in all groups as follows: 35.5% of participants had graduate or professional degree, 32.3% had bachelor's degree, 19.4% had some college, associate's degree or technical degree, and 12.9 had high school degree or equivalent ($\tilde{\chi}^2 = 11.01$, $p = .08$).

5.4.1 Understanding about at-home DNA testing

Before the intervention, data indicated that 52.4% of Group 1, 47.6% of Group 2, and 33.3% of the control group had no previous knowledge regarding at-home DNA testing. The rest of the participants showed diverse degrees of awareness and experience with this technology.

Some participants brought up companies like 23andMe, which garner DNA samples from various biological materials such as saliva and hair. These samples are then matched against a comprehensive database to extract ancestry information. Some participants indicated that there are a limited number of tests available, with a primary focus on ancestry, although some tests can also be used for diagnostic purposes. Among the participants, 4.8% from Group 1, 23.8% from Group 2, and 9.5% from the control group reported having personally conducted an at-home DNA test in the past. These individuals shared their experiences with the testing process, underlining how the results delivered lineage details from both sides of their family, as well as information on health predispositions. They noted that these health data points were periodically updated as the company collected more data.

Overall, the responses from the rest of the participants presented a broad range of familiarity with specific companies, the sources of DNA samples, and the potential uses of DNA analysis.

5.4.2 Interest in taking at-home DNA testing

Prior to the intervention, data showed that 66.7% of participants in each of Group 1, Group 2, and the Control Group were inclined to undertake at-home DNA testing to gain better understanding on various aspects of their personal and ancestral information. These participants mentioned various motivations, from a desire to gain valuable health insights to the potential uncovering of lost relatives. The appeal of exploring their ancestry, genetic predispositions, and family history from the convenience of

their homes was another common sentiment, as was the potential health-care decision guidance that could stem from such testing. By performing the Kruskal-Wallis test, we found no statistically significant difference among groups ($p=1.000$). The remaining 33.3% of participants from each group showed no interest in at-home DNA testing, citing predominant concerns about data privacy and security. Their fears revolved around possible misuse or unauthorized access to their genetic data by large corporations and potential negative ramifications like denial of life insurance if the test results revealed any health issues. Some expressed satisfaction with their current knowledge about their family background and health and saw no need for further exploration.

Post-intervention, interest in at-home DNA testing dropped to 42.9% in Group 1 and 38.1% in Group 2. Their declining interest was mainly tied to growing concerns about privacy and the potential risks involved. Sentiments ranged from doubts about data protection, the potential for misuse of DNA, and fears of inadvertently violating their own and their family's privacy, to the perception of risks outweighing the benefits. Some participants mentioned having already undergone testing and hence saw no added value, while others voiced reservations about making their data public. A few participants considered DNA testing to have too many uncertainties compared to the benefits offered. For the control group, interest remained steady at 66.7% even after navigating the policy and reading the text explanations. By performing the Kruskal-Wallis test, we found no statistically significant difference among groups ($p=.144$). We performed Cochran's Q test for comparing pre and post intervention interest to take an at-home DNA testing. We found significant difference for ($p=.014$) in Group2 and we did not find significant difference for Group1 ($p=.059$). Table 5.1 presents the results for pre and post intervention or navigation stage.

Overall, while all groups initially showed strong interest in at-home DNA testing, Group 2 experienced a significant decrease post-intervention due to increased aware-

Table 5.1: Interest in taking about at-home DNA testing

	Interest in taking at-home DNA testing			
	Pre-Intervention or navigation	Post-Intervention or navigation	Difference in %	Pre Vs Post-Intervention Cochrans' Q Test p-value
Group 1 (n =21)	66.7%	42.9%	-23.8%	<i>Not Significant</i>
Group 2 (n =21)	66.7%	33.1%	-33.6%	<i>.014</i>
Control Group (n =21)	66.7%	66.7%	0%	<i>Not Significant</i>
Kruskal-Wallis Test p-value	<i>Not Significant</i>	<i>Not Significant</i>		

ness of potential risks and privacy issues followed by Group1. In contrast, the control group maintained their interest, indicating that their perceived benefits from such testing remained strong.

5.4.3 Perceived Benefits of at-home DNA testing

The primary motivations for conducting at-home DNA testing, according to participants across the three groups, revolve around two significant themes: ancestry and health. These themes highlight the dual purpose of DNA testing as an investigative tool for personal history and a preventative health measure. In the realm of ancestry, 41.3% of respondents in total (across all the groups) expressed interest in understanding their lineage and ethnicity. They expressed enthusiasm for uncovering unknown family members, tracing family origin, and determining ethnic makeup. Uncovering this information is perceived as not only a personal journey but a way to share findings with family and possibly build or extend family trees. Health insights were the primary area of interest for 58.7% of respondents. Participants see DNA testing as an opportunity to uncover underlying health issues, genetic predispositions, and potential future health risks. The ability to take proactive steps towards healthcare was another common thread, with participants appreciating the opportunity to plan and potentially adjust lifestyle choices based on the findings. Minor motivations included the convenience of at-home testing, mentioned by 9.5% of participants, and the ability to control personal health data, mentioned by 5.2% of respondents. These points underline the shift towards more autonomous healthcare management.

There were, however, some dissenting voices, with 3.1% of respondents seeing no benefit to DNA testing. This segment expressed skepticism about the value of such testing, indicating that the information might not be particularly useful or that they would prefer professional testing environments over at-home methods.

5.4.4 Perceived Concerns of at-home DNA testing

Before the intervention, only 42.9% of Group 1, 38.1% of Group 2, and 33.3% of the control group voiced concerns about at-home DNA testing. The main anxieties revolved around data privacy and potential misuse of their genetic information. Participants were wary about how their results would be stored, accessed, and potentially shared with third parties. The comprehension of ownership and usage rights over their genetic data, along with the risks of unauthorized access or hacking, surfaced as significant concerns. While some participants resigned themselves to an assumed lack of control over their data, others underscored the necessity for additional information and stringent regulations to ensure responsible use of genetic data by the companies. Despite these concerns, an overall relaxed attitude towards DNA data sharing was observed. However, after the intervention, the situation dramatically changed. Concerns about at-home DNA testing jumped to 100% for both Group 1 and Group 2. The intervention likely heightened their awareness of the potential risks and privacy issues associated with DNA testing, leading to unanimous increase in apprehension.

Post navigation and after reading policy-related information, 38.1% of the control group started expressing concerns. This increase, compared to the pre-navigation stage, likely came from a better understanding of potential risks and policy implications surrounding DNA data sharing by reading textual information about the policy such as the opt in and opt out policy where law enforcement can access their data. This group still had the least amount of concern compared to Groups 1 and 2. Table 5.2 presents the results for pre and post intervention or navigation stage.

In summary, while concerns were relatively low across all groups prior to the inter-

Table 5.2: Perceived Concerns of at-home DNA testing

Perceived Concerns of at-home DNA testing			
	Pre-Intervention or navigation	Post-Intervention or navigation	Difference in %
Group 1 (n =21)	42.9%	100%	+57.1%
Group 2 (n =21)	38.1%	100%	+61.9%
Control Group (n =21)	33.3%	38.1%	+4.8%

vention, there was a substantial increase in apprehension post-intervention in Group 1 and Group 2 participants. Meanwhile, the control group, despite maintaining a relatively lower level of concern, did still show a noticeable increase in apprehension after being exposed to more information about data privacy and genetic testing policies.

5.4.5 Interest in sharing DNA data in GEDmatch

Before the intervention, 57.1% of Group 1, 52.4% of Group 2, and 42.9% of the control group participants expressed interest in sharing their DNA data in public genealogy databases. They were drawn by the prospect of uncovering distant family connections, exploring their genetic heritage, and expanding family trees. Some individuals were also motivated by potential societal benefits such as advancements in medical research and groundbreaking treatments derived from big data insights. By performing the Kruskal-Wallis test, we found no statistically significant difference among groups when we asked these questions: whether they were interested in sharing DNA data in GEDmatch ($p = .646$). However, the remaining participants from each group - 42.9% from Group 1, 47.6% from Group 2, and 57.1% from the control group - were not interested to share their data on GED match. Their resistance primarily came from concerns about potential misuse of their personal data, particularly by private companies. They feared their genetic information could be exploited for targeted advertising, especially for health-related products or services. In addition, they expressed apprehension about potential encounters with newly discovered relatives that could lead to the uncovering of undesirable truth.

After the intervention, interest to share DNA data on GEDmatch, a public genealogy database, dropped dramatically to 14.3% in Group 1 and 23.8% in Group 2. The primary reason for this decrease was concerns about data privacy, particularly the risk of hacking and unauthorized access, as this database is publicly accessible. The participants were also worried about the implications for their family's DNA data privacy as their DNA data sharing decision could affect other in the family. Notably, some of these participants were still comfortable to take a DNA test with at-home DNA testing companies. They reported higher trust in these at-home DNA testing companies than public genealogy databases. After reviewing the policy and navigation information, interest in data sharing among the control group participants also slightly dropped to 38.1% ($p = .655$). The main reasons cited were fears of potential hacking and the possibility that their data could be accessed by law enforcement agencies, as indicated in the opt-in and opt-out policy. By performing the Kruskal-Wallis test, we found no statistically significant difference among groups when we asked these questions: whether they were interested in sharing DNA data in GEDmatch after intervention among the groups ($p = .209$). By performing Cochran's Q test, we found significant difference ($p = .003$) within Group1, within Group2 ($p = .007$), pre-intervention and postintervention interest in sharing DNA data in GEDmatch. Table 5.3 presents the results for pre and post intervention or navigation stage.

In summary, while there was initially a moderate level of interest in sharing DNA data in public genealogy databases across all groups, the post-intervention stage saw a significant decrease in willingness to share DNA data in Group 1 and Group2 due to heightened privacy concerns and the realization of potential risks.

Table 5.3: Interest in sharing DNA data in GEDmatch

	Interest in sharing DNA data in GEDmatch			
	Pre-Intervention or navigation	Post-Intervention or navigation	Difference in %	Pre Vs Post-Intervention Cochrans' Q Test p-value
Group 1 (n =21)	57.1%	14.3%	-42.8%	.003
Group 2 (n =21)	52.4%	23.8%	-28.6%	.007
Control Group (n =21)	42.9%	38.1%	-4.8%	<i>Not Significant</i>
Kruskal-Wallis Test p-value	<i>Not Significant</i>	<i>Not Significant</i>		

5.4.6 Perceived Benefits and risks of GED match

Comparing and contrasting the three groups before and after the intervention and post navigation it's apparent that there was a clear shift in perspective. Pre-intervention, all groups identified various benefits of sharing DNA data on public genealogy databases. These included uncovering new familial connections, becoming aware of potential health risks, gaining insight into one's biological lineage, determining geographic origins, accessing information about specific demographic groups, aiding in potential medical research, getting to know one's genetic makeup, and facilitating adopted individuals in their search for biological parents and lost relatives. The comprehensive data analysis offered by public genealogy databases, due to their integration of data from multiple companies and tests, was also seen as a benefit. The additional advantage of aiding in the capture of serious criminals was recognized post-intervention.

However, the respondents also expressed concerns related to the sharing of DNA data on these platforms. These concerns involved possible harassment from DNA matches, the accuracy of relatedness determinations, tracking by tech companies, data sharing with third-party apps, sale and retention of personal data, uncomfortable revelations about relatives, and the risk of personal information exposure.

After the intervention, Group1 and Group2 developed additional concerns regarding potential risks associated with insurance companies that could lead to increased premiums, involvement in police investigations, and the exposure of private DNA

data of relatives. The control group, which was not subjected to any intervention, maintained its initial perspectives throughout the study.

5.4.7 Intention of sharing their personal information in GEDmatch:

Additionally, we inquired with participants regarding their willingness to share personal information in the event that they create a profile on GEDmatch, upload their data, and adjust their privacy settings accordingly. Table 5.4 presents the results for pre and post intervention or navigation stage.

Sharing real name: Before intervention, participants within all three groups - Group 1, Group 2, and the Control Group - displayed a tendency to use their real names when registering on GEDmatch, citing transparency as a key motivation. The belief that there was nothing to hide was common among them, implying that their actual identity was already accessible online, hence a sense of resigned privacy. In Group 1, 80.9% of participants expressed no compulsion to hide their identity. They thought that a fake name wouldn't provide effective privacy protection. Group 2 showed a slightly lesser inclination, with 61.9% of participants opting to use their real names. They highlighted their desire for recognition and connection, with an understanding that their real names were likely known to the company, given they'd already provided their DNA. In the control group, 76.1% of participants underlined the convenience of using their real names and the minimal perceived privacy risks, despite some concerns about potential compromises to anonymity. By performing the Kruskal-Wallis test, we found no statistically significant difference among groups ($p = .357$).

Post-intervention, a notable shift was seen within Group 1 and Group 2. The scores dropped significantly to 57.1% in Group 1 and 33.3% in Group 2, indicating that the intervention had influenced their opinions and behaviors around online privacy. Conversely, the Control Group, the score rose to 90.4%. They mentioned they felt the site is safe to use and perceived their data would be well protected and using the

Table 5.4: Intention of sharing their personal information in GEDmatch

Intention of sharing their personal information in GEDmatch				
	Pre-Intervention or navigation	Post-Intervention or navigation	Difference in %	Pre Vs Post-Intervention/ navigation Cochrans' Q Test p-value
Using real name in registration page				
Group 1 (G1) (n =21)	80.9%	57.1%	-23.8%	p=.025
Group 2 (G2) (n =21)	61.9%	33.3%	-28.6%	p=.014
Control Group (CG) (n =21)	76.1%	90.4%	+14.3%	p=.046
Kruskal-Wallis Test p-value	Not significant	G2 vs CG p < .001 G1 vs CG p = .012		
Not using an alias or prefer using their real name as profile name				
Group 1 (G1) (n =21)	76.2%	33.3%	-42.9%	p = .003
Group 2 (G2) (n =21)	61.9%	52.4%	-9.5%	Not significant
Control Group (CG) (n =21)	57.1%	52.4%	-4.7%	Not significant
Kruskal-Wallis Test p-value	Not significant	Not significant		
Using regular email address				
Group 1 (G1) (n =21)	71.4%	42.8%	-28.6%	p=.034
Group 2 (G2) (n =21)	85.7%	33.3%	-52.4%	p<.001
Control Group (CG) (n =21)	57.1%	57.1%	0%	Not significant
Kruskal-Wallis Test p-value	Not significant	Not significant		

real name would be helpful to find more real genetic connections. By performing the Kruskal-Wallis test, we found statistically significant difference between Group 2 vs control group ($p < .001$) and Group 1 vs Control group ($p = .012$). By performing Cochrans' Q test, we found significant difference ($p=.025$) within Group1, within Group2 ($p=.014$), pre-intervention and postintervention interest in using real name in GEDmatch. Post navigation, by performing Cochrans' Q test, we found significant difference ($p=.046$) within control group as their interest went up.

Using alias name: Before intervention or navigation, when considering their registration process on GEDmatch, the majority of participants from each group

showed a preference for using their real names, not using an alias. In Group 1, 76.2% of participants preferred this approach, as did 61.9% in Group 2 and 57.1% in the Control Group. These participants saw no necessity in concealing their real identities and prioritized transparency and connection. They believed that using an alias could obstruct their purpose of joining the website, i.e., finding relatives. They highlighted the importance of due diligence in understanding the company's background and safeguarding their privacy before sharing their data. By performing the Kruskal-Wallis test, we found no statistically significant difference among groups ($p = .531$).

In contrast, a noticeable segment of participants favored using an alias, citing privacy considerations as their main reason. These participants expressed concerns over privacy issues, a perceived lack of control over the database, and a desire to limit their visibility to other users. For this group, adopting an alias emerged as a strategy to preserve their privacy, deter easy tracking or identification, and hinder potential DNA matches from discovering their identity, thus maintaining anonymity within the platform.

Post-intervention, a shift in preferences became evident. In Group 1, the use of real names dropped to 33.3%, whereas in Group 2, it fell slightly less to 52.4%. This shift shows that the intervention made a substantial difference in Group 1, heightening their privacy concerns and altering their behaviors. In contrast, the Control Group, which did not receive any intervention, saw a slight reduction to 52%. By performing the Kruskal-Wallis test, we found no statistically significant difference among groups ($p = .563$). By performing Cochran's Q test, we found significant difference ($p = .003$) within Group 1 pre-intervention and postintervention interest in favor of using an alias in GEDmatch.

Despite a substantial number of participants still preferring to use their real names, the intervention had a noticeable impact on the preference for using an alias to en-

hance privacy and maintain anonymity.

Using regular email address: A majority of participants in all three groups - Group 1, Group 2, and the Control Group - initially displayed a preference for using their regular email addresses for GEDmatch registration. The main reason were included convenience, direct communication, and simplicity of managing a single email account. Prior to the intervention, 71.4% in Group 1, 85.7% in Group 2, and 57.1% in the Control Group echoed these sentiments. By performing the Kruskal-Wallis test, we found no statistically significant difference among groups ($p = .127$).

However, a minority of participants in each group opted for alternative or dedicated email addresses, aiming to separate their personal information from other online activities to minimize risks such as hacking or unnecessary exposure. This group pointed out the importance of anonymity and caution, fearing potential identification by other users via their email addresses.

Post-intervention, apparent variations between the groups became evident. In Group 1, the percentage of participants favoring their regular email addresses decreased to 42.8%, and in Group 2, the percentage reduced even further to 33.3%. These changes indicate the tangible impact of the intervention on Groups 1 and 2, heightening their privacy consciousness and prompting a shift in behaviors. By performing the Kruskal-Wallis test, we found no statistically significant difference among groups ($p = .301$). By performing Cochran's Q test, we found significant difference ($p = .034$) within Group 1, Group 2 ($p < 0.001$) pre-intervention and postintervention.

Conversely, the Control Group, which received no intervention, maintained a relatively stable score, with 57.1% still willing to use their regular email addresses post-intervention. This consistency highlights the efficacy of the intervention in influencing privacy-conscious behavior in Groups 1 and 2.

5.4.8 Privacy settings: GEDmatch

In this section, we will discuss the privacy setting options chosen by participants in pre-intervention or navigation stage and post-intervention or navigation stage. Table 5.5 presents the results for pre and post intervention or navigation stage.

Putting own name as donor: Among the participants from Group 1, Group 2, and the Control Group, 80.9%, 76.1%, and 71.4% respectively, demonstrated a strong preference for choosing their own name as the donor of their DNA data. This inclination was primarily driven by the belief that it was their sample and using their real names would ensure accuracy. Participants also emphasized that since their names were already associated with their accounts, they saw no reason to hide their identities. A shared perspective was that using their real names was ethically correct since the DNA belonged to them. By performing the Kruskal-Wallis test, we found no statistically significant difference among groups ($p = .772$).

However, the discussions also unveiled significant privacy concerns and potential risks associated with uploading raw DNA data. These concerns were prevalent across all groups, despite the majority leaning towards using their real names. Participants underscored the importance of investigating the company's reputation and retaining control over their personal information. Among the voiced concerns were fears of their DNA and personal data being linked on the internet, potential information leaks, privacy infringements, and the risk of being tracked easily. Group 2 participants expressed additional concerns about the lack of control over the company's database and the necessity to scrutinize privacy regulations thoroughly.

Postintervention, however, a clear shift was observed in Group 1 and Group 2. In Group 1, the percentage of participants using their real names dropped to 42.8%, while in Group 2, it further decreased to 33.3%. In contrast, the Control Group, without any intervention, showed a relative stability with 71.4% still choosing to use own name as donor, thus underlining the effectiveness of the intervention in

Table 5.5: Privacy setting options chosen by participants

Privacy settings Preferences				
	Pre-Intervention or navigation	Post-Intervention or navigation	Difference in %	Pre Vs Post-Intervention/ navigation Cochrans' Q Test p-value
Putting own name as donor				
Group 1 (G1) (n =21)	80.9%	42.8%	-23.8%	p=.005
Group 2 (G2) (n =21)	76.1%	33.3%	-28.6%	p=.003
Control Group (CG) (n =21)	71.4%	71.4%	+14.3%	Not significant
Kruskal-Wallis Test p-value	Not significant	G2 vs CG p =.006 G1 vs CG p = .032		
Preference for Opt in option				
Group 1 (G1) (n =21)	52.4%	33.3%	-19.1%	p =.025
Group 2 (G2) (n =21)	52.4%	33.3%	-19.1%	p=.025
Control Group (CG) (n =21)	42.9%	33.3%	-9.6%	Not significant
Kruskal-Wallis Test p-value	Not significant	Not significant		
Data Sharing privacy policy Option - Public				Friedmans' Test p-value
Group 1 (G1) (n =21)	23.8%	9.5%	-14.3%	p=.020
Group 2 (G2) (n =21)	19%	4.8%	-15.8%	p=.014
Control Group (CG) (n =21)	42.9%	42.9%	0%	Not significant
Kruskal-Wallis Test p-value	Not significant	Not significant		

modulating behavior in Groups 1 and 2. By performing the Kruskal-Wallis test, we found statistically significant difference between Group 2 vs control group ($p = .006$) and Group 1 vs Control group ($p = .032$). By performing Cochrans' Q test, we found significant difference ($p=.005$) within Group1, Group 2 ($p = 0.003$) preintervention and postintervention.

Choosing Opt In/Opt Out option: Preintervention or navigation within the participant groups, 52.4% from Group 1 and Group 2, and 42.9% from the Control Group wanted to opted in, citing various reasons. These included a desire to connect with others, to contribute to law enforcement investigations, and to support research.

Confidence in their innocence, iterating they are not criminals, the potential beneficial use of their DNA, the pursuit to uncover connections, increase their chances of being found, and maximizing the GEDmatch experience were additional reasons they mentioned. By performing the Kruskal-Wallis test, we found no statistically significant difference among groups ($p=.646$).

Postintervention, a notable change was observed. The percentage of participants who chose to opt in decreased to 33.3% in Group 1, Group 2, and the Control Group. The rest of the participants across all groups opted out, reporting distrust in law enforcement and a strong inclination to maintain the privacy of their DNA data. They spoke about possible data misuse, hackings, fraud, errors in law enforcement, and being identified and scrutinized for criminal investigation. By performing Cochran's Q test, we found significant difference ($p=.025$) within Group 1 and Group 2 pre-intervention and postintervention.

The percentage changes signify that the intervention amplified participants' concerns, with some speculating that if they opted in, insurance companies could more readily access their data than if they opted out.

Interestingly, the Control Group, which did not receive any intervention, also displayed a decline in opting in. This could suggest that reading the policy and digging deeper into the terms and conditions can have an impact on people's privacy perception and can increase their privacy-conscious decisions.

Choosing privacy option: When it came to privacy preferences while uploading DNA data, participants across the groups displayed varied choices. In the pre-intervention or navigation for the research option, 47.6% of Group 1, 47.6% of Group 2, and 28.6% of the Control Group chose this. These participants talked about the value of contributing to research and supporting scientific progress. They expressed no reservations about sharing their data for research purposes and demonstrated trust in its appropriate handling. The private option appealed to 28.6% in Group 1, 33.3%

in Group 2, and 28.6% in the Control Group. These participants had concerns about their data becoming public and desired to retain control over their personal information. Their primary considerations were privacy, limited access, and the ability to isolate their data from public view. Regarding the public option, it was preferred by 23.8% of participants in Group 1, 19% in Group 2, and 42.9% in the Control Group. These participants aimed to identify DNA relatives, maximize their chances of matching with others, and broaden their family connections. By performing the Kruskal-Wallis test, we found no statistically significant difference among groups.

While similarities existed in participants' reasons for their privacy preferences, notable differences were also present. For example, Group 1 and Group 2 participants stressed the significance of contributing to research and connecting with family members more than those in the Control Group. In contrast, the Control Group demonstrated a more balanced distribution of preferences, with some emphasizing research, others valuing privacy, and a substantial number favoring public connections.

Post-intervention, there was a discernible shift in preferences. For the research option, the percentage of participants choosing it dropped to 28.6% in Group 1, but slightly increased to 33.3% in Group 2. For the public option, a sharp decline was observed, with only 9.5% in Group 1 and 4.8% in Group 2 opting for it. The Control Group, which had not received any intervention, maintained relatively stable percentages for the research option. This shift emphasizes the intervention's effectiveness in making Groups 1 and 2 more privacy-conscious. By performing Friedmans' test, we found significant difference ($p=.020$) within Group 1 for choosing public option and Group 2 ($p=.014$) pre-intervention and post-intervention.

5.4.9 Understanding of potential access to the DNA data

In a study examining participants' perceptions of data accessibility in public genealogy databases by law enforcement, third parties, insurance companies, and potential breaches, distinctions emerged within and between Group 1, Group 2, and the Control

Table 5.6: Understanding of potential access to the DNA data

Understanding of potential access by different entities				
	Pre-Intervention or navigation	Post-Intervention or navigation	Difference in %	Pre Vs Post-Intervention /navigation Cochrans' Q Test p-value
Understanding of potential law enforcement access				
Group 1 (G1) (n =21)	42.8%	90.5%	+47.7%	p=.005
Group 2 (G2) (n =21)	47.6%	95.2%	+47.6%	p=.003
Control Group (CG) (n =21)	57.1%	85.7%	+28.6%	p=.02
Kruskal-Wallis Test p-value	Not significant	Not significant		
Understanding of potential access by Insurance				
Group 1 (G1) (n =21)	23.8%	81%	+57.2%	p <.001
Group 2 (G2) (n =21)	23.8%	81%	+57.2%	p <.001
Control Group (CG) (n =21)	28.5%	45%	+16.5%	Not significant
Kruskal-Wallis Test p-value	Not significant	G2 vs CG p = .002 G1 vs CG p = .002		
Understanding of potential data breaches				
Group 1 (G1) (n =21)	66.6%	90.5%	+23.9%	p =.025
Group 2 (G2) (n =21)	66.6%	100%	+33.4%	p =.014
Control Group (CG) (n =21)	76.1%	85.7%	+9.6%	Not significant
Kruskal-Wallis Test p-value	Not significant	Not significant		
Understanding of potential access by third parties				
Group 1 (G1) (n =21)	28.5%	90.5%	+62%	p <.001
Group 2 (G2) (n =21)	33.3%	85.7%	+52.4%	p <.001
Control Group (CG) (n =21)	28.5%	42.8%	+14.3	Not significant
Kruskal-Wallis Test p-value	Not significant	G2 vs CG p = .001 G1 vs CG p = .001		

group.

Pre-intervention, a larger percentage of the Control group (57.1%) believed in the accessibility of their data to law enforcement, followed by Group 2 (47.6%), with Group 1 (42.8%) having the least belief. For third-party access, Group 2 (33.33%) had the most belief, then the Control group (28.5%) and Group 1 (28.5%). When considering insurance companies' access, the Control group (28.5%) had the most belief, followed by Group 1 and Group 2 (both at 23.8%). Regarding data breaches, the Control group (76.1%) was most concerned, followed by Group 1 and Group 2 (both at 66.66%). By performing the Kruskal-Wallis test, we found no statistically

significant difference among groups.

Post-intervention, perceptions altered notably in Groups 1 and 2, while the Control group remained relatively stable due to the lack of intervention. There was a marked rise in belief in law enforcement's access to data in Group 1 (90.5%) and Group 2 (95.2%), with a less pronounced increase in the Control group (85.7%). The belief in third-party access saw a surge in Groups 1 (90.5%) and Group 2 (85.7%), while the Control group's belief (42.8%) remained comparatively stable. Participants across all groups consistently expressed concerns about potential data breaches, with Group 2 now showing a maximum belief of 100%, followed by Group 1 (90.5%) and the Control group (85.7%).

Notably, belief in insurance companies' access to data soared in Group 1 and Group 2 (both at 81%) post-intervention, whereas the Control group saw a moderate rise to 45%. The considerable change in Groups 1 and 2 underscores the intervention's influence on participants' perceptions. By performing the Kruskal-Wallis test, we found statistically significant difference between Group 2 vs control group ($p = .001$) and Group 1 vs Control group ($p = .001$) when we asked: Do you think third parties can have access to your data in public genealogy databases? Similarly, by performing the Kruskal-Wallis test, we found statistically significant difference between Group 2 vs control group ($p = .002$) and Group 1 vs Control group ($p = .002$) when we asked: Do you think insurance can have access to your data in public genealogy databases? Applying the Cochran's Q test, we detected a significant variation ($p = .005$) within Group 1 and Group 2 ($p = .003$) after post-intervention, along with the control group ($p = 0.002$) after textual policy explanation, concerning the comprehension of possible law enforcement access after navigation. Additionally, through the execution of the Cochran's Q test, we uncovered a significant change ($p < 0.001$) within Group 1 and Group 2 ($p < 0.001$) after the intervention, in relation to understanding potential insurance access. In the case of potential third-party access, the Cochran's Q test

revealed a significant alteration ($p < 0.001$) within Group 1 and Group 2 ($p < 0.001$) in post-intervention phases. Lastly, when considering potential data breaches, the Cochran's Q test identified a notable variation ($p = .025$) within Group 1 and Group 2 ($p = 0.014$), during post-intervention stages. Table 5.6 presents the results for pre and post intervention or navigation stage.

The Control group's modest increase in belief percentages may stem from enhanced understanding of the platform and its policies. For instance, the uptick in belief in insurance company access may reflect the lack of explicit restrictions in the policy. The increased belief in law enforcement access could be attributed to clarification of the opt-in and opt-out policy, stating the potential use of DNA in law enforcement cases upon opting in.

Overall, this change demonstrates the interventions' capacity to shift beliefs regarding data accessibility in public genealogy databases, emphasizing the importance of informed decision-making in data sharing. Meanwhile, the Control group's relative stability highlights the role of explicit platform policies and user understanding in shaping perceptions.

5.4.10 Willingness to share DNA data

In analyzing responses on a 4-point Likert scale that gauged attitudes towards sharing DNA data with diverse entities such as law enforcement, research organizations, insurance companies, the government, and third parties, differences emerged between Group 1, Group 2, and the Control group both within and among these groups. The scale ranged from "Extremely unlikely" to "Extremely likely."

Prior to intervention, Group 1 showed a high mean score of 3.04 (median = 3), signaling a notable propensity to share their DNA data with law enforcement. They also displayed a substantial willingness to partake in research, reflected by a mean score of 3.6 (median = 4). On the other hand, a lower mean score of 1.38 (median = 1) for sharing DNA data with insurance companies suggested hesitancy in this area.

Group 1 members also displayed a similar level of inclination (3.04) (median = 3) for sharing DNA data with the government as with law enforcement, while sharing data with third parties showed moderate interest with a mean score of 2.04 (median = 2).

Post-intervention, a decline in mean scores across all entities except for research indicated that the intervention influenced the attitudes of Group 1 towards sharing their DNA data. The mean score for sharing data with law enforcement, the government, insurance companies and third parties fell to 1.95 (median = 2), 1.95 (median = 2), 1.14 (median = 1) and 1.14 (median = 1), respectively, illustrating decreased willingness. Despite a marginal dip to 3.0 (median = 3), the persistence of a high score for research purposes suggested that participants' willingness in this area was less affected by the intervention.

Before the intervention, Group 2 demonstrated a moderate willingness to share DNA data with law enforcement (mean = 2.66) (median = 3), a high inclination towards research (mean = 3.38) (median = 4), and a low propensity to share data with insurance companies (mean = 1.38) (median = 1), the government (2.57) (median = 3), and third parties (1.66) (median = 2). Post-intervention, the mean scores for sharing data with law enforcement, the government, insurance companies, and third parties dropped to 1.95 (median = 2), 1.2 (median = 1), 1.0 (median = 1), and 1.04 (median = 1), respectively, indicating decreased willingness. However, the score for sharing data for research purposes marginally increased to 3.42 (median = 4), signifying a robust willingness in this domain even after the intervention. Table 5.7 presents the mean scores, median and changes in mean score and median score pre and post intervention or navigation stage.

Table 5.7: Willingness to share DNA data

Willingness to share DNA data different entities				
	Pre-intervention or navigation	Post-intervention or navigation	Difference in Mean Score	Change in Median
With Law Enforcement - Mean Score (Median)				
Group 1 (n =21)	3.04 (3)	1.95 (2)	-1.09	3 to 2
Group 2 (n =21)	2.66 (3)	1.95 (2)	-0.71	3 to 2
Control Group (n =21)	3.0 (3)	2.7 (3)	-0.3	No change
With Research Organization - Mean Score (Median)				
Group 1 (n =21)	3.6 (4)	3.0 (3)	-0.6	4 to 3
Group 2 (n =21)	3.38 (4)	3.42 (4)	+0.04	No change
Control Group (n =21)	3.28 (4)	3.38 (4)	+0.1	No change
With Insurance Companies - Mean Score (Median)				
Group 1 (n =21)	1.38 (1)	1.14 (1)	-0.24	No change
Group 2 (n =21)	1.38 (1)	1.0 (1)	-0.38	No change
Control Group (n =21)	1.47 (1)	1.47 (1)	0	No change
With Third Parties - Mean Score (Median)				
Group 1 (n =21)	2.04 (2)	1.14 (1)	-0.9	2 to 1
Group 2 (n =21)	1.66 (2)	1.04 (1)	-0.62	2 to 1
Control Group (n =21)	1.66 (1)	1.47 (1)	-0.19	No change
With Government - Mean Score (Median)				
Group 1 (n =21)	3.04 (3)	1.95 (2)	-1.09	3 to 2
Group 2 (n =21)	2.57 (3)	1.2 (1)	-1.37	3 to 1
Control Group (n =21)	3.0 (3)	3.0 (3)	0	No change
<i>Likert Scale Scores</i>				
<i>1 - Extremely Unlikely, 2 - Somewhat Unlikely, 3 - Somewhat Likely, 4 - Extremely Likely</i>				

Comparatively, the Control group, which did not receive any interventions, showed relatively stable mean scores both pre and post the evaluation period. Their scores ranged around the moderate level of 3.0 (median = 3) for sharing data with law enforcement and the government, a high level of 3.28 (median = 4) for research, and a low level of 1.47 (median = 1) and 1.66 (median = 1) for sharing data with insurance companies and third parties, respectively. After the navigation and textual explanation, the Control group's scores showed minor fluctuations.

The evidence suggests that the interventions significantly impacted the attitudes of Group 1 and Group 2 towards sharing their DNA data, particularly with entities such as law enforcement, the government, and third parties, and less so for research purposes. The Control group exhibited minimal shifts in their attitudes, emphasizing that interventions can be critical in shaping perspectives on data sharing, privacy, and access.

5.4.11 Understanding the issue of Informed consent

Before the intervention, the understanding of participants across all groups regarding the interconnected nature of DNA data within a family context was notably limited. None of them talked about how their sharing of DNA data could impact other genetic relatives or genetic family members. The participants seemed to lack comprehension of the intricate relationship that exists between their own DNA data and those of their relatives.

Following the intervention, a substantial shift was observed in the participants' comprehension. A notable proportion of individuals across both Group 1 (80.95%) and Group 2 (95.23%) started discussing the relational aspect of DNA data. They were able to articulate how their individual decisions regarding DNA data could implicate the privacy of not only their existing family members but also their ancestors and future generations. This newfound awareness demonstrates a profound understanding of the shared nature and privacy implications of DNA data within a familial context.

In contrast, there was no change in perception and understanding within the control group. The participants of this group demonstrated a consistent stance in their opinions and comprehension, both pre- and post-navigation. Their perception remained static, thus confirming the impact of the intervention or risk communication on Group 1 and Group 2 participants understanding of the subject matter.

This contradiction undoubtedly indicates the efficacy of the intervention in improving the comprehension of participants in Group 1 and Group 2 regarding the interconnected nature of DNA data. The intervention, as evidenced by the results, appears to have played a pivotal role in enabling participants to grasp the multifaceted implications of their decisions related to DNA data on the privacy of their relatives - both current, past, and future.

5.5 Limitations

While this study has delivered valuable findings, it is not devoid of constraints. A significant limitation lies in the size and demographics of our sample. We were only capable of recruiting 21 individuals per group, which could have impacted our results and hindered the breadth of their applicability. Further, the population involved in the research was predominantly composed of students and university affiliates, with a total participant count that was relatively small. This severely curtails the ability to generalize the findings to larger, more diverse demographics. Our sample consisted mainly of young, well-educated individuals who are likely more tech-savvy and privacy-aware, possibly influencing their understanding and response to risk communication. Additionally, the restricted sample size hampered our ability to perform a robust statistical analysis. This could partly account for the observed lack of substantial differences in the perceptions and behavioral tendencies between Group 1 and Group 2 participants. Despite these limitations, the research delivers critical insights into the efficacy of video risk communication, with Groups 1 and 2 showing heightened privacy consciousness following the intervention. The control group, on the other hand, demonstrated stable decisions, perceptions, and risk-benefit evaluations. Moving forward, our aim is to engage a larger and more diverse participant base to enhance our comprehension of video risk communication's effectiveness, and its temporal and spatial impact on different audiences. We also plan to employ statistical analysis to gauge its effectiveness and expand the generalizability of our findings.

5.6 Discussion

Our study aimed to explore the influence of targeted risk communication on individuals' willingness to undergo at-home DNA testing and share their DNA data. The results highlighted that initial responses without specific risk and benefit information showed significant interest in at-home DNA testing. This suggests a knowledge gap

among participants and a general mistrust in at-home DNA testing, emphasizing the importance of effective risk communication. Following the intervention, we noticed a decline in interest in at-home DNA testing among Groups 1 and 2. This change illustrates how awareness of potential risks can significantly influence decision-making processes. Meanwhile, the control group's interest levels remained relatively constant, further highlighting that perceived benefits can maintain interest without information on risks. Post navigation, there is an increase in concerns regarding at-home DNA testing in the Control Group. This observation suggests that even exposure to policy information can lead to heightened apprehension about potential risks.

Our findings also pointed to a significant decrease in participants' willingness to share their DNA data in public genealogy databases post-intervention in both Group 1 and Group 2 once they learned about how their DNA sharing decisions could impact their genetic relatives and there have been hacking instances in recent past leading to disclose of sensitive genetic data online. The primary driver of this reluctance was data privacy concerns. This trend was observed even in the Control Group, with fears of potential hacking and access by law enforcement agencies leading to decreased data-sharing willingness suggesting understanding the platform and policies by spending extra/conscious time on it could also impact people's decision-making and can lead to attitude changes. An interesting pattern emerged concerning the use of real names and regular email addresses during GEDmatch registration. Initially, participants in all groups displayed a high inclination to use real names during the registration process. Post-intervention, however, we observed a clear shift in behavior among participants in Groups 1 and 2, indicating heightened privacy consciousness due to the intervention. This trend was also observed in the preference for using regular email addresses, with a decline in Groups 1 and 2 and a stable preference in the Control Group post-intervention. Furthermore, when it came to privacy settings, we noted a significant shift in behavior among participants in Groups 1 and 2 post-intervention.

There was a clear preference shift from using real names to aliases for the DNA data donors, emphasizing the effectiveness of targeted risk communication.

A notable decrease in the preference for the opt-in option was observed post-intervention across all groups. This suggests that targeted risk communication amplified participants' concerns about their DNA data privacy and influenced their decisions to maintain this privacy. Similarly, we observed a clear shift in privacy preferences while uploading DNA data post-intervention. There was a decline in choosing public and research options in Groups 1 and 2, indicating that our intervention successfully made these groups more privacy-conscious.

We did not find any significant difference between the Group 1 and Group 2 which indicates that the time of risk communication did not make a significant impact of people's perceptions and informing them about the risks and benefits of DNA data sharing.

This underscores the need for providing balanced information about the potential benefits and risks of at-home DNA testing and data sharing. Clear explanations of privacy policies and data protection measures are integral to this process, as our findings suggest that understanding these elements can significantly influence individuals' attitudes.

5.7 Conclusion

Our study's findings shed light on the crucial role of targeted risk communication in shaping public attitudes and behaviors towards at-home DNA testing and data sharing. After introducing comprehensive and transparent information regarding potential risks and benefits, we observed a significant shift in participants' decision-making. This highlights the importance of clear, concise, and accessible information in guiding individuals to make informed decisions about their DNA data.

Our study has critical implications for a variety of stakeholders, including those engaged in providing DNA testing services, privacy policymakers, and public health

communicators. For those offering DNA testing services, these results can guide efforts to increase transparency and clarity in conveying information about potential risks and benefits. For privacy policymakers, this study highlights the importance of clear and comprehensive policies to protect individuals' genetic data. For public health communicators, these findings underscore the importance of effective risk communication in fostering informed decision-making among the public.

In conclusion, as at-home DNA testing and data sharing continue to gain popularity, it is of paramount importance to engage in effective risk communication and education efforts. This approach can help ensure public understanding, foster informed decision-making, and promote a balance between the potential benefits of DNA testing and data sharing and the necessary safeguarding of individuals' privacy.

CHAPTER 6: Discussion and Conclusion

The industry of Direct-to-Consumer (DTC) genetic testing has seen a significant growth, primarily due to a general lack of understanding among the general populace regarding the potential hazards associated with sharing their genetic data. Many individuals resort to these tests out of sheer curiosity or feel empowered by the wealth of personal information they can gain access to, such as knowledge about their ancestry or predisposition to certain health conditions. However, the realm of commercial DNA testing exists within a largely unregulated space. Companies operating in the DTC genetic testing domain are not bound by the Health Insurance Portability and Accountability Act (HIPAA), which is the primary law safeguarding health information privacy. Without comprehensive regulation and vigilant oversight of the industry, it has come to light through research that the privacy policies of key players in the industry, such as 23andMe and Ancestry, fail to adequately inform consumers about the risks associated with sharing their genetic information with DTC genetic testing firms.

These risks could potentially include receiving erroneous or undesirable health information reports, susceptibility to data breaches, and the possibility of misuse of data. Furthermore, there are no significant regulations in place that dictate the access law enforcement agencies have to genetic samples. The burgeoning of DTC genetic testing companies has paved the way for third-party entities like GEDmatch, which provide services to interpret consumer genetic data. These services, which can range from matching users to genetic relatives to providing customized diet and fitness plans, to delivering health risk assessments, remain largely unregulated and, thus, pose considerable threats to privacy and security. The lack of regulation sur-

rounding DTC tests is a matter of serious concern, with the potential to infringe on individuals' right to privacy. DTC genetic testing can equip individuals with insights into their health risks, information which could potentially be used by insurance companies to deny coverage or to impose higher premiums. This possibility creates the potential for insurance discrimination, which could deter people from participating in medical research.

Of utmost importance is the fact that when an individual decides to undertake a DNA test, the repercussions could also affect all other genetic relatives of that person. In effect, the decision to test and share one's DNA is also a decision made on behalf of their genetic family members, which includes parents, siblings, offspring, and even future generations. This factor is a critical consideration consumers need to account for when sharing their DNA data. If individuals were equipped with a better understanding and awareness of the risks and benefits of DNA data sharing, they could make more informed decisions. To formulate an effective communication strategy, it is critical to gain an understanding of users' perspective on the security and privacy aspects of at-home DNA testing.

To further delve into these complexities, my studies looked into the following:

- How do users understand and interpret the inter-connected nature of DNA data? What are the motivations, perceived benefits, and risks of undertaking an at-home DNA test?
- Are there any differences in perception and attitude between users who have already taken a DNA test vs. those who have not?
- How do users perceive the act of sharing DNA data online? Is there any difference in perception between sharing DNA data with testing companies vs. open databases?
- Do users have a good understanding of the current policies, rules, or laws of their

respective testing companies that have shared their DNA and the existing laws of the USA? What kind of regulations and laws for online DNA data sharing do they prefer?

- What is the most effective way to communicate the risks and benefits of DNA data sharing? How does the intention to share data change once users are informed about both the risks and benefits?

In our comprehensive exploration of user perceptions about at-home DNA testing and sharing data in public genealogy databases, our aim was to understand the rising trend of DNA data sharing on these platforms. In our study, we collected user views about commercial DNA tests, their expected benefits, concerns, and perceptions about third-party tools or public genealogy databases like GEDmatch. The research revealed a significant lack of understanding about the sensitive nature of DNA data among the participants. It was found that users often underestimated the extent of DNA information they were sharing when they took such tests, thinking that they were only sharing a fraction of their DNA. Moreover, they were largely unaware of the interconnectedness of DNA among family members, which made them vulnerable to being identified without their consent due to a relative's decision to share DNA data.

This lack of awareness and misunderstanding was particularly prevalent among participants who had no prior experience with these tests. However, even those who had taken these tests before showed a certain level of complacency about the potential risks, which may be a result of a sense of resignation about their data that had already been shared. Their prior experiences seemed to shape their perception of privacy, leading to what can be termed as a "low expectation of privacy." Regardless of this, there was a shared sense of vulnerability across all groups and a desire for more transparency and control. Participants were particularly concerned about the potential misuse of their data by entities such as law enforcement agencies or insurance

companies, despite recognizing the benefits of these tests, such as gaining health-related information or finding family members.

With these insights in mind, in our second study our goal was to design an effective risk communication method to help users make more informed decisions about DNA testing and data sharing. We initiated a multi-phase participatory design study, involving participants in the design process to ensure that the results cater to the needs and expectations of the stakeholders. The aim was to understand and gather requirements for designing effective risk communication messages and to involve users in creating these methods. In the first phase of this study, we collected the needs of the participants and collaborated with users to create initial five designs or methods to communicate the risks of DNA data sharing to the users. The second phase of the study mainly focused on refining these designs with user feedbacks and finalizing the best two methods. The third phase of our research was focused on assessing the effectiveness of two different forms of risk communication, infographic and video in shaping participants' understanding, recall, and intention to share DNA data in relation to at-home DNA tests and participation in open genealogy databases. Although, Our analysis found no significant differences between the two groups in terms of easy-to-understand and sharing intention criteria, but the video format generally outperformed the infographic in terms of content recall. Additionally, it's important to note that a larger proportion of participants who viewed the video message expressed reluctance to undergo the test or share their DNA data, compared to those who viewed the infographic message.

In our subsequent study, we aimed to explore the influence of targeted risk communication on individuals' willingness to undergo at-home DNA testing and share their DNA data. The results highlighted that initial responses without specific risk and benefit information showed significant interest in at-home DNA testing. After the intervention, we noticed a decline in interest in at-home DNA testing among the

treatment groups. This change illustrates how awareness of potential risks can significantly influence the decision-making process. Meanwhile, the control group's interest levels remained relatively constant, further highlighting that perceived benefits can maintain interest even without information on risks. After navigating through the information, there is an increase in concerns regarding at-home DNA testing in the control group. This observation suggests that even exposure to policy information can lead to heightened apprehension about potential risks. We also found that a significant decrease in participants' willingness to share their DNA data in public genealogy databases after the intervention in both treatment groups, once they learned about how their DNA sharing decisions could impact their genetic relatives and that there have been instances of hacking that led to the disclosure of sensitive genetic data online. The primary driver of this reluctance was data privacy concerns. This trend was observed even in the control group, where fears of potential hacking and access by law enforcement agencies led to a decrease in the willingness to share data, suggesting that spending extra time understanding the platform and policies could also impact people's decision-making and lead to changes in attitude.

An interesting pattern emerged concerning the use of real names and regular email addresses during GEDmatch registration. Initially, participants in all groups displayed a high inclination to use real names during the registration process. However, after the intervention, we observed a clear shift in behavior among participants in two treatment groups, indicating heightened privacy consciousness due to the intervention. This trend was also observed when we asked them about their willingness to share their personal information, such as the preference for using regular email addresses, with a decline in two treatment groups and a stable preference in the control group after the intervention. Furthermore, when it came to privacy settings on GEDmatch, we noted a significant shift in behavior among participants in two treatment groups after the intervention. There was a clear preference shift from using real

names to aliases for DNA data donors, clearing choosing more privacy-preserving options emphasizing the effectiveness of targeted risk communication.

Similar pattern was seen in the preference for the opt-in option was observed after the intervention across all groups. This suggests that targeted risk communication amplified participants' concerns about their DNA data privacy and influenced their decisions to maintain this privacy. Similarly, we observed a clear shift in privacy preferences while uploading DNA data after the intervention. There was a decline in choosing public and research options in two groups, indicating that our intervention successfully made these groups more privacy-conscious. This underscores the need for providing balanced information about the potential benefits and risks of at-home DNA testing and data sharing. Clear explanations of privacy policies and data protection measures are integral to this process, as our findings suggest that understanding these elements can significantly influence individuals' attitudes.

6.1 Contributions

In this section, I provide an overarching description of my contributions to the literature pertaining to DNA data privacy and risk communication.

- **In-Depth Understanding of Users' Perceptions Towards At-Home DNA Testing and Genomic Data Sharing:** In our initial study, we discovered that the vast majority of participants were not conscious of the familial interconnect- edness of DNA. The concept that one's DNA can reveal information about one's relatives was a surprising revelation for most. This ignorance is consistent with prior studies [11, 28], where users exhibited limited understanding of the poten- tial risks associated with DNA data sharing. While some privacy concerns were identified, these apprehensions were not sufficient to deter participants from un- dergoing DNA testing. Further investigation, as conducted by Baig et al. [28], revealed that users, particularly those who opted for at-home DNA testing, gen- erally possessed only basic knowledge about DNA data. To build upon these

findings, our research delved deeper into the perceptions of both adopters and non-adopters of at-home DNA testing. our study exposed a clear knowledge gap regarding the perceptions and attitudes of these users towards different entities accessing their data. We found a pervasive sense of unease among both groups. Non-adopters felt vulnerable, recognizing that their privacy could be compromised by others' decisions. On the other hand, those who had taken the tests exhibited an attitude of resignation towards data privacy, often expressing regret for their prior decisions. The insights gained from our study are of significant value to lawmakers, direct-to-consumer companies (DTCs), public genealogy databases, and researcher for DNA privacy policy development.

- **Effective ways of risk communication:** Risk communication refers to the dissemination of information regarding potential hazards and the uncertainties associated with them to individuals, communities, or groups who may be impacted. In our research, we examined effective methods for communicating the risks inherent in sharing DNA data online. In line with prior research [45, 46], our study confirmed the constructive influence of storytelling on learning and fostering secure behavior. We found that narratives containing instructive elements directly affected participants' actions, while stories outlining significant threats shaped their thinking. We also expanded upon existing findings [51, 52, 55] by exploring the efficacy of visually communicating risks. Visual representation of data has been shown to be effective in uncovering data patterns, enabling comparisons, and grabbing attention. In our study, participants reported that infographics, combining visuals and data, facilitated an easy understanding of the risks associated with DNA data sharing. The combination of pictures and data greatly aided their comprehension and consequently served as an effective medium for communicating the risks associated with sharing DNA data online. In addition, consistent with previous research [81, 63, 64, 98, 99],

our findings suggested that video-based methods can be among the most preferred and effective means of risk communication. Such methods play a critical role in heightening awareness and improving understanding of the risks associated with DNA data sharing. Our research provides an effective DNA sharing risk communication method to help users understand the benefits, risks, and implications of sharing DNA data online, thereby facilitating informed decision-making filling the void of DNA privacy risk communication.

6.2 Design Implications

Design considerations encompass the essential factors and elements that need to be deliberated prior to making substantial decisions regarding the design solution. Typically, they are symbolically represented through carefully selected and clearly interpreted icons. Drawing from my research outlined in Chapter 4, I have extracted the subsequent design implications.

- Video is more effective for conveying the nuanced risks and benefits of DNA data sharing - Videos, with their combination of audio-visual elements, are highly effective mediums for conveying complex information like the risks and benefits associated with DNA data sharing. They can present layered information in an easily digestible manner, using visual aids to break down complex topics. They offer a dynamic way to show real-world scenarios, metaphors, or animations to explain abstract concepts, making it easier for viewers to understand the potential impacts of sharing their DNA data.
- Video design needs to contain natural human voice for more relatability - Incorporating a natural human voice in the video enhances the communication's effectiveness. It humanizes the message, making it more relatable and engaging. A natural voice can convey a range of tones and emotions that text or visuals alone may not fully express. This can help in emphasizing important points,

explaining complex aspects in a simple way, and building a connection with the viewer, ultimately improving comprehension and retention.

- Female voice is more preferred than a male voice - My research has suggested that female voices are often perceived as more comforting, trustworthy, and clear. They can evoke a sense of warmth and empathy, which can make the content more relatable and accessible to viewers. A female voice-over can help in creating an inviting, non-intimidating environment that encourages the viewer to pay attention and understand the message better. However, the choice of voice should always consider the target audience's preferences.
- Balanced contents of risk and benefit in the message - Striking a balance between the risks and benefits in the video is crucial to provide an unbiased perspective. Presenting only the benefits may seem promotional, while focusing solely on risks could instill unnecessary fear. The video should aim to inform viewers about both aspects evenly, allowing them to make an informed decision. This can also establish trust, as it shows the creators' transparency and objectivity about the subject.
- Video should not exceed more than 3 mins to keep audience engaged - With dwindling attention spans, keeping the video concise and to the point is key to ensuring viewer engagement. Videos that are three minutes or shorter can convey the necessary information without overwhelming or losing the viewer's attention. To achieve this, the content should be well-structured, with a clear and engaging narrative that delivers the key points effectively within this time frame.
- Examples with data and statistics are inevitable for risk communication to aid users's understanding - Using examples with data and statistics can greatly enhance the effectiveness of risk communication. These elements can substantiate

the points being made and provide a concrete foundation for understanding the risks involved. Graphical representations of data and statistics can simplify the interpretation of complex information and allow viewers to grasp the magnitude or relevance of the risks and benefits associated with DNA data sharing.

- Risk communication should not use technical language or jargons, should be aimed at an average audience - The use of jargon or highly technical language can alienate viewers, especially those without specialized knowledge in the field. For effective communication, the language should be accessible and easily understood by the average person. This includes simplifying complex concepts, avoiding or clearly defining technical terms, and using everyday language. This approach ensures that the video's message can reach a broader audience, and can aid in the viewers's understanding of the risks and benefits involved.

6.3 Future Work

We have learned a lot through the investigations presented in this thesis, but there is still work that needs to be done to expand our understanding of how different messages affect users' perceptions and the adoption of security and privacy tools. Our recent research study highlights the divergence in perspectives regarding online DNA data sharing among individuals of different races and nationalities. For instance, our findings show that African Americans often feel vulnerable upon realizing that law enforcement agencies can access genealogical databases to aid in crime resolution. This critical insight sets the groundwork for a more in-depth examination of the perceptions held by various racial and national groups about at-home genetic testing. Research [100, 101] has consistently shown that cultural backgrounds and races significantly influence opinions and perceptions of diverse technologies or systems. As such, my research interest lies in exploring these differences in users' perceptions across varying nationalities and racial groups. By collating and summarizing their

viewpoints, I aim to propose design guidelines, policies, rules, and laws that prioritize data privacy and cater to these diverse user experiences.

I plan to delve into key factors like risk communication and the spatial and temporal effects of risk communication. I aim to extend my previous study by involving a broader participant pool. Moreover, I consider our results exploratory and cannot be generalized for all users, since it was limited to a specific population, such as university students and employees or individuals with good technology skills. We believe that the validity of the findings could be improved by targeting less tech-savvy or low-income people from different countries with a larger sample size.

My ultimate goal is to develop a privacy-conscious platform. This platform would allow users to maintain control over their data privacy even after sharing their genetic information on public genealogy databases. The larger objective behind this research is to advance medical science by providing researchers with valuable genetic data while also offering adoptees the possibility to find their biological families. However, such benefits should not come at the expense of user's privacy. Ensuring that DNA privacy isn't compromised once data is shared is essential. Similarly, non-adopters should not feel vulnerable or victimized by others' decisions. Balancing these different aspects is a crucial challenge that my research hopes to address.

6.4 Conclusion

In conclusion, the surge of the Direct-to-Consumer (DTC) genetic testing industry is largely propelled by curiosity and a quest for personal insights. Yet, there is a concerning lack of understanding about potential privacy and security risks associated with sharing genetic data. Our research provides an in-depth examination of this critical issue, revealing a significant knowledge gap in the public's understanding of DNA data sharing and the implications thereof. We observed a substantial lack of knowledge about the interconnections of genetic data and its impact on privacy, with many participants underestimating the extent of DNA data they were sharing.

Equally concerning is the fact that the DTC genetic testing industry operates in a largely unregulated space, exposing users to various risks such as data breaches, misuse of data, and potential insurance discrimination. Our study also underscored the need for effective risk communication methods to help users make informed decisions. We found that the use of video as communication tool not only increased understanding but also influenced the decision-making process. Furthermore, targeted risk communication proved to be crucial in amplifying participants' concerns about their DNA data privacy, resulting in an increase in privacy-conscious behaviors. This research provides valuable insights for lawmakers, DTCs, public genealogy databases, and researchers, paving the way for the design and implementation of robust security and privacy mechanisms. Overall, this study underlines the urgent need for greater transparency, better regulations, and more comprehensive risk communication in the domain of DTC genetic testing and data sharing. These efforts will be crucial in ensuring the protection of individuals' privacy rights while harnessing the benefits of this rapidly evolving field.

REFERENCES

- [1] L. Sahoo, M. Shehab, E. A. Qahtani, and J. Dev, “Nobody wants my stuff and it is just dna data, why should i be worried,” in *Privacy Symposium 2022: Data Protection Law International Convergence and Compliance with Innovative Technologies (DPLICIT)*, pp. 155–178, Springer, 2022.
- [2] “Cnn-gedmatch privacy.” <https://www.cnn.com/2019/05/27/us/genetic-genealogy-gedmatch-privacy/index.html>. Accessed: 2020-03-30.
- [3] “More than 26 million people have taken an at-home ancestry test.” Accessed: 07/25/22.
- [4] S. C. Nelson, D. J. Bowen, and S. M. Fullerton, “Third-party genetic interpretation tools: a mixed-methods study of consumer motivation and behavior,” *The American Journal of Human Genetics*, vol. 105, no. 1, pp. 122–131, 2019.
- [5] J. Cohen, “Gedmatch review: What is gedmatch?.” <https://resources.selfdecode.com/blog/gedmatch-review-what-is-gedmatch/>. Last accessed 28 September 2021.
- [6] N. Moray, K. E. Pink, P. Borry, and M. H. Larmuseau, “Paternity testing under the cloak of recreational genetics,” *European Journal of Human Genetics*, vol. 25, no. 6, pp. 768–770, 2017.
- [7] “Genetic information discrimination.” Accessed: 2020-11-07 02:28:36.
- [8] “U.s. department of health and human resources. u.s. department of health human services. âhipaa security guidance.â.” Accessed: 2020-11-07 02:28:36.
- [9] “Janet dolgin lois shepherd, bioethics and the law, 219 (4th ed. 2019).” Accessed: 2020-11-07 02:28:36.
- [10] “Louisiana state legislature.” Accessed: 2020-11-07 02:28:36.
- [11] D. Saha, A. Chan, B. Stacy, K. Javkar, S. Patkar, and M. L. Mazurek, “User attitudes on direct-to-consumer genetic testing,” in *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, pp. 120–138, IEEE, 2020.
- [12] M. J. Culnan and P. K. Armstrong, “Information privacy concerns, procedural fairness, and impersonal trust: An empirical investigation,” *Organization science*, vol. 10, no. 1, pp. 104–115, 1999.
- [13] T. Dinev and P. Hart, “An extended privacy calculus model for e-commerce transactions,” *Information systems research*, vol. 17, no. 1, pp. 61–80, 2006.

- [14] N. Bol, T. Dienlin, S. Kruikemeier, M. Sax, S. C. Boerman, J. Strycharz, N. Helberger, and C. H. De Vreese, "Understanding the effects of personalization as a privacy calculus: Analyzing self-disclosure across health, news, and commerce contexts," *Journal of Computer-Mediated Communication*, vol. 23, no. 6, pp. 370–388, 2018.
- [15] H. Krasnova, N. F. Veltri, and O. Günther, "Self-disclosure and privacy calculus on social networking sites: The role of culture: Intercultural dynamics of privacy calculus," *Wirtschaftsinformatik*, vol. 54, pp. 123–133, 2012.
- [16] S. Lee, H. R. Ha, J. H. Oh, and N. Park, "The impact of perceived privacy benefit and risk on consumers's desire to use internet of things technology," in *Human Interface and the Management of Information. Information in Applications and Services: 20th International Conference, HIMI 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part II 20*, pp. 609–619, Springer, 2018.
- [17] D. Kim, K. Park, Y. Park, and J.-H. Ahn, "Willingness to provide personal information: Perspective of privacy calculus in iot services," *Computers in Human Behavior*, vol. 92, pp. 273–281, 2019.
- [18] E. Princi and N. C. Krämer, "Acceptance of smart electronic monitoring at work as a result of a privacy calculus decision," in *Informatics*, vol. 6, p. 40, MDPI, 2019.
- [19] S. E. Gollust, E. S. Gordon, C. Zayac, G. Griffin, M. Christman, R. Pyeritz, L. Wawak, and B. A. Bernhardt, "Motivations and perceptions of early adopters of personalized genomics: perspectives from research participants," *Public health genomics*, vol. 15, no. 1, pp. 22–30, 2012.
- [20] A. Childers, *Adoptees' experiences with direct-to-consumer genetic testing: emotions, satisfaction, and motivating factors*. PhD thesis, University of South Carolina, 2017.
- [21] N. M. Baptista, K. D. Christensen, D. A. Carere, S. A. Broadley, J. S. Roberts, and R. C. Green, "Adopting genetics: motivations and outcomes of personal genomic testing in adult adoptees," *Genetics in Medicine*, vol. 18, no. 9, pp. 924–932, 2016.
- [22] E. Vayena, E. Gournay, J. Streuli, E. Hafen, and B. Prainsack, "Experiences of early users of direct-to-consumer genomics in switzerland: an exploratory study," *Public Health Genomics*, vol. 15, no. 6, pp. 352–362, 2012.
- [23] Y. Su, H. C. Howard, and P. Borry, "Users's motivations to purchase direct-to-consumer genome-wide testing: an exploratory study of personal stories," *Journal of Community Genetics*, vol. 2, no. 3, p. 135, 2011.

- [24] M. D. Khan R, "Rumors of the death of consumer genomics are greatly exaggerated | genome biology | full text." <https://doi.org/10.1186/gb4141>. Accessed: 2020-03-30 04:35:04.
- [25] J. M. Bollinger, R. C. Green, and D. Kaufman, "Attitudes about regulation among direct-to-consumer genetic testing customers," *Genetic testing and molecular biomarkers*, vol. 17, no. 5, pp. 424–428, 2013.
- [26] M. Schaper, S. WÃ¶hlke, and S. Schicktanz, "âi would rather have it done by a doctorââlaypeopleâs perceptions of direct-to-consumer genetic testing (dte gt) and its ethical implications," *Medicine, health care, and philosophy*, vol. 22, no. 1, pp. 31–40, 2019. Copyright - Medicine, Health Care and Philosophy is a copyright of Springer, (2018). All Rights Reserved; Last updated - 2019-02-22.
- [27] R. M. Hendricks-Sturup and C. Y. Lu, "Direct-to-consumer genetic testing data privacy: key concerns and recommendations based on consumer perspectives," *Journal of personalized medicine*, vol. 9, no. 2, p. 25, 2019.
- [28] K. Baig, R. Mohamed, A.-L. 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*, pp. 1–13, 2020.
- [29] B. S. Weir, A. D. Anderson, and A. B. Hepler, "Genetic relatedness analysis: modern data and new challenges," *Nature Reviews Genetics*, vol. 7, no. 10, pp. 771–780, 2006.
- [30] M. Marks and T. Li, "Dna donors must demand stronger protection for genetic privacy," *Stat*, 2018.
- [31] J. Marchini and B. Howie, "Genotype imputation for genome-wide association studies," *Nature reviews. Genetics*, vol. 11, p. 499â511, July 2010.
- [32] S. H. Lee, D. Harold, D. R. Nyholt, M. E. Goddard, K. T. Zondervan, J. Williams, G. W. Montgomery, N. R. Wray, and P. M. Visscher, "Estimation and partitioning of polygenic variation captured by common snps for alzheimer's disease, multiple sclerosis and endometriosis," *Human molecular genetics*, vol. 22, no. 4, pp. 832–841, 2013.
- [33] M. Humbert, E. Ayday, J.-P. Hubaux, and A. Telenti, "Addressing the concerns of the lacks family: Quantification of kin genomic privacy," in *Proceedings of the 2013 ACM SIGSAC Conference on Computer Communications Security, CCS â13*, (New York, NY, USA), p. 1141â1152, Association for Computing Machinery, 2013.
- [34] J. Kaiser, "Agency nixes decode's new data-mining plan," *Science*, vol. 340, no. 6139, pp. 1388–1389, 2013.

- [35] M. Edge and G. Coop, “Attacks on genetic privacy via uploads to genealogical databases,” 2019.
- [36] N. A. F. F. H. J. W. J. J. A. W. R. O. A. S. He, D. and . E. Eskin, “Identifying genetic relatives without compromising privacy.” <https://genome.cshlp.org/content/early/2014/03/09/gr.153346.112.abstract>. Accessed: 2020-03-31.
- [37] Ney, “Genotype extraction and false relative attacks security risks to third-party genetic genealogy services beyond identity inference,” 2020.
- [38] D. Nyholt, C.-E. Yu, and P. Visscher, “On jim watson’s apoe status: Genetic information is hard to hide,” *European Journal of Human Genetics*, vol. 17, no. 2, pp. 147–149, 2009.
- [39] G. Bansal, D. Gefen, *et al.*, “The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online,” *Decision support systems*, vol. 49, no. 2, pp. 138–150, 2010.
- [40] P. Dahlstrøm, E. Fauchald, and M. L. Benjamin Fimreite, “Users knowledge and attitudes towards data collection in activity trackers,”
- [41] X. Bellekens, A. Hamilton, P. Seeam, K. Nieradzinska, Q. Franssen, and A. Seeam, “Pervasive ehealth services a security and privacy risk awareness survey,” in *2016 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (CyberSA)*, pp. 1–4, 2016.
- [42] J. Zhao, S. Ha, and R. Widdows, “Building trusting relationships in on-line health communities,” *Cyberpsychology, Behavior, and Social Networking*, vol. 16, no. 9, pp. 650–657, 2013.
- [43] X. Zhang, S. Liu, X. Chen, L. Wang, B. Gao, and Q. Zhu, “Health information privacy concerns, antecedents, and information disclosure intention in online health communities,” *Information & Management*, vol. 55, no. 4, pp. 482–493, 2018.
- [44] Z. Deng and S. Liu, “Understanding consumer health information-seeking behavior from the perspective of the risk perception attitude framework and social support in mobile social media websites,” *International journal of medical informatics*, vol. 105, pp. 98–109, 2017.
- [45] E. Rader, R. Wash, and B. Brooks, “Stories as informal lessons about security,” in *Proceedings of the Eighth Symposium on Usable Privacy and Security*, pp. 1–17, 2012.
- [46] K. Pfeffer, A. Mai, E. Weippl, E. Rader, and K. Krombholz, “Replication: Stories as informal lessons about security,” in *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, pp. 1–18, 2022.

- [47] C. Fennell and R. Wash, “Do stories help people adopt two-factor authentication?,” *Studies*, vol. 1, no. 2, p. 3, 2019.
- [48] K. R. Fulton, R. Gelles, A. McKay, Y. Abdi, R. Roberts, and M. L. Mazurek, “The effect of entertainment media on mental models of computer security,” in *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*, pp. 79–95, 2019.
- [49] A. P. 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, SOUPS ’12*, (New York, NY, USA), Association for Computing Machinery, 2012.
- [50] J. V. Sancho, B. M. Ochoa, and J. C. Domínguez, “Aproximación a una taxonomía de la visualización de datos,” *Revista Latina de Comunicación Social*, no. 69, pp. 486–507, 2014.
- [51] U. Hoffrage, S. Lindsey, R. Hertwig, and G. Gigerenzer, “Communicating statistical information,” 2000.
- [52] I. M. Lipkus and J. G. Hollands, “The visual communication of risk,” *JNCI monographs*, vol. 1999, no. 25, pp. 149–163, 1999.
- [53] B. P. Sarma, N. Li, C. Gates, R. Potharaju, C. Nita-Rotaru, and I. Molloy, “Android permissions: a perspective combining risks and benefits,” in *Proceedings of the 17th ACM symposium on Access Control Models and Technologies*, pp. 13–22, 2012.
- [54] H. R. Lipford, A. Besmer, and J. Watson, “Understanding privacy settings in facebook with an audience view,” *UPSEC*, vol. 8, pp. 1–8, 2008.
- [55] K. Marett, A. L. McNab, and R. B. Harris, “Social networking websites and posting personal information: An evaluation of protection motivation theory,” *AIS Transactions on Human-Computer Interaction*, vol. 3, no. 3, pp. 170–188, 2011.
- [56] R. W. Rogers, “Cognitive and psychological processes in fear appeals and attitude change: A revised theory of protection motivation,” *Social psychophysiology: A sourcebook*, pp. 153–176, 1983.
- [57] A. C. Johnston and M. Warkentin, “Fear appeals and information security behaviors: an empirical study,” *MIS quarterly*, pp. 549–566, 2010.
- [58] K. Witte, “Putting the fear back into fear appeals: The extended parallel process model,” *Communications Monographs*, vol. 59, no. 4, pp. 329–349, 1992.
- [59] R. W. Rogers, “A protection motivation theory of fear appeals and attitude change1,” *The journal of psychology*, vol. 91, no. 1, pp. 93–114, 1975.

- [60] A. Vance, D. Eargle, K. Ouimet, and D. Straub, “Enhancing password security through interactive fear appeals: A web-based field experiment,” in *2013 46th Hawaii International Conference on System Sciences*, pp. 2988–2997, IEEE, 2013.
- [61] S. Boss, D. Galletta, P. B. Lowry, G. D. Moody, and P. Polak, “What do systems users have to fear? using fear appeals to engender threats and fear that motivate protective security behaviors,” *MIS Quarterly (MISQ)*, vol. 39, no. 4, pp. 837–864, 2015.
- [62] J. Jansen and P. Van Schaik, “Persuading end users to act cautiously online: A fear appeals study on phishing,” *Information & Computer Security*, 2018.
- [63] Y. Albayram, M. M. H. Khan, T. Jensen, and N. Nguyen, “â... better to use a lock screen than to worry about saving a few seconds of timeâ: Effect of fear appeal in the context of smartphone locking behavior,” in *Symposium on Usable Privacy and Security (SOUPS)*, 2017.
- [64] E. Al Qahtani, M. Shehab, and A. Aljohani, “The effectiveness of fear appeals in increasing smartphone locking behavior among saudi arabians,” in *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)*, pp. 31–46, 2018.
- [65] A. Baig, “What is Sensitive Data Exposure How to Avoid It? - Securiti — securiti.ai.” <https://securiti.ai/blog/sensitive-data-exposure/>. [Accessed 23-Mar-2023].
- [66] S. Kim, J. Cowley, and M. S. Wogalter, “Emphasis terms for warning directives on compliance intent,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 51, no. 9, pp. 569–573, 2007.
- [67] C. Bravo-Lillo, L. F. Cranor, J. Downs, S. Komanduri, and M. Sleeper, “Improving computer security dialogs,” in *Human-Computer Interaction – INTERACT 2011* (P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, and M. Winckler, eds.), (Berlin, Heidelberg), pp. 18–35, Springer Berlin Heidelberg, 2011.
- [68] S. Egelman, “My profile is my password, verify me! the privacy/convenience tradeoff of facebook connect,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’13, (New York, NY, USA), p. 2369â2378, Association for Computing Machinery, 2013.
- [69] S. Patil, R. Schlegel, A. Kapadia, and A. J. Lee, “Reflection or action? how feedback and control affect location sharing decisions,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 101–110, 2014.
- [70] D. Lindegren, F. Karegar, B. Kane, and J. S. Pettersson, “An evaluation of three designs to engage users when providing their consent on smartphones,” *Behaviour & Information Technology*, pp. 1–17, 2019.

- [71] R. H. Thaler and C. R. Sunstein, *Nudge: Improving decisions about health, wealth, and happiness*. Penguin, 2009.
- [72] E. K. Choe, J. Jung, B. Lee, and K. Fisher, “Nudging people away from privacy-invasive mobile apps through visual framing,” in *IFIP Conference on Human-Computer Interaction*, pp. 74–91, Springer, 2013.
- [73] B. Ur, F. Alfieri, M. Aung, L. Bauer, N. Christin, J. Colnago, L. F. Cranor, H. Dixon, P. Emami Naeini, H. Habib, *et al.*, “Design and evaluation of a data-driven password meter,” in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 3775–3786, 2017.
- [74] Y. Wang, P. G. Leon, K. Scott, X. Chen, A. Acquisti, and L. F. Cranor, “Privacy nudges for social media: An exploratory facebook study,” in *Proceedings of the 22nd International Conference on World Wide Web, WWW ’13 Companion*, (New York, NY, USA), p. 763–770, Association for Computing Machinery, 2013.
- [75] F. Raber, A. De Luca, and M. Graus, “Privacy wedges: Area-based audience selection for social network posts,” in *Twelfth Symposium on Usable Privacy and Security ({SOUPS} 2016)*, 2016.
- [76] S. Coleman, “The minnesota income tax compliance experiment: replication of the social norms experiment,” *Available at SSRN 1393292*, 2007.
- [77] H. Masaki, K. Shibata, S. Hoshino, T. Ishihama, N. Saito, and K. Yatani, “Exploring nudge designs to help adolescent sns users avoid privacy and safety threats,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI ’20, (New York, NY, USA), p. 1–11, Association for Computing Machinery, 2020.
- [78] A. Besmer, J. Watson, and H. R. Lipford, “The impact of social navigation on privacy policy configuration,” in *Proceedings of the Sixth Symposium on Usable Privacy and Security*, pp. 1–10, 2010.
- [79] A. Frik, N. Malkin, M. Harbach, E. Peer, and S. Egelman, “A promise is a promise: The effect of commitment devices on computer security intentions,” in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2019.
- [80] H. Almuhiemedi, F. Schaub, N. Sadeh, I. Adjerid, A. Acquisti, J. Gluck, L. F. Cranor, and Y. Agarwal, “Your location has been shared 5,398 times! a field study on mobile app privacy nudging,” in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, pp. 787–796, 2015.
- [81] J. Abawajy, “User preference of cyber security awareness delivery methods,” *Behaviour & Information Technology*, vol. 33, no. 3, pp. 237–248, 2014.

- [82] F. Mouton, M. M. Malan, L. Leenen, and H. S. Venter, "Social engineering attack framework," in *2014 Information Security for South Africa*, pp. 1–9, IEEE, 2014.
- [83] W. R. Flores and M. Ekstedt, "Shaping intention to resist social engineering through transformational leadership, information security culture and awareness," *computers & security*, vol. 59, pp. 26–44, 2016.
- [84] K. Beckers and S. Pape, "A serious game for eliciting social engineering security requirements," in *2016 IEEE 24th International Requirements Engineering Conference (RE)*, pp. 16–25, IEEE, 2016.
- [85] C. Posey, T. L. Roberts, and P. B. Lowry, "The impact of organizational commitment on insidersâ motivation to protect organizational information assets," *Journal of Management Information Systems*, vol. 32, no. 4, pp. 179–214, 2015.
- [86] Y. Ding, P. Meso, and S. Xu, "Protection motivation driven security learning," 2014.
- [87] C. McCoy and R. T. Fowler, "'you are the key to security': Establishing a successful security awareness program," in *Proceedings of the 32nd Annual ACM SIGUCCS Conference on User Services*, SIGUCCS '04, (New York, NY, USA), p. 346â349, Association for Computing Machinery, 2004.
- [88] M. Siponen, S. Pahlila, and M. A. Mahmood, "Compliance with information security policies: An empirical investigation," *Computer*, vol. 43, no. 2, pp. 64–71, 2010.
- [89] T. Gundu and S. V. Flowerday, "The enemy within: A behavioural intention model and an information security awareness process," in *2012 Information Security for South Africa*, pp. 1–8, IEEE, 2012.
- [90] "Free qualitative data analysis software | qda miner lite." Accessed: 2021-10-12 06:41:14.
- [91] "more-than-26-million-people-have-taken-an-at-home-ancestry-test." Accessed: 2021-09-06 10:01:10.
- [92] Y. Albayram, M. M. H. Khan, and M. Fagan, "A study on designing video tutorials for promoting security features: A case study in the context of two-factor authentication (2fa)," *International Journal of Human-Computer Interaction*, pp. 1–16, 2017.
- [93] M. Harbach, M. Hettig, S. Weber, and M. Smith, "Using personal examples to improve risk communication for security & privacy decisions," in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pp. 2647–2656, ACM, 2014.

- [94] J. J. Francis, M. Johnston, C. Robertson, L. Glidewell, V. Entwistle, M. P. Eccles, and J. M. Grimshaw, "What is an adequate sample size? operationalising data saturation for theory-based interview studies," *Psychology and health*, vol. 25, no. 10, pp. 1229–1245, 2010.
- [95] A. S. for Public Affairs, "System Usability Scale (SUS) | Usability.gov — usability.gov." <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html>. [Accessed 06-Jul-2023].
- [96] J. N. Madhuri, "Use of audio visual aids in teaching and speaking," *Research Journal of English Language and Literature*, vol. 1, no. 3, pp. 108–122, 2013.
- [97] N. Tugut, B. Yesildag Celik, and A. Yilmaz, "Health literacy and its association with health perception in pregnant women," *Journal of Health Literacy*, vol. 6, no. 2, pp. 9–20, 2021.
- [98] E. Al Qahtani, L. Sahoo, Y. Javed, and M. Shehab, "' why would someone hack me out of thousands of students': Video presenter's impact on motivating users to adopt 2fa," in *Proceedings of the 27th ACM on Symposium on Access Control Models and Technologies*, pp. 139–150, 2022.
- [99] E. Al Qahtani, L. Sahoo, and M. Shehab, "The effectiveness of video messaging campaigns to use 2fa," in *International Conference on Human-Computer Interaction*, pp. 369–390, Springer, 2021.
- [100] L. Baruh, E. Secinti, and Z. Cemalcilar, "Online privacy concerns and privacy management: A meta-analytical review," *Journal of Communication*, vol. 67, no. 1, pp. 26–53, 2017.
- [101] D. A. Briley, M. W. Morris, and I. Simonson, "Reasons as carriers of culture: Dynamic versus dispositional models of cultural influence on decision making," *Journal of consumer research*, vol. 27, no. 2, pp. 157–178, 2000.

CHAPTER 7: Appendix

7.1 APPENDIX A: Supplementary data for Chapter 3

7.1.1 Screening Survey

First Name:

AGE:

Major or Field of Profession:

Email ID:

Have you ever done DNA testing? Option 1: Yes Option 2: No

(Note: if the response is Yes) Where did you do it?

- Option 1: 23andMe
- Option 2: MyHeritage
- Option 3: AncestryDNA
- Option 4: OTHERS

Why did you decide to do it?

- Option 1: Family Trees & Ancestry Research
- Option 2: Health information
- Option 3: Personal Identity
- Option 4: Participate in Research
- Option 5: OTHERS

7.1.2 Interview Questions

7.1.2.1 First part: Experienced Group

So In the survey, You mentioned that you did a DNA test,

- Why did you decide to do it?
- What benefits did you perceive from it?
- Do you have any concerns after doing it?
- Have you ever shared your testing result online or with any platforms? (**Note: if the response is Yes**)
 - Why did you choose to share?
 - What benefits did you get from doing it?
 - How are you feeling after doing it?
 - Do you have any concerns about sharing your DNA data online?

7.1.2.2 First part: Non-Experienced Group

- Do you know about DNA testing? (**Note: if the response is Yes**)
 - Could you explain what you know about DNA testing?

(Note: if the response is No. A description of DNA testing is provided by Interviewer) *A genealogical DNA test is a DNA-based test that looks at specific locations of a person's genome, in order to determine the level and type of the genetic relationship between individuals, verify ancestral genealogical relationships to estimate the ethnic mixture of an individual as part of genetic genealogy. For example 23 and me. There are various at-home DNA test kits available. The results may include health predispositions or conditions along with your ancestry too.*

- DNA test kits are available in supermarkets, If you get a big discount or get as a gift, would you decide to do the testing?

(Note: if the response is Yes.)

- why would you take it?
- What benefits do you think you can get from it?
- Do you have any concerns about it?

(Note: if the response is No)

- why would you not take it?
- Do you have any concerns about it?

Second part: Experienced and Non-Experienced Groups.

7.1.2.3 Note: Both groups watched GEDmatch tools demo video

Video transcription: *This is the home page of the GED match. You basically get your DNA tested in different consumer direct-to-consumer companies like 23 me or ancestry. And then, you get a raw DNA data sequence that you can upload here to find your genetic relatives or your biological parents or to complete your family tree. When you upload your DNA data, you get a kit number; the kit number is assigned to you. GEDmatch has many tools. I'm just going through pure popular tools, for example. One too many DNA comparisons, when you enter your kit number here, It basically compares your DNA data with all other DNA data that is present in the GED match, and it displays 3000 closes to match here. I want to know further about them. I can contact them through names and emails. And if I want to see how our chromosomes are related to each other, I can just click here. It gives me a 1-1 comparison. This is my kit number. This is the person I wanted to compare with. And I can actually see how our chromosomes are related to each other. Then the next tool we would be talking about is GEDmatch Forum. Users can ask and answers questions or concerns here or if they need any kind of help. The next tool we would talk about is GED COMM. You can enter someone's first name and last name, and you can actually see*

their details. Let me enter John smith. Here are all the users with the name John Smith. Now, let me choose one John smith to find out more details. I just click on their GED come. So this is the john smith details the birth, death, places, where and who the father, mother, and their children are. When you click the pedigree chart, you get other details like their family members.

- Would you share your DNA result online?

(Note: if the response is Yes)

- What are your expected benefits if you share?
- Do you have any concerns about sharing your DNA data online?

(Note: if the response is No)

- Why would you not share?
- Do you have any specific concerns about sharing your DNA data online?

- Is there any functionality you would be interested in using?
- Any functionality would you restrain yourself from using? and why?
- How frequently do you think you will visit the site?
- What do you think will happen with your data?
- Who do you think has access to your data?
- Rating: On a scale of 0-5 how would you rate your concern about your data being handled on those sites?
- Anything else would you add to what we have discussed so far?

Note: Description provided by interviewer : OPT-IN & OPT-OUT POLICY

There is a policy in GED match that is opt-in and opt-out opt in means your DNA that would be available for comparison to any raw data in GED match database. So any your data would be compared to any user. And opt-out means your DNA data would be available for comparison to any raw data in GED match database except DNA kit's identified as being uploaded for law enforcement investigation of a crime.

- Would you opt-in/ opt-out?
- What is the rationale behind your decision?
- Do you have any specific concerns?

**7.1.2.4 Note: Crime-related scenario - Golden state Killer Case
provided by interviewer**

Joseph James DeAngelo â known as the âGolden State Killer” for his spree of rapes and murders across California in the 1970s and 80s. Law enforcement solved this long 40 years pending case in April 2018 by using the GED match database. The investigative team had uploaded a DNA profile of him found in the crime site to GEDmatch by setting up a fake account to search for matching. Found matches on GED match which are his relatives and this information was used to zone in and capture the killer who was arrested and convicted.

- What is your opinion on law enforcement using the database in this case?

Provided by interviewer: *In this case, D’Angelo has not shared his DNA in GED Match but some of his relatives have shared their DNA with the site, which enabled the tracking and identification of D’Angelo by zoning in on his relativesâ records. That means that if anyone in your family or relative shares their DNA, it automatically shares or exposes part of your DNA sequence. Whenever people*

choose to share their DNA data, they implicitly share DNA data about others in their family tree which includes their present, past, and future relatives.

- What do you think about it?
- How would you feel if you were tracked or surveilled involuntarily by the law agencies? (**Note:Both open-ended and Likert scale**)
- How do you feel about sharing your own DNA sequence with law enforcement agencies? (**Note:Both open-ended and Likert scale**)
- How would you feel about your genetic relative or family sharing their DNA data? (**Note:Both open-ended and Likert scale**)
- What do you think your family will feel if you share your own DNA data?
- Follow up question: how do you think your family will feel about it? (**Note:Both open-ended and Likert scale**)

Note: Health-related scenario

7.1.2.5 Provided by interviewer:

Genetic data can help in research, medicine, or health fields.

- How would you feel about sharing your data for this purpose? (**Note:Both open-ended and Likert scale**)
- How would you feel about your genetic relatives or family sharing their DNA data for this purpose? (**Note:Both open-ended and Likert scale**)
- If you share your DNA data for these purposes, what do you think your family would feel about it? (**Note:Both open-ended and Likert scale**)
- Follow up question: how do you think your family will feel about it? (**Note:Both open-ended and Likert scale**)

7.1.2.6 Provided by interviewer:

Genetic data can reveal hereditary diseases (diseases passed from generation or family), expected, present or past illness of a person for example probability of breast cancer or prostate cancer.

- How would you feel about sharing your data in this case? (**Note:Both open-ended and Likert scale**)
- How would you feel about your family member sharing their DNA data in this case? (**Note:Both open-ended and Likert scale**)
- If you share your DNA data, what do you think your family would feel about it? (**Note:Both open-ended and Likert scale**)
- Follow up question: how do you think your family will feel about it? (**Note:Both open-ended and Likert scale**)

7.1.2.7 Provided by interviewer:

There has been debate in some parts of the world whether genetic data should be accessible to insurance companies or not. If allowed insurance companies can pull your genetic data while resolving a claim.

- How would you feel about sharing your DNA data with insurance companies? (**Note:Both open-ended and Likert scale**)
- How would you feel about insurance companies pulling your genetic data while resolving one of your claims? (**Note:Both open-ended and Likert scale**)
- How would you feel about your family sharing their DNA data with insurance companies? (**Note:Both open-ended and Likert scale**)

- If you share your DNA data with the insurance companies, what do you think your family would feel about it? (**Note:Both open-ended and Likert scale**)
- Follow up question: how do you think your family will feel about it? (**Note:Both open-ended and Likert scale**)

7.1.2.8 Last part: Experienced and Non-Experienced Groups.

- What kind of setting changes or policies or preference changes would make these platforms more privacy-preserving?
- To what extent would you expect DNA data sharing will be popular in the future?
- What are the privacy expectations from this type of platform in the future?
- Would you use this if your friends or family members or people around you are using it?
- What situation would you like to use?

7.1.3 Demographics of the participants

Table 7.1: Experienced Group

	Gender	Age	Education or Occupation	Testing company
E1	F	59	Accounting	23&me
E2	F	52	Facility management	AncestryDNA
E3	F	29	education	AncestryDNA
E4	F	54	International Studies	AncestryDNA
E5	M	20	Electrical engineer	AncestryDNA
E6	F	22	Art, education	23&me
E7	F	28	Religious studies	AncestryDNA
E8	F	31	Executive assistant	AncestryDNA
E9	F	22	Religious studies	CRI Genetics
E10	F	59	Office manager	AncestryDNA
E11	F	54	Anthropology/FT Staff	23&me AncestryDNA
E12	F	39	Physics	23&me AncestryDNA
E13	M	56	Emergency planner	23&me
E14	F	38	Career coach	23&me
E15	F	21	Computer science	AncestryDNA
E16	F	20	Civil engineering	AncestryDNA
E17	F	18	Psychology and Spanish	23&me
E18	F	24	MS Health Informatics	23&me
E19	F	52	Human Resources	AncestryDNA
E20	F	35	Education	23&me
E21	F	57	Accounting	AncestryDNA
E22	F	23	Anthropology	AncestryDNA
E23	F	65	Librarian	AncestryDNA
E24	F	34	Psychology	23andMe
E25	F	59	Office manager	AncestryDNA
E26	F	27	Biology	AncestryDNA
E27	F	25	Fine Arts	23andMe
E28	F	47	Research Compliance	AncestryDNA
E29	F	33	Library faculty	AncestryDNA
E30	F	59	Finance	23andMe AncestryDNA

Table 7.2: Non-experienced Group

	Gender	Age	Education or Occupation
NE1	F	22	Mass Media and Journalism
NE2	F	40	Translation. InterpretOR
NE3	F	46	MBA
NE4	F	20	Math Teacher
NE5	F	24	Educational Leadership
NE6	M	45	IT Analyst
NE7	F	24	Social work
NE8	F	57	Graduate School
NE9	F	21	Mathematics
NE10	F	29	Data science
NE11	F	21	Computer Science
NE12	F	43	Teaching
NE13	F	24	Biology
NE14	F	48	DSBA
NE15	F	39	Data Analyst
NE16	F	24	Public Administration
NE17	M	21	Computer Science
NE18	M	25	Exercise science
NE19	M	23	Computer Science
NE20	F	27	Graphic Design
NE21	M	31	Public policy
NE22	F	42	Academic Advisor
NE23	M	50	IT
NE24	F	51	higher ed admin
NE25	F	20	Poli Sci/Sociology
NE26	M	38	Public health
NE27	F	27	Electrical engineering
NE28	F	19	Meteorology
NE29	F	19	Computer Science
NE30	M	18	Computer Science

7.2 APPENDIX B: Supplementary data for Chapter 4

7.2.1 Video transcripts

Personal story video - *Hello everyone! Today I am going to talk about and share my experience of consumer DNA tests. I bought the DNA test kit from a supermarket for less than 100 bucks. I registered the kit online with the company and then mailed my DNA sample. Once the results were ready, they notified me. The results gave me deep insight into my ancestry such as where my ancestors are, my heritage, and what kind of ancestry compositions. It also gave me detailed reports about my future health risks, and what I am predisposed to. Not only about health predispositions, but they also talked about my traits and wellness which helped me to make better lifestyle choices. I was also able to find my DNA relatives online and connect with them. But Recently, I watched news saying people don't realize that DNA is very sensitive information. Giving away DNA can end up in the hands of unknown third-party companies; there are hardly any prohibitions. The only major law that protects users' genetic privacy is GINA which is Genetic Information Nondiscrimination Act. But GINA does not cover long-term illness or life insurance. So, DNA access by insurance could lead to raised premiums or denial of services. Also through the DNA test, you get to know your hereditary diseases, which means diseases or health conditions run in the family. That means when I decided to share my DNA data, I implicitly shared DNA data about my present, past, and future relatives, about part of their health details too without their consent. Sharing genetic information could also lead to genetic discrimination. This reminds me of one instance that African Americans were discriminated against as they have a higher chance of sickle cell conditions. Also with more research, I found that law enforcement uses DNA databases to solve crimes. They do familial searches; familial searches means they try to find matches in the database; if any of your long-distance cousin is found, they can surveil all relatives to find the convict. Lastly, I just want to tell you guys, if you really need to take the*

test, take it as there are tons of good reasons such as an adoptee could find his family but before that, just weigh the risks and benefits for you.

Conversational story video - F - *Hey, I got this ancestry DNA testing kit, they are on sale.*

M - What would you do with that?

F - we can know ancestry, health predispositions, find relatives, and traits. This would help us to take better decisions about our lifestyle choices and we can also take preventive actions by learning about our future health risks.

M - Cool, I can gift this one to my aunt. That could be a big help to her, she is adopted, and she always wants to know about her biological family.

M - Hey, why don't you browse about these DNA tests? We can get some more ideas about it

F - Sure!

M - Did you find anything?

F - (Reading article) Ya, I m looking at some articles, it seems pretty concerning. These DNA databases are used by law enforcement, at least accessed in 70 cases in the past few years. As DNA is inherited, if I share my DNA, then I am sharing my family's DNA too without their consent. So everyone related to you could be dragged involuntarily to police investigations.

F - (After a while) Think about if insurance access to your health information, it could be a big problem, they could deny to cover you. Breast cancer and diabetes run in my family, they could refuse to cover me, maybe my sister too you never know.

F - (Reading the article again) It says these tests do not fall under health privacy acts like HIPPA and also the major genetic privacy act GINA does not cover long-term health issues or even life insurance. Also, revelation of DNA could lead to genetic discrimination. for instance, in the past African Americans were treated unfairly due to the stigma of higher chances of sickle cell condition in their genes. Lately, there

have been many instances of hacking that exposed millions of people's genetic data. These data are also accessed by third parties including commercial companies.

M - Oh my god! Good that we browsed, I think it's always good to know both the benefits and risks before deciding.

Informational video - *Today we are talking about the benefits and risks of taking at-home DNA testing and sharing your DNA data with private companies. Consumer DNA testing kits are becoming increasingly popular. More than 26 million people have taken an at-home ancestry test. It's a good way to find out more about your genetic makeup, family history, genetic relatives, ancestry, ethnicity and health risks, traits which could help you to make better choices in lifestyle or take preventive actions by knowing health predispositions. But privacy researchers have shown concerns as DNA is a very sensitive and private data of an organism. DNA molecule is the ultimate identifier of an organism that holds the entire biological instructions of an individual. Putting your DNA information on the internet or public databases could lead to non-intended data access and sharing with third parties or even access by insurance companies which can lead to denial of insurance or raised premiums. Lately, there are many instances of hacking which led to millions of DNA data being exposed online. Furthermore, the revelation of DNA data could lead to risks like genetic discrimination. For example, in the past African Americans were treated unfairly due to the stigma of sickle cell conditions. Ancestry services do not fall under health privacy acts like HIPPA, also the only major law that protects genetic privacy is Genetic Information Nondiscrimination Act, called GINA. But GINA does not cover life insurance or long-term illnesses. So predisposition to long-term health problems could lead to denial of insurance. Commercial pharmaceutical companies could control the drug market by learning about genetic or health conditions. Health conditions in the family, such as diabetes and cancer, are revealed and could be accessed by insurance. Moreover, these databases are used by law enforcement to solve cold cases.*

More than in 70 cases public DNA databases are used by police. As DNA is inherited in the family, relatives youâve never met can take DNA tests that could drag you to police investigation or involuntarily surveilled. So when you choose to share your DNA data, you share your present, past, and future familyâs DNA data too These tests compromise the genetic privacy not just of people who choose to take the tests, but also their distant relatives who havenât consented to anything. Currently, More than 70% of Europeans or Americans could be identified as some of their family member has taken this test. So, you really need to think carefully about whether it's something that can really benefit you as now there are a ton of good reasons such as learning more about ancestry, genetic family. or is it just for fun as we don't necessarily know all the ways in which this data could be used?

7.3 APPENDIX C: Survey questions for Chapter 5

Ask about their interest, concerns in sharing their DNA data

- Explain at home DNA testing
- Are you interested in taking at-home DNA testing? Yes, Why?, NO Why?
- What benefits do you think would you get if you do at-home DNA testing? Or what interests you the most?
- Do you have any concerns about taking an at-home DNA test?

Show the video of ged match explaining what it does and how it does: existing video of gedmatch

- Are you interested in sharing your DNA data in public genealogy databases if you take a test? Yes, Why? NO Why?
- What benefits do you think would you get if you share DNA data in public genealogy databases ? Or what interests you the most?
- Do you have any concerns about sharing DNA data in public genealogy databases?

Pre survey

Motivation

If you plan to take at-home DNA testing, then why would you?

- Ancestry (1- Not at all interesting, 2- little interesting 3- moderately interesting, 4- very interesting, 5- Highly interesting)
- DNA relative finding (1- Not at all interesting, 2- little interesting 3- moderately interesting, 4- very interesting, 5- Highly interesting)
- Health predisposition (1- Not at all interesting, 2- little interesting 3- moderately interesting, 4- very interesting, 5- Highly interesting)

- Wellbeing and lifestyle (1- Not at all interesting, 2- little interesting 3- moderately interesting, 4- very interesting, 5- Highly interesting)
- Traits (1- Not at all interesting, 2- little interesting 3- moderately interesting, 4- very interesting, 5- Highly interesting)

Show the video of ged match registration

This is the registration page and privacy policies of GEDmatch. Please tell me how would like to fill up in each field

- Would you use your real name? If yes, why? If no. why?
- Would you use an alias? If yes, why? If no. why?
- Would you use your regular email address? If yes, why? If no. why?
- Would you use your name as the donor? If yes, why? If no. why?
- Do you think you can upload othersâ information here? If yes, why? If no. why?
- Which Privacy option would you choose? And why?
- Opt out - what does this mean? What is the exposure these people have?
- Opt in- what does this mean? What is the exposure these people have?
- Research - what does this mean? What is the exposure these people have?
- Private- what does this mean? What is the exposure these people have?

Awareness

- Who do you think can have access to your data?
- Do you think law enforcement can have access to your data in public genealogy databases?

- Do you think third parties can have access to your data in public genealogy databases?
- Do you think insurance can have access to your data in public genealogy databases?
- Do you think public genealogy databases can be breached?
- Who covers these tests? HIPAA FDA GDPR GINA NONE ALL OF IT

Sharing with other entities (Four point likert scale - Extremely likely to - Extremely Unlikely)

- I would share my DNA data with law enforcement
- I would share my DNA data for research
- I would share my DNA data with insurance companies
- I would share my DNA data with government
- I would share my DNA data with third parties

par **Post survey**

Motivation

If you plan to take at-home DNA testing, then why would you?

- Ancestry (1- Not at all interesting, 2- little interesting 3- moderately interesting, 4- very interesting, 5- Highly interesting)
- DNA relative finding (1- Not at all interesting, 2- little interesting 3- moderately interesting, 4- very interesting, 5- Highly interesting)
- Health predisposition (1- Not at all interesting, 2- little interesting 3- moderately interesting, 4- very interesting, 5- Highly interesting)

- Wellbeing and lifestyle (1- Not at all interesting, 2- little interesting 3- moderately interesting, 4- very interesting, 5- Highly interesting)
- Traits (1- Not at all interesting, 2- little interesting 3- moderately interesting, 4- very interesting, 5- Highly interesting)

Show the video of ged match registration

This is the registration page and privacy policies of GEDmatch. Please tell me how would like to fill up in each field

- Would you use your real name? If yes, why? If no. why?
- Would you use an alias? If yes, why? If no. why?
- Would you use your regular email address? If yes, why? If no. why?
- Would you use your name as the donor? If yes, why? If no. why?
- Do you think you can upload othersâ information here? If yes, why? If no. why?
- Which Privacy option would you choose? And why?
- Opt out - what does this mean? What is the exposure these people have?
- Opt in- what does this mean? What is the exposure these people have?
- Research - what does this mean? What is the exposure these people have?
- Private- what does this mean? What is the exposure these people have?

Awareness

- Who do you think can have access to your data?
- Do you think law enforcement can have access to your data in public genealogy databases?

- Do you think third parties can have access to your data in public genealogy databases?
- Do you think insurance can have access to your data in public genealogy databases?
- Do you think public genealogy databases can be breached?
- Who covers these tests? HIPAA FDA GDPR GINA NONE ALL OF IT

Sharing with other entities (Four point likert scale - Extremely likely to - Extremely Unlikely)

- I would share my DNA data with law enforcement
- I would share my DNA data for research
- I would share my DNA data with insurance companies
- I would share my DNA data with government
- I would share my DNA data with third parties