Understanding Context of Use and Perceptions of Usability of Cosegregation Analysis Tool AnalyzeMyVariant

Aarti Swaminathan

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2018

Committee:

Annie T. Chen

Brian Shirts

Program Authorized to Offer Degree:

Biomedical and Health Informatics

University of Washington

Abstract

Understanding Context of Use and Perceptions of Usability of Cosegregation Analysis Tool AnalyzeMyVariant

Aarti Swaminathan

Chair of the Supervisory Committee:

Annie T. Chen

Biomedical Informatics and Medical Education

Calculating the genetic risk for a disease with allelic variants of unknown significance can be a complicated task. AnalyzeMyVariant is a tool designed for genetics experts that uses pedigree data from families with genetic variants of unknown significance, to calculate likelihood ratios that a variant fits pathogenic or benign patterns. In this study, we performed a two-part evaluation to understand the context within which genetics experts might use this tool and assess their initial usability perceptions.

First, we surveyed existing literature to develop an instrument to assess perceptions of usability based on constructs of usability, quality, and safety. The instrument consisted of scaled

as well as open-ended questions assessing users' perceptions relating to each of the constructs of interest, with regard to their experience with AnalyzeMyVariant. We used the instrument to collect data from 57 genetic experts and trainees who were recruited via email invitations. The second part of our evaluation was comprised of semi-structured interviews with six genetics experts to identify work contexts in which users might use the tool and further delve into issues faced in using the tool. These interviews were inductively coded and major themes identified using the constant comparative method. Based on these findings, we provide recommendations for future improvement of the tool.

This work has importance in the consideration of the varying needs of genetics professionals and how they use cosegregation analysis in their work, and the difference between requirements for research- and clinically-focused work. The results could also inform the future development of other tools developed for experts in a wide area of scientific fields, particularly with regard to the attention that must be paid to experts' context of use, background knowledge, and the intended applicability of results.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor and committee chair Dr. Annie T. Chen of the Biomedical Informatics and Medical Education at the University of Washington School of Medicine. She provided invaluable guidance throughout the process, while consistently allowing this paper to be my own work.

Secondly, I would also like to thank Dr. Brian Shirts, my second thesis committee member, for allowing me to conduct this project on his tool AnalyzeMyVariant and providing guidance, support and clarification about all matters related to genetics and AnalyzeMyVariant.

I would also like to thank members of the department of Biomedical Informatics and Medical Education at UW Medicine, the team working on AnalyzeMyVariant, and members of Dr. Shirts' lab who helped us with pilot testing the survey tool. Without their passionate participation and input, the survey study could not have been successfully conducted.

Finally, I must express my very profound gratitude to my parents and family for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

1 INTRODUCTION AND BACKGROUND

Given that the human genome encompasses an incredible amount of variation from individual to individual, with many of these variations still ostensibly producing overlapping phenotypes, analyzing the variation in genes poses a great challenge. With specific regard to using genetics to characterize and predict disease risks for conditions like breast cancer, the disease likelihoods of many genetic variants remain uncharacterized. Usually, this is because these variants do not occur very frequently in the population. Uncharacterized variants in genes with known clinical associations are termed variants of unknown or uncertain significance.

AnalyzeMyVariant [1] is an online tool developed for genetics specialists to use pedigree information from family-scale genetics studies to characterize the significance of genetic variants of hitherto unknown significance, for eight different cancer genes. This is done by cosegregation analysis of the pedigree data, using established penetrance values for each gene. Cosegregation analysis is an established statistical method, but its use has historically been limited to a small number of statistical geneticists with experience using specialized software [2], [3]. AnalyzeMyVariant was developed with the intent of allowing a larger group of both research and clinical genetics practitioners (e.g. genetic counselors and statistical geneticists) to perform quantitative cosegregation calculations.

To truly be of value, however, perceived usability of the tool is of great importance, especially in an increasingly competitive online space for services and an increasingly tech-savvy user population. When a user is satisfied with the tool, he/she is more likely to continue to use that tool more frequently, and also more likely to recommend that tool to others. In this way, perceived usability is vital to the greater adoption of AnalyzeMyVariant as a tool and eventually to its overall success with its target audience.

In this study, we seek to comprehensively assess perceived usability for the website. While many assessment tools for websites have been developed for a wide variety of contexts, including e-commerce and informational sites, and using a variety of approaches, the literature on models developed for this kind of website tool is lacking. The use considerations of this tool form an interesting intersection of e-commerce and informational use. Although not an e-commerce website, some of the transactional nature of e-commerce websites is present. Users are required to upload their clients' personal information right at the beginning of the interaction, which requires that they trust the integrity and accuracy of the tool quickly. At the same time, the tool is meant to provide important information to users who upload data for analysis, explaining not only the results generated, but also the methods used in the calculation of the result. Therefore, it contains both a transactional element, similar to e-commerce, as well as an element of information provision, similar to informational websites. A model that seeks to assess either one or the other type of website cannot be used to sufficiently analyze this tool in entirety. In order to address this, we chose to conglomerate different extant models to create one suitable for the context of AnalyzeMyVariant. In the next section, we present an overview of the tool, followed by a brief review of general literature assessing usability, before summarizing the foundational structure of the model used.

1.1 Overview of the tool

As genetic testing for hereditary diseases become more common, genetic variants that have an unknown clinical significance can be problematic for genetic counselors who are trying to manage their patients' risk of these conditions, especially cancer.

Given that there is no standardized way to classify these variants of unknown significance, and that these variants may have a large or small amount of evidence that may be consistent or contradictory from multiple other sources, interpretation of such variants, and clear presentation to patients of the implications of these results can be a rather difficult task. Clinicians and genetic counselors often need to rely on a multitude of different information sources to assess the potential pathogenicity of the variant [4], ranging from checking online databases to see how the variant has been classified by various sources [5], to assessing functional studies to look at family history. These methods also differ from condition to condition, as certain types of cancer might have more physiological indicators than others [4]. Moreover, some of these rare variants may be found only in single families, precluding the ability to find out more about using population-based strategies [6].

In such cases, family-based cosegregation analysis can be a valuable asset to clinicians looking to understand how to present genetic test results to patients [7] [8]. Robust statistical methods exist to quantifiably and concretely inform genetic counselors of the exact level of risk [3] [9]. Furthermore, these quantitative results can be validated and combined with other quantitative data, to be able to present robust evidence to re-classify a variant [7].

However, genetic counselors often lack the statistical expertise to run these tests themselves, and therefore end up analyzing these results only by qualitative means, coming up with qualitative assessments of whether the variant segregates with the disease or not [2]. Since these methods are qualitative, they can often differ widely across institutions and clinics, which makes it harder to maintain a standard of pathogenicity classification of genomic variants.

The tool being evaluated in this study, AnalyzeMyVariant, allows users to calculate quantified risks of a specific disease for a genetic variant of unknown significance, using genetic

information from families with these variants of unknown significance. Using pre-defined models of disease penetrance within the population, the tool tracks the allelic segregation of the variant through the family, along with the occurrence of the disease in family members, to generate likelihood ratios that a particular variant is pathogenic [10]. It uses 2 separate robust statistical methods, the full likelihood Bayes factor (FLB), and cosegregation likelihood ratios (CSLR), both of which outperform other, simpler methods of analysis, such as counting meiosis [3] [9] [11].

Genetics experts who seek to use this tool can create a pedigree file of their client's family variant information, according to the formatting instructions provided. Upon uploading the results in an Excel or text file format, and selecting a gene to run the calculation with, the tool will generate results as shown in figures 1 and 2. Results shown in both figures are for the example file that is provided on the website and the BRCA2 gene, which is one of eight different cancer gene options currently available. Figures 1 and 2 show the interface which users interact with.

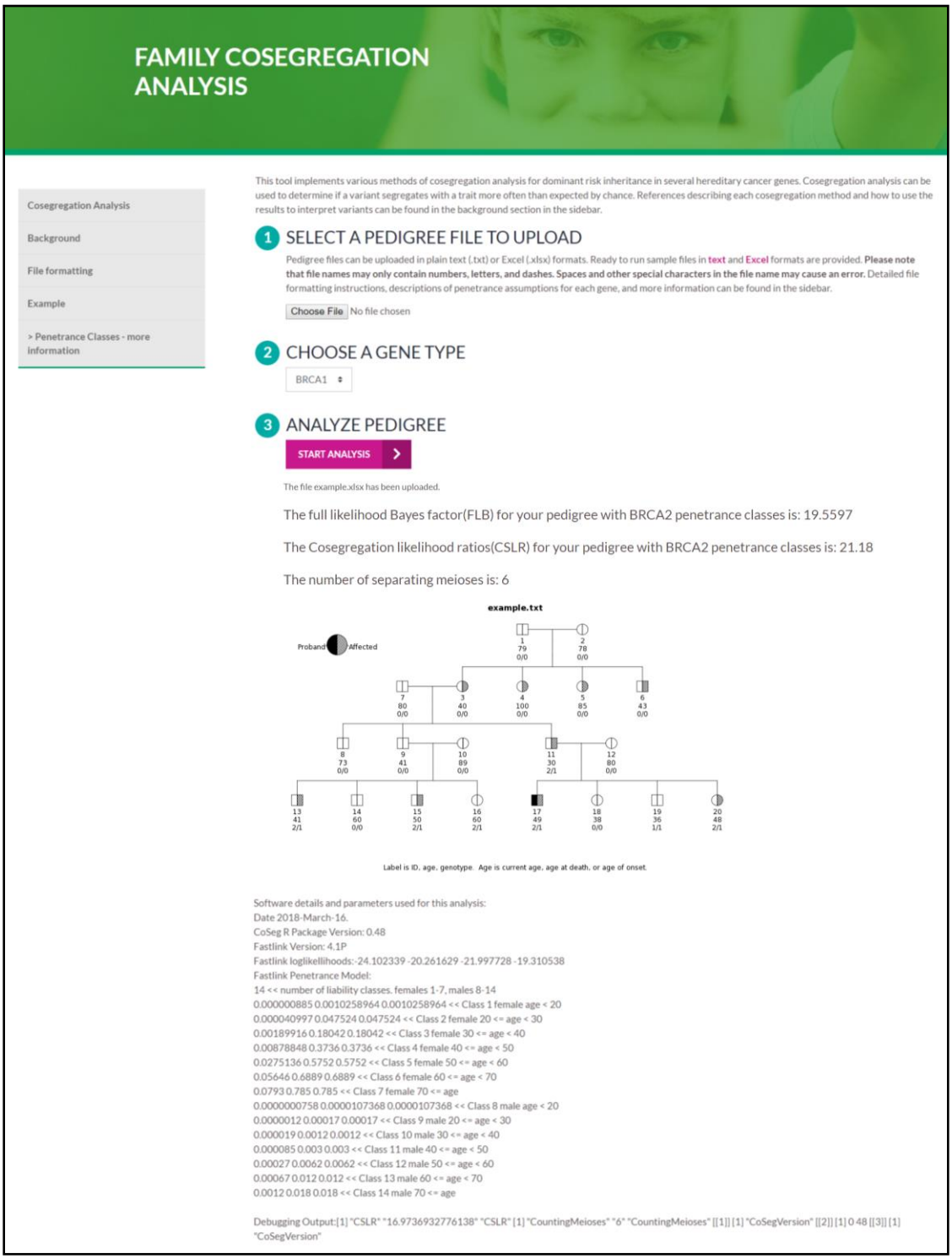

Figure 1. AnalyzeMyVariant *interface – main page, where users can upload files, select genes, and view numerical and graphical results, along with software parameters used in the analysis. The extra sidebar on the left allows navigation to other pages.*

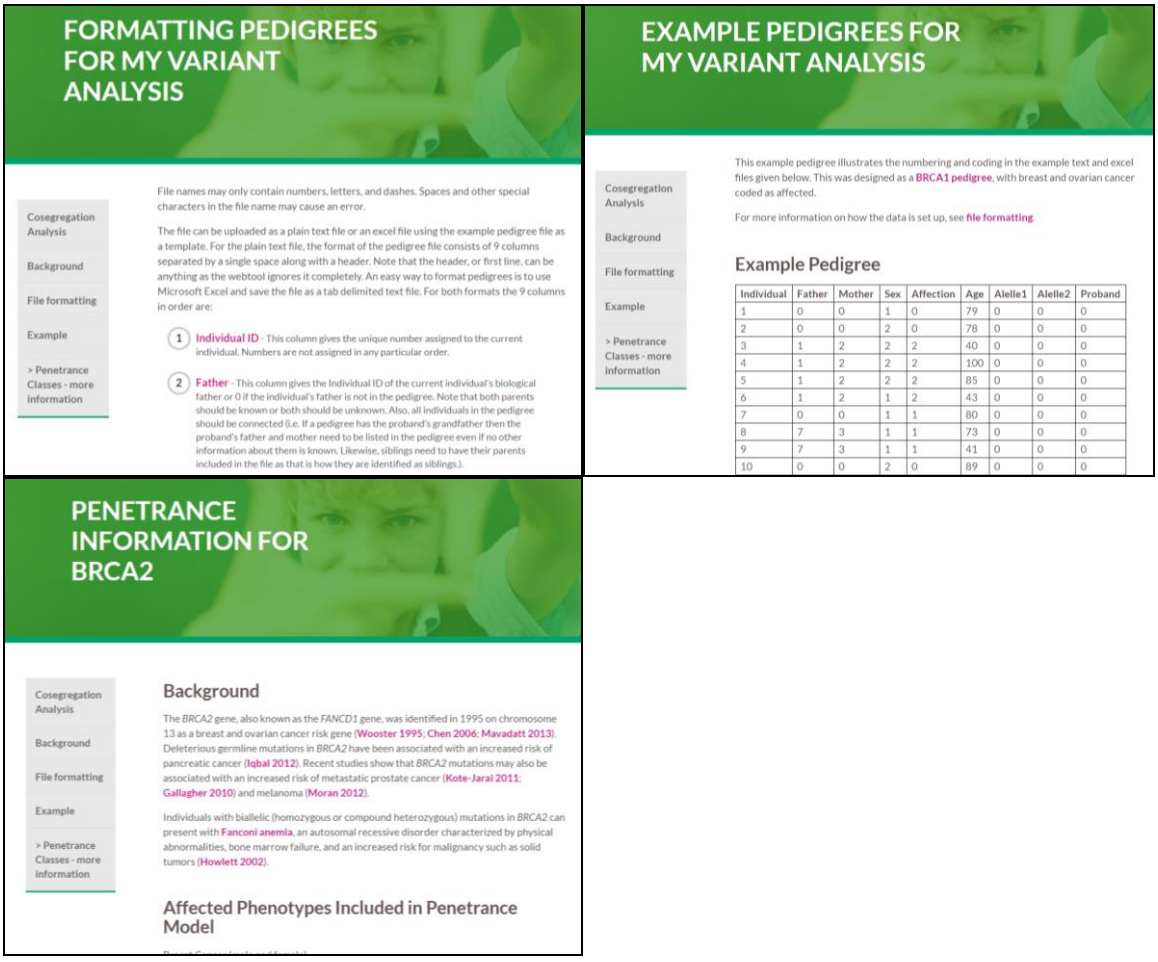

Figure 2. Other pages in the AnalyzeMyVariant *website. (Clockwise from top left): instructions on how to format pedigree files for use with the tool; a detailed description of the example file; and penetrance information for each gene the website currently provides analysis for - shown here is the information for BRCA2.*

By allowing genetic counselors with limited statistical knowledge to perform complex statistical analyses quickly and simply, this tool could potentially be a great asset for experts in the genetics community, both on the large scale of research to help re-classify a variant, and on

the smaller scale of helping clinical counselors to understand how to present test results to patients and modify treatment recommendations accordingly. Assessing how target users actually find the application and understanding whether it provides all that is needed to enable counselors to quantitatively assess and understand disease risk, is therefore of vital importance to promoting greater adoption of the tool. In order to do this, we conducted a literature review to develop a survey instrument to assess all the factors involved.

1.2 Literature review

Previous literature has sought to assess a variety of different constructs in evaluating quality in an online realm. These constructs include trust, satisfaction, credibility, quality and usability. Evaluations of these constructs have also been conducted in a variety of different settings. Of these settings, the two major categories are informational and e-commerce websites [12], although special cases similar to AnalyzeMyVariant also exist. We employed a search of journal databases, using keywords like “assessment of website/online tool”, “usability”, “quality”, “user experience”, “satisfaction”, “trust”, as well as keywords signifying other constructs of interest, searching for models that assessed the constructs mentioned above. Journal databases included general search engines like Google Scholar and Semantic Scholar, as well as more specific databases like PubMed and ACM Digital Library.

Findings from major studies are summarized in table 1.

Table 1. Constructs evaluated on various platforms and their components

Main construct	**Components of model**	**Reference**
Trust	Credibility, ease of use and risk	Corritore et al. [13]
	Process (IT capability), performance (user perception of efficacy) and purpose (user perception of intentions)	Söllner et al. [14]
Quality	Reliability, competence, responsiveness, ease of use, security, and product portfolio	Yang & Jun [15]
	Content usefulness, informational adequacy, usability, accessibility, privacy/security and interaction	Yang et al. [16]
	User interface, responsiveness, need fulfillment and security	Gummerus et al. [17]
Usability	Consistency, navigability, supportability, learnability, simplicity, interactivity, readability, content relevance, credibility and telepresence	Lee and Kozar [18]
	Efficiency, effectiveness, satisfaction, learnability, productivity, safety, trustfulness, accessibility, usefulness and universality	Seffah et al. [19]
	Ease of use, responsiveness, need fulfillment, privacy, personalization, visual appearance, information quality, trust, interactivity, advertising, playfulness and technology integration	Chiou et al. [20]
	Efficiency in use, learnability, rememberability, reliability in use, user satisfaction	Constantine &

Main construct	**Components of model**	**Reference**
		Lockwood [21]
	Speed of performance, time to learn, retention over time, rate of errors by users, subjective satisfaction	Shneiderman et al. [22]
	Efficiency of use, learnability, memorability, errors/safety, satisfaction	Nielsen [23]
	Throughput, learnability, attitude	Preece et al. [24]
	Effectiveness/speed, time to learn, retention, effectiveness, attitude	Shackel & Richardson [25]
Credibility	Situational factors (those that orient an unfamiliar user); visual design factors (interface design and colors); and user variable factors (degree of familiarity users have with the subject matter, their understanding of the system)	Fogg & Tseng [26]
	Web use motivation and information completeness	Dutta Bergman [27]
Satisfaction	Task relevant cues, involvement, atmospheric responsiveness	Eroglu et al. [28]
	Informational and system dimensions: expectation, perceived	McKinney et

Main construct	Components of model	Reference
	performance and disconfirmation for each dimension	al. [29]
	Ease of use, customer service, web community, privacy, web expertise, experiential quality, flow, informational quality	Jaiswal et al. [12]

It is interesting to note that these models were created for different purposes, and therefore have different approaches. For example, in assessing trust, Corritore et al. [13] discuss trust in HCI systems and different kinds of trust (slow and swift), as well as a life-cycle of trust, where accuracy promotes trust but errors lead to loss of trust, with the loss becoming greater with each subsequent perceived error. Cues of trust include website interface and design, overall look, grammatical correctness and professionalism, conveying expertise, providing comprehensive information, projecting honesty, lack of bias and shared values between the website and the user. By contrast, Söllner et al. [14] contrast such human defining factors to dimensions of performance, process and purpose. The *performance* dimension reflects the capability of the IT artifact in helping the user to achieve his goals. The *process* dimension reflects the user's perception regarding the degree to which the IT artifact's algorithms are appropriate. Finally, the *purpose* dimension reflects the user's perception of the intentions the designers of the IT artifact had. Both of these models look at a cross-sectional view of trust at a point in time. Sillence et al. [30], on the other hand, describe a longitudinal model of trust, where trust is initially assessed by heuristic, superficial analysis, then systematic evaluation and finally

to a stage where information is integrated across sites and sources and longer-term consultation is performed.

It is also of note that many of the main constructs are interlinked, even within separate models. Gummerus et al.'s [17] four components of quality (user interface, responsiveness, need fulfillment and security) together contribute to trust, and eventually satisfaction, in a larger model. Similarly, credibility appears as a component in various studies, but is also independently evaluated by researchers as a larger construct in its own merit [25] [26] [27]. Yang et al. [16] use these dimensions of system and informational quality and combine them with elements of the Technology Acceptance Model (TAM) and the six proposed quality dimensions mentioned in table 1.

Although the majority of literature found was based on informational and e-commerce websites, some unique tools similar to AnalyzeMyVariant have also been evaluated. A study was conducted by Cohn et al. [31] to evaluate Health Heritage, an online tool to "guide users through the collection of their family health history by relative, generate a pedigree, complete risk assessment, stratification, and recommendations for 89 conditions" [32]. Key measures related to (a) the initial experience and satisfaction of the Health Heritage users, (b) completeness and accuracy of family health history information and (c) subsequent appropriateness of risk assessments were assessed for this study. Measures used to assess the website were: time taken to complete filling out the information, degree of completeness of information entered, quality, satisfaction and whether the information was "worth the time and effort".

Shyr et al. [33] also conducted a usability study of exome analysis software. While a model was not developed to assess usability, many themes in the issues brought up by users as

part of think-aloud evaluations are similar to what has been seen in other models: navigation, layout, consistency, information completeness, overall ease of use, and system response time.

As can be seen in the models so far discussed, many of the dimensions discussed in various models tend to overlap. Some of the larger constructs even fit inside each other, with trust appearing within models of quality and usability, and credibility appearing within models of trust and usability.

After conducting a literature review of different models that assessed perceptions of usability, we then developed a set of constructs to use in our tool, taking into consideration previous models, as well the expected context within which users might make use of this tool. This process, as well as the constructs chosen, are described in the methods section below.

2 METHODS

The methods section of this paper is divided into two sections. First, we first describe the methods used to build the survey, test it, recruit participants and analyze the data. We then describe the process of creating a contextual interview protocol, identifying participants, and analyzing interview data. Insights gained from the data using these methods are then described in the remainder of the paper.

2.1 Survey methodology

2.1.1 Selection of constructs

Taking into consideration the medium of the assessment tool and the situational use of the tool, we systematically went through constructs from different studies, as defined by each of those studies, and judged their appropriateness for our purpose.

This purpose was to create a survey tool that could effectively assess participants' experiences in getting what they needed from a technical site designed for professional calculations and understanding what the barriers were to their being able to use this website. Since we expect this to be a professionally valuable asset in their line of work, for use with real data from real families, constructs such as 'playfulness' were considered of lesser importance. Also, since this is a very specific service being provided to a particular population of users, constructs that seek to acquire information about adapting it to a larger population, like 'universality', were again deemed to be of less value. Similarly, since the first stage of the study was a survey tool assessing participants' experiences in their first interaction with this tool, longitudinal constructs such as 'learnability over time' would be more difficult/abstract to evaluate. We also considered other factors like the conceptual overlaps seen in the general

evaluation models explored, their comparability to the limited research on evaluations of tools that were similar to AnalyzeMyVariant, and the context of development of each model and validation of the different approaches. Thus, we arrived at the following list of constructs, listed and defined below in table 2.

Table 2. Construct definitions

Construct	**Definition**	**Source**
Privacy	The degree to which access to sensitive information is controlled.	[12] [16] [20]
Security	The degree to which safety and security are maintained.	[15] [16] [17] [19] [23]
Information completeness	The extent to which information is comprehensive and up-to-date.	[12] [16] [20] [27] [33]
Information relevance	The extent to which information is relevant to users' needs.	[12] [18] [20]
Need fulfillment	The extent to which the website can actually satisfy the user's requirements.	[17] [20]
Reliability	The accuracy, efficiency and technical dependability of the calculator.	[15] [21]
Ease of use	The extent to which the user finds it simple to interact with the website.	[12] [13] [15] [20]

In some cases, single constructs in other studies became multiple constructs in ours. For example, the construct of 'informational quality' [12] [20] was defined in some studies as

encompassing both information completeness and information relevance of information, while other studies defined these separately. Again, considering the context of use, we chose to include them separately. Conversely, while some studies used multiple constructs to assess ease of use of a tool, such as navigability, readability and simplicity [18] in table 1, we chose to evaluate ease of use as a single construct, so as not to make the survey too lengthy. Based on the identified constructs, the survey was designed, as discussed below.

2.1.2 Survey design

Based on the preceding literature review, we developed questions to assess each construct in table 2. Since we wanted a combination of quantitative feedback to gauge and compare user response, and quantitative feedback to identify leverage points for improvement, we developed multiple-choice and open-ended questions for each construct. The initial set of questions is shown below in table 3.

Multiple choice questions consisted of statements to which participants could select answers from a Likert-type scale of agreement (totally disagree, partially disagree, neutral, partially agree, totally agree). Open-ended questions followed a basic why/what format. Phrasing of the questions was based on previous studies, which are cited in the last column on table 3. Other questions for which relevant survey examples could not be found were then phrased accordingly, for consistency of language.

Table 3. Survey questions

Construct	**Item**	**Source**
Ease of use	I found the website clear and easy to use	[18] [34]
	I found the results clear and easy to understand.	[35]

Construct	Item	Source
	What, if anything, makes it difficult to use?	
Output relevance	I found the output relevant.	[35]
	Why not?	
Need fulfillment	The tool told me exactly what I needed to know.	[35]
	Why not?	
Reliability	I found the tool to be: - Accurate - Efficient - Reliable	[35]
	What were your reasons for giving the above ratings?	
Information completeness	I found the information provided on the website to be thorough and complete.	[27] [18]
	Why not?	
Security	I feel secure in my interaction with the website.	[36]
	Why not?	
Privacy	The website protects my right to control access to the personal information that I uploaded.	[12]
	What could be changed to improve your sense of privacy?	
Trust	I feel like I can trust this tool.	[36]
	If you did trust the tool, what aspects contributed to that? If you didn't trust it, what was missing from the tool/wrong with it?	[37]

Construct	Item	Source
Satisfaction	I was satisfied with my experience with the tool.	
	What could be added/changed to improve your experience?	

In order to break up the survey into sections that would encourage participants to think about different aspects of perceived usability as they went through the website and filled out the survey, we turned back to the literature to figure out how to group different constructs together. We developed sections based on how different studies classified these constructs. The categorizations are shown in figure 3. Where different studies had different classifications, the conflict was resolved by taking into account the following factors: the extent to which the model was used (inferred from the number of citations of each paper), the context of development, the context of validation, and all the places where constructs appeared together (relationships between constructs). While privacy and security were often categorized under quality, we chose to split them into a different section to prevent the quality sections becoming too long (since reliability itself had three parts).

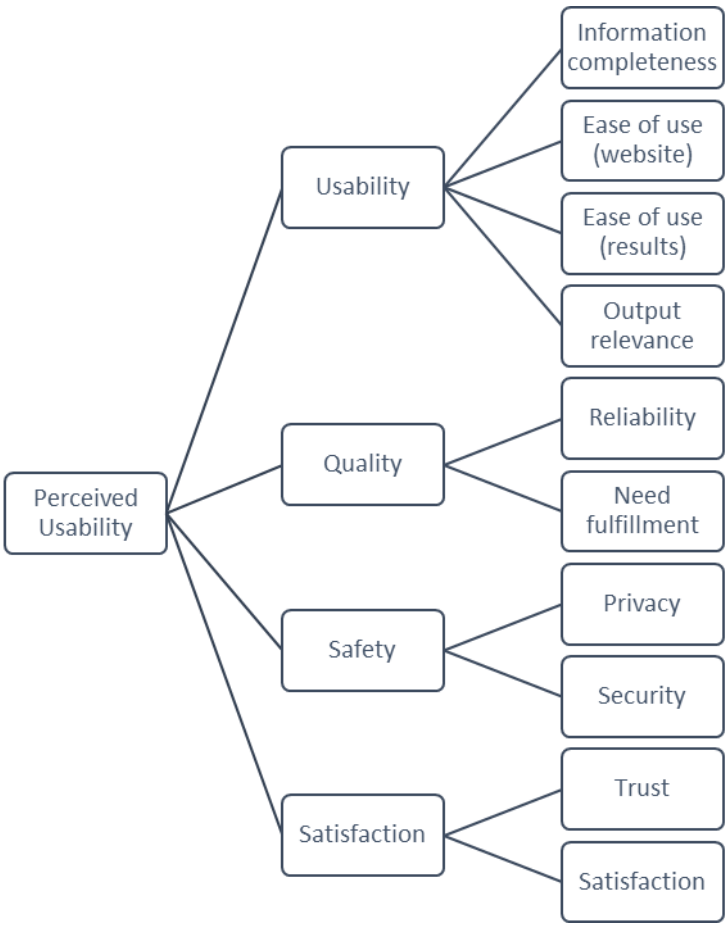

Figure 3. Hierarchy of constructs

The survey was developed using the REDCap electronic data capture tools hosted at the University of Washington. REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies, providing 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources [38].

The final survey (after edits made based on pilot tests) is included in Appendix C of this document and also may be found at: https://redcap.iths.org/surveys/index.php?s=PAP7R3XFHE.

To conduct our study with human subjects, we also applied for, and received approval from the University of Washington Institutional Review Board. The IRB number is STUDY00003809.

2.1.3 Pilot testing of surveys

Prior to rolling out the survey to intended participants, we conducted pilot tests. The purpose of this was twofold. Firstly, we wanted to make sure there were no technical difficulties in accessing the links in the invitations sent out to prospective participants, and that both the site and the survey tool were functioning without errors. Secondly, we wanted to assess whether the survey was designed well enough to capture the information we wanted to collect. In short, we wanted to make sure that it was functional, reliable, user-friendly, intuitive, and responsive.

To this end we conducted two rounds of pilot tests. In the first round, four students and researchers from Dr. Brian Shirts' lab (including Dr. Shirts himself) were sent invitations and asked to go through the study. They were chosen as pilot study participants because they had a sufficient grounding in the site or the science to be able to test the site and the survey, but whose prior involvement with the project excluded them from the actual study. They gave us qualitative feedback about issues they had with the site or the survey tool, either by email, or in response to the open questions in the survey.

Feedback from this round of pilot testing led to some changes to the website: errors with the example file were resolved, and an additional tab was added to the site, explaining the data in the example file, so that participants would have some understanding of the background of the data in the example file. Other feedback about the survey led to the modification of instructions on the survey splash page for greater clarity, and the addition of extra questions to the

Background Information section on the survey to assess participants' level of experience, as well as their areas of specialization within genetics. In addition, based on the feedback, the survey questions were modified to be less vague, and the number of open questions was reduced to one per page.

In the first iteration of the second round of pilot testing, we reached out to Dr. Sean Mooney from the Biomedical Informatics and Medical Education department at the University of Washington, for permission to reach out to his students. While they were unable to participate in the pilot studies, we did get feedback from some students regarding doubts that they had after reading the invitation, about how data they uploaded in the study would be used and whether they needed to upload personal information. Based on this feedback, we made some modifications to the content of the email invitation to be sent out to participants, to clarify any doubts about data use as a result of participating in the study.

In the second iteration of the second round of pilot testing, we reached out to seven graduate students from the Biomedical Informatics and Medical Education department. Once again, they were sent email invitations and asked to go through the study and provide qualitative feedback about their experience and any issues they faced. In addition, we provided a list of questions for them to answer about their experience:

- How long did it take you to complete the survey?
- Were the instructions clear?
- What did you think of the length? Was it too long or short?
- What did you like about it?
- What did you dislike?

- Were there any questions where you felt that you needed more information or where you were confused?
- Do you have any other suggestions or feedback?

Based on the feedback from this round of testing, some more changes were made to improve the technical functionality of the site: error messages were modified to be more informative to participants so that they knew what had gone wrong, and a loading bar was added to let participants know that their upload had been successful and the calculation was in progress. The software parameters section on the results page was also formatted to be more readable and spaced out further away from the graphical results, to prevent confusion and clutter. Some more changes to the splash page and invitation were made, to improve clarity, readability and shorten the length of the text.

2.1.4 Survey participants and recruitment

Since the tool itself was developed for genetics experts, we chose to limit our sample for the study to three kinds of participants: genetic counselors and trainees, statistical geneticists and trainees, and undergraduate and graduate-level genetics students from various universities.

To recruit participants with characteristics similar to our target users, we looked for sources to find established genetic counselors and trainees. We identified accredited genetic counseling organizations like the National Society of Genetic Counsellors (NSGC) and the American Board of Genetic Counselors (ABGC), and obtained contact information for potential participants from the online public directories of these organizations. We used filters like board accreditation and areas of specialization to target users who were most likely to find this tool of use to their regular work. We recruited students and trainees via program directors of all the

accredited genetic counseling training programs across the country. The number of potential participants from each source is listed in table 4 below. Each of the sources was cross-verified to ensure no duplication of participants across different databases.

Table 4. Participants numbers, by source

Source	Participant Type	Number of Potential Participants
National Society of Genetic Counselors (NSGC)	Genetic Counselors	812
American Board of Genetic Counselors (ABGC)	Genetic Counselors	439
Accredited Genetic Counseling Training Programs	Counselors and Trainees	51 programs
Statistical and Computational Genomics Class (UW GS 559)	Students	25 students

Recruitment was done by e-mail invitations to prospective participants, which briefly explained the study and invited them to participate. The content of the email sent out to prospective participants can be found in appendix A of this document. On accessing the link in the email, participants were then directed to a consent page (appendix B) where the study was explained further, and participants were asked to consent before continuing. Links to the website and the survey itself were also included on the consent page. At the end of the survey, they were asked to leave their contact information if they were interested in participating in any future studies.

To motivate participants to take part in the study, we created a raffle draw for participants, with a $100 gift card to be awarded to a single winner at the end of data collection. Participants could enter themselves in the draw by entering their email address at the end of the survey. (Email addresses entered here were stored separately from the rest of the data, to preserve anonymity.)

2.1.5 Survey data analysis

Scaled questions were analyzed using descriptive statistics and visualizations of response distribution to assess what the population of participants felt about each aspect of the website (e.g. relevance, reliability etc.). Responses to the scaled questions were on a 5-point Likert-type scale of agreement (completely disagree, partially disagree, neutral, partially agree and completely agree). These scales were treated as ordinal rather than continuous, since the intervals between values/categories cannot be presumed equal [39] and many other criteria for considering these items to be continuous (such as assumptions about skewness and homogeneity of the population) are not present [40]. Therefore, the descriptive statistics used were medians, quartiles, frequency and percentage of responses for each category, for each construct. We intended to compare responses for the major participant groups, but we were unable to collect enough data from all the subsets of participants to allow separate analysis and comparison of each subset.

Open-ended responses in the surveys were qualitatively analyzed. Responses were downloaded from REDCap in MS Excel format (.xlsx), converted to MS Word format (.doc) and imported into ATLAS.ti. Themes were inductively coded in ATLAS.ti and analyzed using the constant comparative method [41].

Finally, all of these results were collated and written up in a form to be disseminated and used to inform refinement of the website itself.

2.2 Interview methodology

2.2.1 Interview procedures

From preliminary analysis of the survey data, we realized that more information about the context within which target users might need to utilize this tool was needed. Additionally, while participants had mentioned some interesting usability issues, more detail about these issues would have been beneficial for understanding how to redesign the tool to suit users. Thus, we decided to add semi-structured interviews to explore both these areas. The interviews each consisted of 4 parts (and the transcript is available in appendix F):

- **Introduction**: Participants were asked about their profession, level of experience and what their major area of expertise was.
- **Context of use**: Participants were asked for detail about their day-to-day tasks, the frequency with which they performed tasks like creating pedigrees, how often these tasks were performed, as well as their degree of familiarity and frequency with which they came across patients with variants of unknown significance, and their usual method of addressing these
- **Using the tool**: Participants then repeated the process of using the tool, as in the survey study, and were asked to think aloud during the process. Usability issues that came up were discussed and explored.

- **Wrap-up**: Participants were asked to sum up their experience and provide any last thoughts about the tool, immediately after using it.

After the interview structure was designed, an amendment to the initial IRB was sought and granted. Depending on the location of the participants, interviews were conducted in person, or remotely via Zoom Meetings. Interviews were audio or video recorded, according to participants' preferences, and transcribed afterwards. De-identification and transcription of interviews were done simultaneously. Transcripts and recordings were both stored on the same secure online server, but stored separately from each other. Transcripts were labeled with participant IDs, but a separate index of participant ID linked to participant names was also kept, which was stored with the interview recordings, away from the deidentified transcripts.

2.2.2 Interview participants and recruitment

Interview participants were genetic counselors and statistical geneticists. Six participants in total were recruited. Recruitment for interviews was done in two ways:

1. Survey participants who had indicated at the end of the survey that they would be interested in follow-up studies, were contacted via email.
2. To enrich representation of user groups for which participation in the survey study was small, we also directly reached out via email to prospective users using online public directories of organizations and institutions and contacts of the site developer (Dr. Shirts).

The content of the emails sent to both old and new participants may be found in appendix A and D of this document. Participants who indicated interest were then scheduled for in-person or remote interviews. Consent forms for interviews were either read and signed at the beginning of the interview (in the case of in-person interviews) or sent beforehand via email and signed and

sent back (in the case of remote interviews). The consent form signed by interview participants can also be found in appendix E of this document.

2.2.3 Interview data analysis

Transcripts of interviews were analyzed in ATLAS.ti. Open coding was done on interview transcripts, and using the constant comparative method [41], thematic categories of interest were developed.

Comparison was done in three stages [42]. First, comparisons were done within each interview to develop categories and label them appropriately, and to understand what the salient insights from each interview were. Next, comparison was done between interviews of subjects that belonged to the same group. Due to the limited number of interviews, grouping was done simply by categorization as statistical geneticists or genetic counselors, even though individuals within each group differed from each other in terms of experience and their day to day job description. The purpose of this round was to solidify the categories developed in the first round and develop them further. Finally, comparison was done between interviews from different groups. The aim of this final round was to establish and identify the differences between the experiences of each group and be able to describe them appropriately.

3 RESULTS

3.1 Survey results

3.1.1 Background and demographics

A total of 57 usable survey responses were collected. Table 5 below shows the number of respondents, split by occupation categories.

Table 5. Occupations of survey participants

Occupation	Number of responses	Number of respondents contacted
Genetic counselors and trainees	52	1,225
Statistical geneticists and trainees	2	109
Genetics students	1	21
Other	2	2

Only surveys which were completely filled out were included in the analysis. On review of partially-filled out survey responses, the maximum drop-off (38%) in response was observed to be where the participants were asked to complete the task of running the calculator. Since no identifiers were used in the survey, this 38% could have been due to respondents losing interest in the study and quitting at this point or could have been respondents who later came back and did fill out the survey (since the REDCap session is not saved or unique to participants, the same user opening the survey link multiple times would have been logged as multiple sessions).
A gender breakdown for each category is shown in table 6 below.

Table 6. Genders of survey participants

Occupation	**Male**	**Female**
Genetic counselors and trainees	4 (7.7%)	48 (92.3%)
Statistical geneticists and trainees	1 (50%)	1 (50%)
Genetics students	0 (0%)	1 (100%)
Other	1 (50%)	1 (50%)
Total	6 (10.5%)	51 (89.5%)

3.1.2 Genetics experience

On average, respondents had 10.39 ± 7.9 years of experience in genetics (including training).

The table below shows the breakdown of specialties of respondents.

Table 7. Specialties of survey participants

Specialty	**Count**
ART/Pre-implantation genetic disorders	3
Adult (including complex disease)	8
At-home/Direct-to-consumer DNA testing	1
Cancer	38
Cardiac	2
Fetal Therapy	0
Hematology	3
Metabolic	1
Molecular/Cytogenetics/Biochemical Tests	3

Specialty	Count
Nephrology	0
Neurogenetic	1
Ophthalmology	1
PGD/Pre-conception	1
Pediatric	13
Personalized Genomic Medicine	3
Prenatal	8
Psych disorders	1
Public health/newborn screening	1
Research	7
Screening (multiple marker)	0
Specialty disease	3
Teratogens	0

Respondents were allowed to specify multiple areas of specialty. The maximum number of respondents were those who listed cancer as one of their areas of specialty, which makes sense, as the tool currently only supports cosegregation analysis for eight cancer genes.

3.1.3 Calculator run success

Table 8 below shows the breakdown of the files the participants used to run the calculator. Most participants used only one type of file. Only one participant tried the calculator

with both the example file and a modified version of the example file. Five participants did not specify which type of file they used.

Table 8. Files used in survey by participants

File type	**Count**
Example pedigree	47
Example pedigree w/ modifications	3
Participant's own pedigree file	3
No response	5

The table below shows a summary of participants' success in running the calculator. While most of the participants were able to run the calculator (73.7%, 42/57), more than a quarter of the participants (26.3%, 15/57) were unable to do so. Of those who were able to run the calculator, three participants reported that the results did not seem correct.

Table 9. Calculator run success of survey participants

		Did the results seem correct, given the family history of the disease and/or the genetic testing results of the family?			
		Yes	**No**	**No response**	**Total**
Were you able to successfully run the calculator?	**Yes (73.7%)**	35	3	4	42
	No (26.3%)	1	4	10	15

Asked to provide feedback about why they were unable to run the calculator, six participants stated that they did not have a pedigree in the format required, and so were unable to run the calculator. Four participants tried to use the example file but received an error for a reason which was unclear to them. The remaining did not provide a reason.

Asked to provide feedback about why the results did not seem correct, the 4 who were unable to run the calculator provided the same reason (could not run calculator). Of the three who could run the calculator but did not think the results were correct, two of them thought the meiosis counting was incorrect, and one thought that the example file was erroneous, as it seemed like individuals 13 and 15 were affected, despite both their parents being unaffected. Relevant portions of both the tabular and visual formats of the example pedigree are shown in table 10 and figure 4. As can be seen, individuals 9 and 10 have four offspring, individuals 13-16, of which 13 and 15 are affected (affection status=2 in table 10). However, the parents, 9 and 10, are *not* both unaffected. While individual 9, the father (sex=1), is unaffected (affection status=1), the mother (individual 10) is not unaffected; her affection status is unknown (affection status=0). While this is clear upon referring to table 10, figure 4 does not clearly show a difference between unknown and unaffected affection status, possibly leading to the confusion that the participant had.

Table 10. Portion of tabular format of example pedigree

Individual	Father	Mother	Sex	Affection Status (0=unknown, 1=unaffected, 2=affected)	Alelle1	Alelle2
9	7	3	1	1	0	0
10	0	0	2	0	0	0
11	7	3	1	2	2	1
12	0	0	2	1	0	0
13	9	10	1	2	2	1
14	9	10	1	1	0	0
15	9	10	1	2	2	1
16	9	10	2	1	2	1

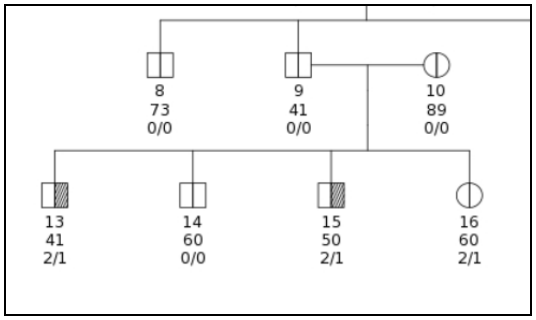

Figure 4. Portion of example visual pedigree. The labels for each individual represent ID, age and genotype (top to bottom). The right half of each square/circle represents affection status, with grey indicating someone affected with the condition.

Additionally, one participant also mentioned that it was unclear which cancers were counted towards the penetrance estimates of each gene, therefore making it impossible to assess the tool, and felt that a tool where one could adjust the penetrance values would be much more useful. For example, the ATM gene is associated with breast, pancreatic, colorectal and gastric cancers, as cited on the AnalyzeMyVariant penetrance information page for the ATM gene (figure 5). However, AnalyzeMyVariant only considers breast cancer in its penetrance model. This excludes the possibility of being able to choose to include these other cancers in the model.

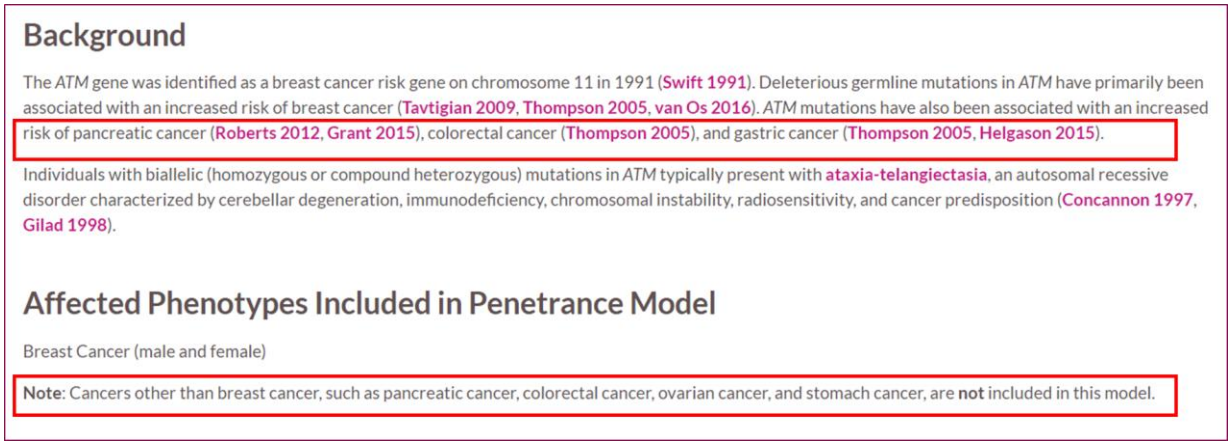

Figure 5. Penetrance information page for ATM gene. Text within red boxes show that the gene is associated with multiple cancers, but that the tool only considers breast cancer.

3.1.4 Survey responses (constructs)

There were four main categories of constructs: usability, quality, safety and satisfaction. Each of these four categories will now be discussed in greater detail.

3.1.4.1 Usability

The first section of the survey, usability, consisted of four scale items: two items of ease of use, information completeness and output relevance. A total of 55 respondents provided

responses to the scaled questions in this category. As can be seen from table 11 below, the median ratings for all four constructs in this section were ‘neutral’ (3 being neutral), and the upper and lower quartiles were in the ‘partially agree’ and ‘partially disagree’ categories respectively. The counts of responses in each category is shown in table 12, and the distribution of results that contributed to this can be seen in figure 6, where the highest peaks in each can be clearly seen. In the cases of information completeness and website ease of use, the highest number of responses fell into the ‘partially agree’ category, albeit by small margins. For output relevance, the most frequently selected response was ‘neutral’, and for results ease of use, the highest number of responses disagreed with the statement that were the results were clear and easy to understand. Explanations for these scores may be found in the open responses where participants provided feedback.

Table 11. Description of usability scaled responses (n=55)

Constructs	Median	Upper Quartile	Lower Quartile
Information completeness	3	2	4
Ease of use (website)	3	2	4
Ease of use (results)	3	2	4
Output relevance	3	2	4

Table 12. Response counts for all constructs of usability (n=55)

Questions	**1 = Completely disagree**	**2 = Partially disagree**	**3 = Neutral**	**4 = Partially agree**	**5 = Completely agree**
Information completeness (I found the information provided on the website to be thorough and complete.)	1 (1.8%)	8 (14.5%)	18 (32.7%)	20 (36.4%)	8 (14.5%)
Ease of use (website) (I found the website clear and easy to use.)	4 (7.3%)	11 (20.0%)	12 (21.8%)	18 (32.7%)	10 (18.2%)
Ease of use (results) (I found the results clear and easy to understand.)	8 (15.1%)	19 (35.8%)	10 (18.9%)	10 (18.9%)	6 (11.3%)
Output relevance (I found the output relevant.)	3 (5.7%)	8 (15.1%)	21 (39.6%)	17 (32.1%)	4 (7.5%)

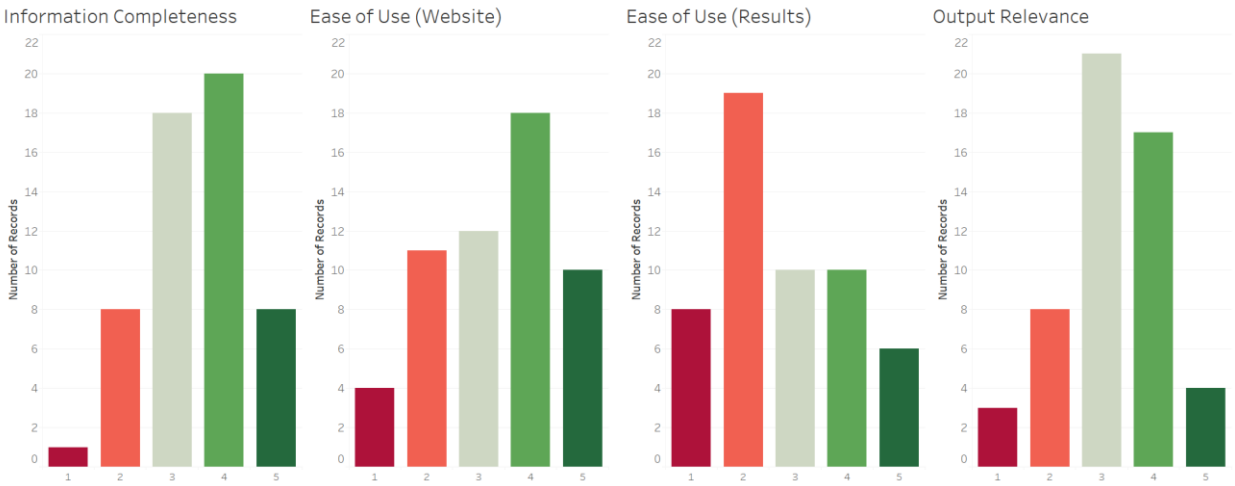

Figure 6. Usability responses

Of the 34 respondents who provided additional feedback about usability, 16 of them expressed confusion about how to interpret the results, as they were unsure what scales the numbers were on, and how to interpret whether the results were clinically significant or not. While the majority wanted reference values of some sort to put the results into context, two participants (both statistical geneticists) stated they were more comfortable with LOD scores, and would prefer that as a result format.

10 participants also stated that the pedigree was 'prohibitively difficult' to make, as they often did not understand intuitively how to do it, and it took them a long time to figure it out. This might be why the majority of participants (as seen in table 8) just used the example file. Participants also stated that, given how difficult it was to make the pedigree file, they were unlikely to want to use this tool, as they didn't have time to do this for all the patients they wanted to analyze.

Other usability feedback included one participant suggesting the ability to look at models of inheritance apart from dominant (currently the tool assumes a dominant mode of inheritance

in its calculation), one participant suggesting that the output below the figure on the results page would be better hidden to those who didn't want to access it explicitly, and two participants who expressed an interest in using this tool for other genes other than the eight defined. These participants wished to be able to define the penetrance classes and values themselves for use with other genes. Lastly, three participants also did mention that they appreciated the clean, user friendly layout of the website.

3.1.4.2 Quality

In the second category, quality, there were four items: accuracy, efficiency, reliability and need fulfillment. A total of 54 respondents provided responses to the scaled questions in this category. As can be seen from table 13, once again the median ratings for all four constructs in this section were neutral, with the quartiles being in the partially agree/disagree categories. However, from the counts in table 14 and the distributions seen in figure 7 below, it is clear that in the need fulfillment construct, as well as the accuracy and reliability constructs, the overwhelming majority marked themselves neutral, more than 50% of participants in the case of accuracy and reliability, and 48% of participants in the case of need fulfillment.

Table 13.Description of quality scaled responses (n=54)

Constructs	Median	Upper Quartile	Lower Quartile
Accuracy	3	2	4
Efficiency	3	2	4
Reliability	3	2	4
Need fulfillment	3	2	4

Table 14. Response counts for all constructs of quality (n=54)

Questions	**1 = Completely disagree**	**2 = Partially disagree**	**3 = Neutral**	**4 = Partially agree**	**5 = Completely agree**
Accuracy (I found the tool to be accurate.)	2 (3.7%)	1 (1.9%)	32 (59.3%)	13 (24.1%)	6 (11.1%)
Efficiency (I found the tool to be efficient.)	2 (3.7%)	5 (9.3%)	16 (29.6%)	19 (35.2%)	12 (22.2%)
Reliability (I found the tool to be reliable.)	2 (3.7%)	1 (1.9%)	36 (66.7%)	9 (16.7%)	6 (11.1%)
Need fulfillment (The tool told me exactly what I needed to know.)	5 (9.3%)	10 (18.5%)	26 (48.1%)	9 (16.7%)	4 (7.4%)

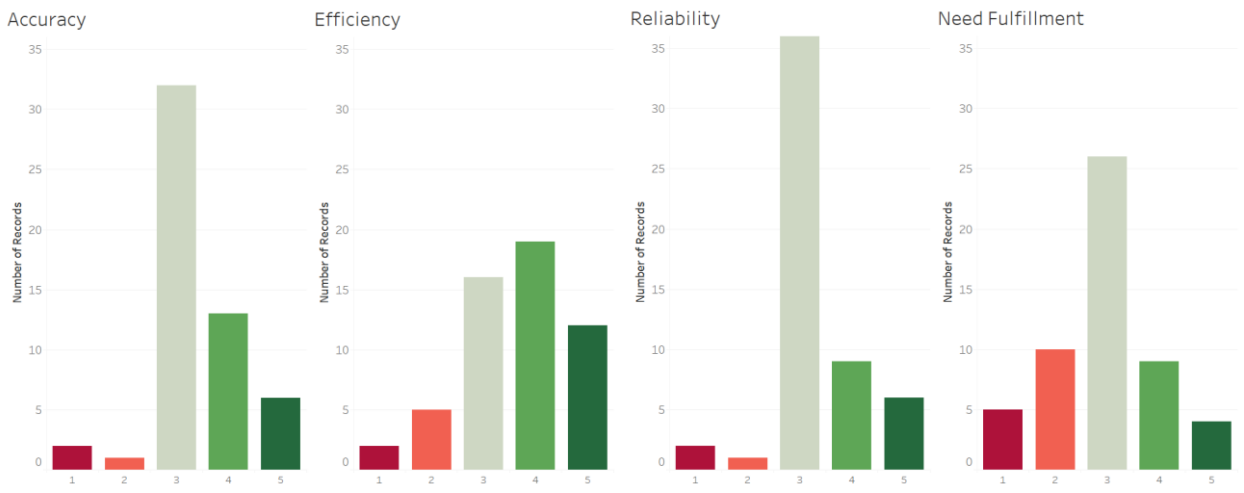

Figure 7. Quality responses

Of the 22 participants who provided qualitative feedback in this section of the survey, the most frequently occurring comment (n=11) was that participants did not understand how to evaluate the constructs, as they did not understand how to interpret the results. Another participant said that not understanding what information was used to put together the penetrance estimates made them wary to use the results anywhere. This notion that they were unable to really interpret the constructs here probably explains why so many responses to the questions were marked as neutral.

Other feedback included three participants again mentioning how difficult the pedigree was to create and how that would prevent them from finding the tool helpful, and one participant bringing up the issue of how some of the genes available for analysis in the tool are actually related to multiple cancers, but the tool only considers a singular affection status for a single condition.

3.1.4.3 Safety

In the third category of the survey, safety was measured with two items: privacy and security. A total of 54 respondents provided responses to the scaled questions in this category. The median responses of both constructs in this section lay in the neutral region (table 15). The counts in table 16 and the graphs in figure 8 show that the majority of responses (exactly 50% in both cases) were neutral. However, 24% and 30% of respondents also completely agreed with both statements of privacy and security, respectively (the exact phrasing of these statements can be seen in table 11).

Table 15. Description of safety scaled responses (n=54)

Constructs	Median	Upper Quartile	Lower Quartile
Privacy	3	2	4
Security	3	2	4

Table 16. Response counts for both constructs of safety (n=54)

Questions	**1 = Completely disagree**	**2 = Partially disagree**	**3 = Neutral**	**4 = Partially agree**	**5 = Completely agree**
Privacy (The website respects my right to control access to the personal information that I uploaded.)	1 (1.9%)	4 (7.4%)	27 (50.0%)	6 (11.1%)	16 (29.6%)

Questions	**1 = Completely disagree**	**2 = Partially disagree**	**3 = Neutral**	**4 = Partially agree**	**5 = Completely agree**
Security (I feel that this website is secure.)	2 (3.7%)	5 (9.3%)	27 (50.0%)	7 (13.0%)	13 (24.1%)

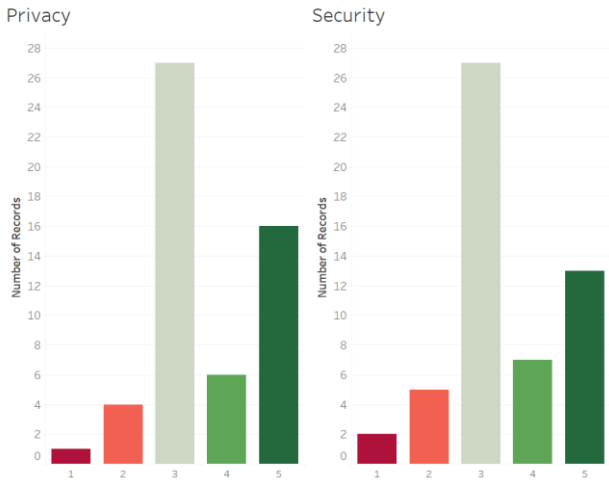

Figure 8. Safety responses

In the 20 feedback responses given by participants about privacy and security, 10 participants said that they had no idea if the site was private or secure, as no overt statements of either were provided anywhere on the site. Three other participants suggested that they would like to see an ‘https’ domain being used for the website, and one participant said that since there were no guarantees or consumer statements of satisfaction listed, they would never trust this website with their data. Additionally, four participants mentioned that it was unclear whether files being uploaded were being retained by the website, as there were no statements of data use, or a way to ‘delete’ the uploaded pedigree.

However, on the other hand, one participant did mention that he would be likely to trust it because it was a University of Washington development. Another mentioned that they would not upload personal health information of clients/subjects as a matter of principle anyway, but that it didn't matter as much, because the pedigree did not require any personal information be uploaded.

3.1.4.4 Satisfaction

The last category was satisfaction, which included only one item: satisfaction. A total of 55 respondents provided responses to the scaled question in this category. For the scaled responses, once again the majority lay in the neutral category (table 18 and figure 9), and the distribution around neutral was fairly even on both sides, with the median response being neutral (table 17).

Table 17. Description of satisfaction scaled responses (n=55)

Construct	Median	Upper Quartile	Lower Quartile
Satisfaction	3	2	4

Table 18. Response counts for satisfaction (n=55)

Questions	**1 = Completely disagree**	**2 = Partially disagree**	**3 = Neutral**	**4 = Partially agree**	**5 = Completely agree**
Satisfaction (I was satisfied with my experience with the tool.)	6 (10.9%)	9 (16.4%)	19 (34.5%)	15 (27.3%)	6 (10.9%)

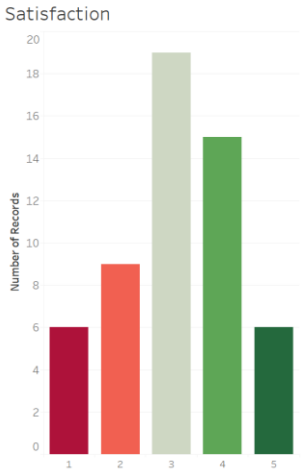

Figure 9. Satisfaction responses

Unlike in the other sections of the survey, there were three open-ended questions in this last section, asking what participants liked least, what they liked most, and what could be added to improve their experience.

What participants liked least:

Of the 33 responses to the question about what they liked least, 15 mentioned how the results were difficult to interpret and understand. One additionally mentioned the lack of clarity about how penetrance estimates were derived.

11 participants reiterated the difficulty they faced in creating the pedigree in the format required, saying it was either hard to understand the instructions or too time consuming to convert from the pedigree formats they already had. Some additionally expressed confusion about aspects of pedigree-making; how to add siblings, how to input mother and father details, and how to add details of individuals for whom only a single parent is known. In using the example file, they also said that having only one example with such a detailed and extensive

family history was less helpful, as it was unlikely to reflect real pedigrees participants might have, which would have fewer individuals and less information known.

One participant mentioned not liking the slow load time of the results, and another mentioned the lack of privacy and security guarantees. Yet another participant wanted the tool to have greater generalizability for more genes or modes of inheritance. Lastly, one participant did not like the image on the main page, as they felt as though they were being 'observed' by the young boy as they uploaded family histories.

What participants liked most:

Of the 35 responses to the question about what participants liked the most, 12 participants said they liked how simple and easy to use the website was, with very few steps required to get results. Three additionally felt that the instructions for uploading and running the tool were quite clear for something that seemed quite complex. 11 participants praised the visual design of the site, saying it looked professional, clean and pleasant to look at. Four participants liked how fast the calculator was in computing results. Two participants said they liked the concept of the website as a novel tool that they would like to use, but that the time taken to format pedigrees would probably prevent them from being able to actually use the tool in their practice.

What could be changed/added:

Finally, 26 participants answered the question about what could be changed/added to improved their experience. 13 participants said they would like to have more information about results, either in the form of explanations, reference ranges, or even different formats, such as a LOD score or as an odds ratio for/against the variant segregating with disease. One participant also suggested that the interpretation of the example pedigree could be provided, to allow users to understand, by example, how to interpret their own uploaded pedigree.

4 participants said they would like to be able to upload data directly from popular pedigree-making tools, such as Progeny, rather than having to create a text or Excel file especially for this tool.

Three participants stated that they would like the additional functionality of being able to manipulate the penetrance estimates being used, so that they could use this tool with other genes.

Other suggestions included: a way to 'save' the loaded results so users could switch back and forth between pages in the website without having to re-run the calculator, a way to export or download results, better error coding so that users could understand why their uploaded files didn't work, and potential reformatting of the 0s, 1s and 2s in the pedigree to '?', '+' and '-', or other more intuitive symbols. One participant also mentioned that some extra information about each available gene might be useful, as for example in the case of ATM, aside from cancer, ATM mutations might have the AT phenotype in the family, or bleeding problems.

3.2 Interview results

Qualitative, semi-structured, one-on-one interviews were conducted with six genetics experts: four genetic counselors (IDs 2, 3, 4 and 5) and two statistical geneticists (IDs 1 and 6). Three of these six participants (IDs 1, 2 and 3) had previously participated in the survey study, while for the other three, this interview was their first exposure to the tool.

The results from these two sets of interviews are described separately, as the feedback from each differs so widely. Within each section, feedback is further divided into four thematic categories: context of use, pedigree creation, interpretation and use of results, and general feedback about the website.

3.2.1 Genetic counselors

3.2.1.1 Context of use

The four genetic counselors interviewed came from different institutions, and as such, the day to day work they performed varied. Two of them worked within large healthcare institutions and performed solely clinical tasks in their day-to-day work (IDs 4 and 5). One (ID 3), although a clinical genetics counselor, worked as part of a smaller tele-counseling clinic. All worked in cancer genetics, and a large proportion of their patients were breast cancer patients. The last one (ID 2) worked as a genetic counseling researcher attached to a large institution.

In the cases of the clinical counselors, most of their workdays were spent in consultations with patients, each lasting approximately an hour. Tasks performed within these consultations included taking family histories from patients (if the details had not been filled out by patients prior to the consultation) to make pedigrees, discussing genetic testing results, and running some cancer risk models (if they took no more than a few seconds to run and interpret). Longer analysis tasks that required more concentration were performed in the downtime between and after consultations. At these times, they would get and update pedigrees based on testing results, run bigger and more complex cancer risk models, look up databases for information about variant pathogenicity and compare results from different labs (most of these results being in the form of multi-gene panels). The researcher, on the other hand, spent most of her days performing research tasks such as organizing materials for exome sequencing studies, screening and talking to participants, interpreting variants and returning results.

Asked about how often they saw patients with variants of unknown significance, their responses ranged from sometimes to fairly often. It was even more likely to see uncertain variants these days, they said, because most genetic testing was done in the form of multi-gene

panels, so patients often ended up information about many more genes than they had gone to test for. However, when it came to having patients with variants of unknown significance, where at least some family members had also been tested, all said that this would happen very rarely. Pushed to give an estimate of how often this might happen, ID 5 said "maybe…*maybe* once a month". Others simply said it was downright unlikely. Variants of unknown significance (VUS) were simply too common and too vague to even be used as a basis for encouraging family members to be tested. Their general modes of treating VUS involved multiple processes of trying to ascertain pathogenicity:

> "Going, looking at the report in detail, looking at the family history to see what makes sense. Looking at the medical history of the patient, it may have been a patient that I ordered the test on, or that's coming to me for interpretation. Going in some of the databases to see if I can get some more information." – ID 4

Based on these, they would then give qualitative recommendations to patients about how treat the VUS, going forward. In most cases, treatment recommendations generally didn't change at all.

On the rare chance they did see individuals with unknown variants whose families had also been tested, their procedure was still to look at a number of different sources to come to a conclusion about how to make clinical recommendations. This would involve looking up the variant in databases like ClinVar, looking at functional studies, amino acid differences, running cancer risk models, and checking up with the labs where testing had been done to ask about how they had classified that variant, particularly if different labs had classified the same variant in different ways. When it came to potentially using this tool, they said they would probably run this along with other cancer risk models in the downtime between consultations.

> "I probably would do it as part of the prepping session so I could look up information on ClinVar, I might look up information on the amino acid difference, what the lab says and then I might run this to see what else is this going to tell me to help me give a better risk assessment for the patient." – ID 3

The goal of all of these tasks was simply to give an idea to counselors of where on the spectrum of pathogenicity the variant may lie. Based on findings from AnalyzeMyVariant and other sources of information, they would then decide whether to either leave clinical treatment unchanged, encourage their patients to get more family members tested, or participate in studies to get more data about the variant. In a lot of cases, insurance was a factor as well. Unless they had sufficient evidence that an individual or their family members needed to be tested, such expensive tests would not be covered by insurance.

3.2.1.2 Pedigree creation

As mentioned briefly above, pedigrees used by genetic counselors were usually created before patient consultations by programs such as Progeny. Patients were prompted by this program to enter various pieces of information about themselves and their family, and the program would collect that information and use it to render a visual pedigree. Later, during the consultation, the counselors might then verbally collect more information from patients and edit the pedigree in Progeny. The pedigrees they worked with were always in a visual graphical form. In making a pedigree from scratch, they would start with the consulting patient, and then radiate outwards along the family tree from there, asking about different relatives, their age, health status and genetic data, if available.

> "If I was taking up family history of you, then everything would be based on you. And you obviously have a mother and father, so really everything is based on the three of you. I would put whether your father and mother are alive, how old they are. How old you are. Do you have children. Then I would ask about your brothers and sisters. How old they are and if they have children. Then I would ask about your mother's brothers and sisters and her parents and if they're living or deceased, what their status is. Then I would ask about your father's relatives, about his brothers and sisters and their parents. And ask about the health status for breast cancer or colon cancer or whatever." – ID 4

Because the pedigrees they were used to working with were completely visual, they all found the process of making the pedigree for AnalyzeMyVariant incredibly non-intuitive. Directed to the site and to the instructions page, they didn't even know how to start the process of creating the pedigree. And when they did open an Excel file, it took them several minutes and several questions to understand how to add individuals. The most difficult part of adding people to pedigree file were the first 3 columns of the individual, mother and father IDs. They did not understand how the process of assigning individual IDs worked, or whether it was based on some hierarchy (for example, ID 4 thought the order might be assigned according to which generation and individual was in in the pedigree). The example pedigree shown below in figure 10 illustrates the confusion.

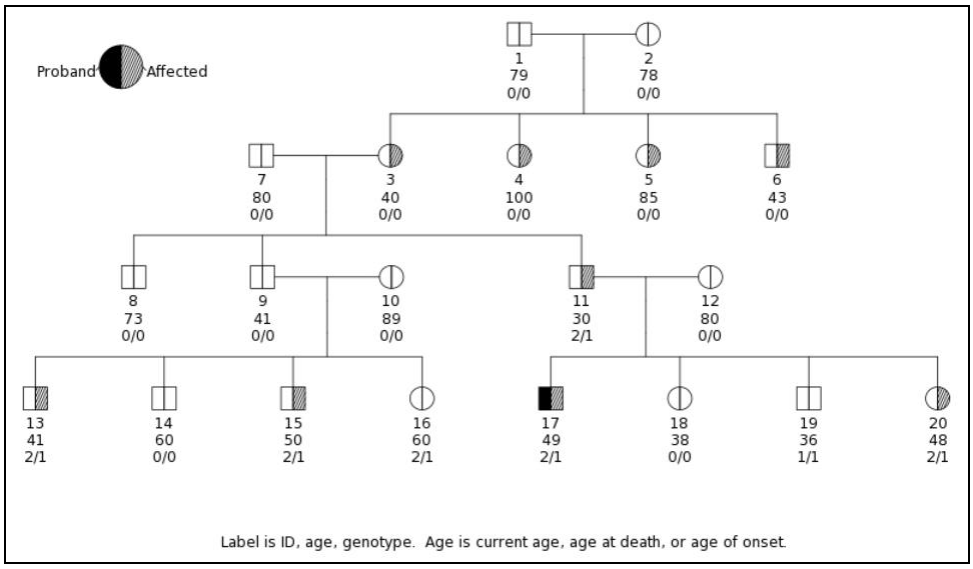

Figure 10. Visual representation of the example pedigree. The top-most number below each individual is the ID. ID assignment in the example pedigree starts from the top of the tree and continues linearly along subsequent generations. By contrast, genetic counselors usually start drawing the pedigree from the proband (in this case individual 17) and move radially outward.

With regards to parents, ID 3 thought the mother and father IDs were assigned based on whether individuals were on the proband's maternal or paternal sides. But while these confusions could be clarified by re-reading the instructions, the biggest confusion they faced was in assigning a "0" for mother and father IDs of individuals at the top of the tree. AnalyzeMyVariant simply uses "0" as unknown, and therefore, for individuals at the top of the tree whose ancestry is unknown, the fields of mother and father are to be assigned with this code. ID 5 thought that "0" was to be assigned to the proband or the proband's parents, and so was consequently confused about how to add their parentage. They also did not know how to add siblings to a pedigree and did not understand how the pedigree used mother and father ID to identify siblings. The other big problem faced was in adding individuals for whom only one parent was known, as often happened with patients. Since AnalyzeMyVariant allows only either both parents to be known, or both parents to be unknown, it was impossible for them to add children of family

members for whom only details about a single parent was known. Other columns in the pedigree also created confusion. The difference between unknown and unaffected affection status was one that troubled ID 4 a lot.

> "Unaffected is a very strong word to use - to me that means that you have asked them and they've been tested and you know for sure that they're not affected. Unknown could mean any number of things, it could be that you haven't asked, or there's no way of finding out. And with different cancers it could be any number of things. With colon cancer I would want to know if you've had colon polyps. Have you had a colonoscopy. So a lot of things would fall into the bucket of unknown." – ID 4

Overall, they all found the process difficult to follow. Of all the counselors, the one who found it easiest to make the pedigree was ID 2. Not only had she filled out the survey (and so had seen the site) before, she also filled out the pedigree by taking a printout of a visual pedigree from one of her cases, and writing down individual ID numbers next to each person on the pedigree, and thereafter used the sheet as a reference to make the pedigree. However, she as well as the others, found it infinitely more difficult than the visual pedigree they were used to making, which took them a minute or two to make. The AnalyzeMyVariant pedigree, on the other hand, took them all about 10-15 minutes to make. Figure 11 shows both types of pedigrees and illustrates how it might be non-intuitive for a genetic counselor to understand and follow the relations between family members purely from the table.

Individual	Father	Mother	Sex	Affection	Age	Alelle1	Alelle2	Proband
1	0	0	1	0	79	0	0	0
2	0	0	2	0	78	0	0	0
3	1	2	2	2	40	0	0	0
4	1	2	2	2	100	0	0	0
5	1	2	2	2	85	0	0	0
6	1	2	1	2	43	0	0	0
7	0	0	1	1	80	0	0	0
8	7	3	1	1	73	0	0	0
9	7	3	1	1	41	0	0	0
10	0	0	2	0	89	0	0	0
11	7	3	1	2	30	2	1	0
12	0	0	2	1	80	0	0	0
13	9	10	1	2	41	2	1	0
14	9	10	1	1	60	0	0	0
15	9	10	1	2	50	2	1	0
16	9	10	2	1	60	2	1	0
17	11	12	1	2	49	2	1	1
18	11	12	2	1	38	0	0	0
19	11	12	1	1	36	1	1	0
20	11	12	2	2	48	2	1	0

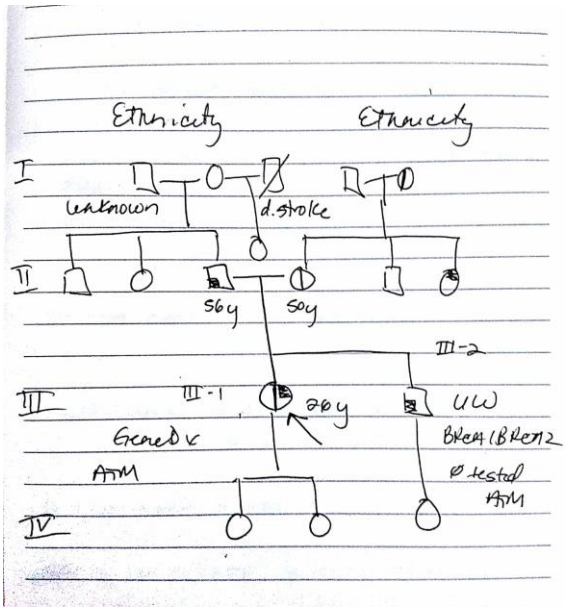

Figure 11. Tabular pedigree (left) as required by AnalyzeMyVariant *and graphical pedigree (right) as hand-drawn by one of the interviewees*

3.2.1.3 Interpretation of results

After counselors finally managed to create a simple pedigree for use with the website, they were then directed to plug it in and look at the results. None of them were able to understand the results at all. Looking at the Full Likelihood Bayes factor, all said they had learnt about Bayes at some point during their schooling and training, but that they hadn't used it for years. They couldn't even begin to understand what the scale of the values were. All of them were completely unfamiliar with the Cosegregation Likelihood Ratio and didn't even have the surface familiarity with the name that they had with the Bayes factor. With regard to the number of meioses, while ID 2 seemed to understand what it meant, enough to know that a result of 1 gave her very little information, ID 4 could not understand what the number of meioses referred to at all.

"The separating meiosis is 6 between who and who? Who is this in reference to?" – ID 4

The output below the graph also went over their heads. None knew what Fastlink was, let alone what the values meant. The only part of the results page they could understand was the graph itself, which rendered a visual version of the pedigree they had entered. Some problems were encountered here as well. It took them some time to realize what the numbers under each individual meant, and all of them missed the sentence below the graph that functioned as a legend. Additionally, in some cases where the pedigree tree wasn't centered, the pictorial legend was rendered on top of the pedigree, obscuring the topmost individual and also making it hard to read the legend (figure 12).

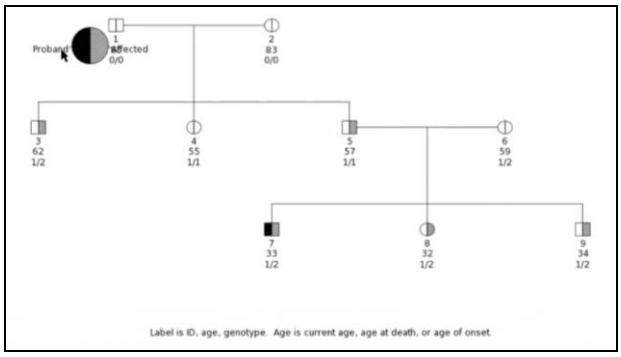

Figure 12. Example instance where the legend is rendered on top of part of the pedigree

The line running through each individual also confused ID 5. She thought it made sense for the proband, whose one side was colored black to identify them as the proband but couldn't understand why all the others had it too. Asked about what could be added to make the output more understandable and usable by them, their suggestions varied. All felt that the factors and ratios being provided needed to be defined, either on the results page itself, or by hyperlinking the result statement to direct users to a simple definition. Additionally, all felt that some kind of context would be needed to make the results actionable. They wanted to know how they could tie

the results back to the family in front of them, as their entire purpose of using the tool was so that it could help them make a decision.

> "What I want to know is, what does it tell me about my patients. Or even a graph to show me whether this is leading me towards being more or less concerned about whether the variant is co-segregating with the disease." – ID 3

Some did take the opportunity to look at the Background and Penetrance Classes information pages but found them uninformative. When ID 2 tried to look at those and then come back to her results tab, she found that she couldn't, and had to re-upload and run the file. This was something she said would be rather frustrating when doing analysis.

How this context could be given differed between participants. IDs 2 and 3 were more familiar with other formats of data like odds ratios or risk penetrance percentages, but were willing to use these figures, provided they could understand how. ID 5 further said that she didn't necessarily need clinical recommendations, as those might change over time, but at least she'd need some idea of the scale of results. ID 4 wanted to be able to put the results in context, not just in terms of the scale of possible results and pathogenicity, but also in terms of the quality of data she entered. She said she would like to see multiple examples of pedigrees with different qualities of data, and the results for each, so that she could understand how good her data needed to be, to get a truly accurate value. Furthermore, she wanted these examples to be specific to each gene and condition, since there was so much variation between conditions.

> "Because this tool is just one piece of the information about a variant, how it cosegregates in that family. And it's only as good as the information I put into it. So, I guess somehow I'd want to know, how the information could be better." – ID 4

In all cases, they simply wanted to understand how to know what to tell patients.

> “I probably would explain [to patients] that there is this tool that helps us to try to figure out a little bit more information. Because patients need layman terms, they want to know, what is this telling them. Actionable information. They get lost in statistics. So if I'm telling them that this variant is likely to be associated with a pathogenic presentation in a family, then I would say, here is what's making me think that way. I may or may not give them a number. I guess it depends on the situation, how savvy the patient is.” – ID 3

3.2.1.4 Other feedback about the website

Lastly, participants explored other pages on the website, and asked if they had any feedback about these parts of the website, such as the background and penetrance information pages. Other miscellaneous concerns also came up during the sessions with the genetic counselors. In some places, they found instructions unclear or misleading. For example, while the example file itself was provided in both Excel and text formats, the file formatting instructions only mentioned text formatted files as being allowed. Errors were not always informative, in particular when uploaded files were misnamed (the website did not allow files with spaces in the name). ID 2 also mentioned that it was important to mention somewhere that the tool assumes a dominant mode of inheritance, something she found was not clarified on the site. Genetic counselors also did not find the Background page very useful, as they said they would never have the time to go and read any of the cited papers.

> “I'm not going to go and look at any of these. If I'm going to use this tool it's going to be because it makes my work less, not more. I'm already doing a lot of work.” – ID 4

Additionally, as ID 2 mentioned, not everyone who wanted to use this tool would have open access to all the journals and papers cited. The information would therefore be completely unavailable to them.

3.2.2 Statistical geneticists

3.2.2.1 Context of use

The two statistical geneticists interviewed (IDs 1 and 6) both worked within large universities. The context within which they expected to use the tool was therefore in the research studies they ran, as opposed to use in clinical treatment. ID 6 worked mainly in association studies, using SNP array data. He also used family data in the context of linkage analysis and expected to use cosegregation analysis as part of his linkage analysis studies. His usual method of analysis would be to look at data collected from the families (usually about a few dozen per project), count up those who had the trait and those who didn't, and run statistical analyses to get an idea of linkage between the genetic trait and the condition. However, his area of expertise wasn't in cancer genetics and, in fact, he generally worked mostly with glaucoma and other eye-related conditions. ID 1, by contrast, did work with cancer studies, and was currently working on studies to identify genes that might be involved in cancers.

In both participants' cases, once again, they did come across variants of unknown significance fairly often but did not usually collect extensive information about families. Even the families specifically recruited for studies tended to be nuclear families, with families having more than two generations appearing only rarely.

Both statistical geneticists also did not directly interact with patients, and were a few steps removed from the actual data collection process from study participants. As a result, the

data they finally got and used in their analysis was in the form of tabular data, usually with identifiers already removed. The format of this data tended to be text tab-delimited files, with columns of IDs, father and mother IDs, sex, affection status and genotype information (very similar to AnalyzeMyVariant's pedigree format). The analyses tended to be quantitative, and on larger scales than a single family. Data from individuals and families would be aggregated to run statistical analyses with. Commonly used calculations in both cases were typically in the form of generating LOD scores, and neither used Full Likelihood Bayes or Cosegregation Likelihood Ratio very often. Furthermore, the purposes of these studies operated on a higher level than clinical treatment and recommendation for single families.

3.2.2.2 Pedigree creation

Statistical geneticists were very comfortable with the instructions and formatting requirements for the AnalyzeMyVariant pedigree. They seemed to instinctively understand how to code the information for each category, and, in fact, didn't really have much to code. In both cases, the data they used to make the pedigree was already formatted in much the same way, and all that was required was some transposition of data.

The only real effort required to make the pedigree was the amalgamation of data that was often stored in multiple files. For example, individual, mother and father IDs, along with sex, were usually stored in one file. Details about phenotype and genotype might then be stored in other files. To make the pedigree, therefore, they would have to siphon information from multiple files into the pedigree file to run on AnalyzeMyVariant. However, neither anticipated much difficulty with doing this. In fact, ID 1 said she'd probably just write a PERL script to transpose the data from the files into the pedigree.

> "So I have the individual, dad ID, mom ID in one file, maybe the sex in one file and the affection status in another one. And then marker data in another file. I would just write a PERL script to grab all the data from all these files and plug it in. So the longest time would be to just write that PERL script. I could create the PERL Script in, maybe, 20 minutes? And then I might modify it and update it and make it better. And then maybe make it more versatile to so I can use it with multiple files so I don't have to go in and change the whole thing every time." – ID 1

The only thing that minorly confused them was the header row that the pedigree required, as that was not standardly expected in the field to be needed for statistical analysis (according to ID 1).

3.2.2.3 Interpretation of results

After creating a simple pedigree, participants then plugged it into the website and ran the calculator. While the results and figure given on the results page weren't entirely foreign to them, they did have some trouble understanding them. ID 1 said she was much more comfortable with LOD scores, and usually did such analysis with only such scores. While she was vaguely familiar with Full Likelihood Bayes (FLB) and Cosegregation Likelihood Ratios (CSLR), she couldn't properly recollect more than a hazy idea of what the score meant, and said she'd have to go back and look it up, and actually wanted more information about the definition of each, as well some basic information about the scale of the values. ID 6 also had come across Bayes before, but was unsure about whether or not the value shown was the log version of the Bayes factor or not, something that was not clarified on the page. He said he would be more likely to use CSLR, if CSLR was the same as a linkage likelihood ratio (another thing he was unsure about, and the page did not offer instant clarification for). He also stated he'd be most

comfortable with a LOD score. Both said they'd need to spend more time looking up the results to understand exactly what they were and how they were calculated, and perhaps understand how to convert it to other formats to suit their needs, such as LOD scores.

ID 1 also raised another doubt she'd had about how the scores were calculated, based on some previous experimentation she'd done with the tool (when she had participated in the survey). When she had deliberately left the proband unspecified in the pedigree and run the calculator, the tool had provided results for FLB and meioses counting, but not for the CSLR, which seemed incorrect to her, as the FLB should not have provided a result either. Moreover, she also felt that the calculation for counting meioses was incorrect, as according to her calculations, the meioses counting should take the proband into consideration, but didn't seem to, as in both cases, the exact same value was given. ID 6, having not taken part in the survey, hadn't had a chance to experiment with the tool, but the pedigree he uploaded failed to give an output for either CSLR or meioses counting, the webpage instead showing the FLB value and unintelligible code for both CSLR and meioses counting.

Both participants liked the visual rendering of the pedigree, and ID 6 thought it was a 'neat' feature if he could get a visualization for free along with the calculation. However, both had issues with the line down the middle of individuals on the pedigree. According to ID 6, this was potentially misleading, as it might lead users to think that two separate traits were being represented in the pedigree, with half of each square/circle being used to show each trait. Both said that typically, the accepted practice with visual pedigrees was to use whole circles colored in, with a small arrow underneath the proband to indicate who the proband was. Additionally, ID 1 pointed out that the pedigree also did not visually distinguish between unknown and unaffected

phenotypes, with both being colored white, and suggested amending the coloring scheme to do this.

> "So in this, one individual I gave an unknown affection status, that's not being shown on this pedigree. There needs to be a way to show that this person is not affected and we know they're not affected, and that other person we don't know. There needs to be a coloring scheme to do that. What we have done in the past is white for unknown and gray for unaffected. When you're working with multiple traits it can become more complicated but since we're working here with only a single trait it can be simplified. So the two affected in this pedigree could be colored in completely black and then just have an arrow pointing to the proband. Then the unknown individual would be white and the unaffected individual would be Gray." – ID 1

In the current version of the pedigree, the proband was represented by half the circle being colored black, affected individuals being colored half grey, and everything else being white. Apart from this, they had no trouble understanding anything else in the pedigree, although they both also faced the issue of the legend being rendered on top of the pedigree.

Both statistical geneticists were also able to understand and follow the output below the graph, although they did find the formatting of it a little confusing. They could not immediately understand what each of the 3 numbers on each line referred to, and only upon looking at the tables on the Penetrance Classes pages did they understand what the order of the values were. They suggested formatting it into a table. ID 1 also said she would additionally like to see the penetrance curves displayed, rather than just having the data in a table. ID 6, on the other hand, expressed interest in being able to define the penetrance classes and values himself, as he worked primarily with glaucoma, and would like to use this tool with such conditions as well, not just for

cancer. He also said that since he often aggregates values and results across many families in a study, the ability to create a master pedigree with multiple families and analyzing them all together would be a useful one. Both also expressed interested in having a downloadable version of the penetrance information output to save somewhere, along with the results and visual pedigree.

3.2.2.4 Other feedback about the website

After running the calculator, both statistical geneticists also took the time to explore some of the other pages on the website. Miscellaneous concerns mentioned included: the lack of a clear clarification that the mode of inheritance being used was only dominant inheritance (IDs 1 and 6), the lack of a statement that the model assumes that homozygotes for the uncertain variant don't exist (ID 1), and the need for more description about how penetrance classes were derived (ID 1). With regards to penetrance classes, ID 1 also found that key citations about the penetrance classes were missing on the individual penetrance pages, particularly citations for where the homozygous normal values were taken from, something she felt should be rectified immediately. Otherwise, both participants liked the clean, neat look of the website, and liked that it was simple and focused on one task.

4 DISCUSSION AND CONCLUSIONS

This study aimed to assess the usability perceptions of two kinds of genetic experts, with regards to one highly specialized cosegregation analysis tool, AnalyzeMyVariant. After developing and deploying a survey to assess various constructs around usability, we further conducted interviews with users to gain more in-depth feedback. In discussing the results from this study, we will first look at the results from each user group separately. Then we will discuss the commonalities among the two groups, as well as the differences and how they might be addressed.

4.1 Genetic counselors' perceptions of usability

Several usability issues were identified in the surveys and interviews that prevented genetic counselors from being able to make full use of the tool. Chief among these were the difficulties that many participants faced in creating the pedigree according to the website's formatting requirements, and the inability of participants to interpret the results. Both these themes recurred throughout the open responses given by survey respondents, as well as in the interviews conducted with experts. The reason why both these elements were so important to users can be linked to the context within which they used this tool. As mentioned briefly in the introduction and explored more fully in the interviews, genetic counselors work in very fast-paced environments, filled with back-to-back consultations with patients, and interspersed with a myriad of tests, research and qualitative analysis conducted using patient results, risk models, and databases of variant information. All decisions require a great deal of thought, care and background research. Decisions involving variants of unknown significance are even more difficult, given the lack of rigid guidelines and frameworks for how to interpret and treat them.

Given all of these factors, it is unreasonable to expect that counselors would have the patience or time to spend more than 5-10 minutes creating a single data sample and running a tool to get information that is often just one tiny piece in a huge puzzle. Therefore, in order for this tool to be useful to them and make their work 'less, not more', both the procedure for inputting data and the process by which they can extrapolate and understand results must be short.

With regards to creating pedigrees, the whole tabular format tended to be unintuitive to counselors, as they were more used to working with visual pedigrees (see figure 10 in the results section). On the other hand, it is unreasonable to imagine that the tool could be modified to run based on a visual pedigree without a software feature that converts visual pedigrees to coded relationships, which is a complex conversion. It is, however, possible, that certain modifications could be made to make the process easier. For example, instructions on the File Formatting tab could be merged with the tabular and visual representations on the Example tab, to allow users to read instructions and see the layout represented in both ways simultaneously. A downloadable template might be provided on the file formatting page, as interviewed counselors were almost always confused about how to go about even starting to create a pedigree file. Other formats of instructions, such as videos, could be provided, as was suggested in some of the survey responses and one of the interviews. The format of the pedigree itself could also be modified to allow users to input names instead of ID numbers (for the ID, mother and father columns), although this would require some modification to the back-end processing needed to run the calculation.

Formatting the pedigree to allow names instead of ID numbers also has potential repercussions for participants' trust of the tool. Most interviewed genetic counselors did not feel trust was a hugely important issue currently, as no information was being entered that could be traced back to actual people. However, if counselors were allowed to enter names instead, they

might then consider trust to be a bigger issue and require more validation of the privacy and security of the site. As illustrated by responses to survey questions about privacy and security, some key changes are necessary to the site to provide a sense of security (such as using a secure server and providing clear statements about data use).

With regards to presentation of the results, it is clear that genetic counselors need far more information than is currently provided on the site to be able to understand the results. The inability to interpret the results was a recurring theme in both surveys and interviews, and all of the interviewed counselors made it clear that they were unlikely to have the time (or in some cases, the access) to go through the cited background publications. Participants both needed definitions of the values being presented to them, as well as some form of context to understand what clinically significant and clinically non-significant results might look like. Suggestions for how this context might be presented varied, from providing a more extensive library of examples, along with explained results, to providing explicit threshold values (although as one interview participant pointed out that threshold values might keep changing).

4.2 Statistical geneticists' perceptions of usability

Our original intention was to collect survey as well as interview data from statistical geneticists, to be able to understand the perspectives and needs of this population. However, we were unable to collect survey data from many statistical geneticists, limiting our ability to fully understand and characterize their needs. Nevertheless, the two interviews we were able to conduct, as well as the few survey responses we did get, did shed some light on how different statistical geneticists' needs were from those of genetic counselors.

Since statistical geneticists often work in the context of large-scale research studies and with large quantities of data, relevance to single families was of less importance, as was the need for clinically relevant recommendations or context. They were interested in being able to manipulate more aspects, such as modifying penetrance values, or analyzing multiple families together. They were also more concerned about aspects such as proper citations and references being provided for penetrance values, and the ability to fully understand how the calculations were derived. Also, while they were more comfortable with the numerical results than genetic counselors, they were still not completely familiar with definitions and scales of the Full Likelihood Bayes factor and Cosegregation Likelihood Ratio and expressed interest in seeing those better defined on the results page. As a result of their greater familiarity with the numerical figures, they also had more questions and concerns about how exactly the calculations were done, especially when the calculations did not work at all, or worked in a way differently than they expected. Since statistical geneticists are likely to want to aggregate these results and use them in other calculations in the future, understanding how exactly the calculation works is very important. By harboring incorrect assumptions, they risk making inaccurate inferences and using erroneous computations on the values. Clarifying the formulas used by the tool somewhere, such as on the Background page (with potentially a link from the results page), is therefore a core need for this user group. By defining and describing the calculation of these figures, statistical geneticists may then become familiar enough with these values to make them comparable to other preferred figures like LOD scores.

Pedigree formatting requirements for AnalyzeMyVariant were much more intuitive to statistical geneticists, since (at least in the cases of those interviewed) their data was often already formatted in a similar way and, at most, data preparation simply required transposition of

data from multiple files into one pedigree. However, they had some issues with the display of the pedigree, discussed in the next section.

4.3 Usability issues common to both user groups

While statistical geneticists found pedigree-making a lot more intuitive, the visual rendering of the pedigree bothered them, as they found the use of the dividing line in individuals to be confusing, as was the lack of clear distinction between unknown and unaffected phenotypes. These issues were not restricted to statistical geneticists, as one of the interviewed genetic counselors also found the line down the middle to be confusing, while the misrepresented phenotype led one of the survey participants to surmise that the results looked incorrect because they misinterpreted an individual with unknown phenotype to be unaffected.

Other minor issues did come up for both sets of users as well. These included uninformative error messages, which prevented some surveyed genetic counselors from being able to run the calculator, and one interviewed statistical geneticist from seeing certain results. Lack of clarification about aspects of the models (assumptions about the non-existence of variant homozygotes, assumptions of dominant modes of inheritance) were also expressed by genetic counselors in surveys and statistical geneticists in the interviews. Finally, the need for statements about data use and privacy, as well as the provision of a secure domain were themes that came up as well. These minor issues can and should be rectified for the benefit of all future users. However, some issues were not common across both groups, and are discussed next.

4.4 Differences between genetic counselors and statistical geneticists, and implications for redesign

Some of the larger issues raised by genetic counselors and statistical geneticists require deeper consideration about the intended use of the tool. Some requirements of each population are directly in conflict, such as the preferences of and comfort with pedigree creation formats. Other requirements, while not directly in conflict, do not fully overlap either.

In defining and explaining the results, statistical geneticists did not require as extensive a contextual description as genetic counselors. Their needs were more mathematical, in terms of understanding how it was calculated, what scale it was on, and how to potentially aggregate it and maybe convert to other formats like LOD scores. Genetic counselors, on the other hand, wanted to get a much clearer idea, not only of the value itself and the scale, but also how they should interpret it, what degree of confidence to place on it, and how to use it in clinical care. As interview ID 3 put it:

> "So I think you have two different populations. If you're looking at direct patient care it needs to be very applicable and actionable at the clinical care level. How to integrate this, what is this really helping me to tell them. Versus if it is a lab or something, or someone that is using it to help reclassify a variant, then they would just want the scientific nitty gritty details."

Not only are their needs more extensive, they are also more subjective and perhaps harder to satisfy, given that information would be used in direct clinical care.

Given all of this, in the future, one or the other user population will have to be prioritized in the consideration of redesign and addition of information to the website. Genetic counselors could be taken as the primary user group, in which case the site should be tailored to provide a

lot more foundational and actionable background information about what the tool does and what it means for their clinical work practically. Uploading information should be more user-friendly and intuitive, possibly matching question-answer styles to fill out information, rather than a tabular file that needs to be coded. Explicit implications should be provided for results, so that counselors quickly understand how to apply results to their larger considerations of treatment for patients. This avenue represents more extensive redesign to the functionality and operation of the website and requires a lot more time to be invested in understanding how to provide clinically actionable information that is accurate and leaves no room for misunderstanding and erroneous application by unqualified users. The trade-off is that the tool can then be used at the ground level to help counselors understand and manage the disease risk of real patients who come to them with variants of unknown significance.

On the other hand, statistical geneticists could be taken as the primary user group. This avenue represents less extensive redesign, as not much about the base functionality of the site would need to be changed. Statistical geneticists would simply need more high-level information about the statistical formulae and assumptions being used to generate results, as well as potentially the option to redefine some values being used (penetrance values). Since the site already seems more tailored to statistical geneticists' expectations and needs, this avenue would be the easier of the two. However, the trade-off is that the impact of a tool like this on regular patients and families would be considerably less, as statistical geneticists are more likely to work on research studies, at a level far above influencing individual clinical treatment. Of course, it must also be remembered that much of our results and conclusions about statistical geneticists are based on a restricted sample of data, and therefore any future redesign that seeks to cater to

the needs of statistical geneticists should perform more extensive usability evaluations with that subset of participants.

While it is technically possible to satisfy both user groups' needs, perhaps by having different 'tabs' for different types of users to use the tool differently, this would add a great deal of complexity and bulk to the site. A common positive theme throughout this study has been users' appreciation of the simplicity of the design and the tool. They liked that it was targeted for one specific purpose, and the lack of bulk in functionality led to it being perceived as efficient and sleek. To attempt to satisfy both user groups would mean drastically reducing the simplicity of the website. The trade-off between attracting more users and keeping the tool simple should be considered in future redesign.

4.5 Implications for methodology and the role of the user experience researcher

As an online cosegregation analysis tool to assess the statistical disease likelihood of a genetic variant of uncertain significance, built for use with eight specific cancer genes, AnalyzeMyVariant is a very niche tool. It was designed to be used by a select few users with a high threshold expertise in genetics, disease and unknown variants. Based on this niche use, we designed a survey usability assessment targeted to these genetics experts, to collect feedback about how to improve the design. However, as we collected data, it became clear that there was more depth to the usability issues than was clear from the survey responses. After expanding the study methodology to include contextual interviews, we were able to gain a lot more insight from user groups, and more importantly, we were able to clearly define and distinguish between the needs and expectations of the two different user groups. While the survey data did not contain enough quantities of responses from both user types to truly compare the responses on a large

scale, even comparing single responses of statistical geneticists and genetic counselors did not offer the kind of clarity of distinction between user needs as was evident with just a few interviews. The real difference came from understanding the context of use.

Context of use is a construct that has been extensively used in the evaluation of usability. Maguire et al. [43] in 2001 extensively discuss the importance of understanding the main goals of the user community as well as the environment in which the tool might be used. Many models and evaluation methods incorporate the consideration of use context as part of other heuristics, such as 'flexibility' for different environments, or even 'ease of use' [44]. However, by restricting a complex consideration like this to one-dimensional heuristics, much of the nuance and complexity is lost. Indeed, when describing context of use, Maguire et al. discuss the importance of evaluating use contexts in two ways: both in laboratory simulated-use situations, as well as field tests, in real contexts. While some studies [45] have found that simulated situations could perform as well, if not better, than real life evaluations in identifying usability issues, many considerations are left out of this comparison. Joyce et al. [44] discuss the importance of 'decoupling' context of use from heuristic evaluations and assessing context in a more nuanced manner.

While much of this research has been done in the realm of mobile technologies, where context of use may be constantly changing, our results demonstrate that the importance of considering context of use separately is present in even web-based niche technologies like AnalyzeMyVariant. While many constructs in the survey like satisfaction or ease of use may be assumed to taken context into consideration, the very fact that it was a task-based study makes it more similar to Maguire's laboratory simulated-use situation. Certainly, many of the issues that subsequently came up in interviews were also mentioned in surveys, but the reasoning and

understanding behind why these were issues for users came only with a more in-depth qualitative exploration of users' work contexts. The interviews, by contrast, which encouraged users to use their own data to make pedigrees and explain how they might use this tool in their work, was therefore more akin to the field evaluation described by Maguire. Only by a combination of the two were we able to form a clear picture of users' needs in this study, reinforcing both Maguire et al.'s as well as Joyce et al.'s conclusions about the need to decouple context of use in the assessment of tools, particularly those that are intended to be used as part of users' everyday work.

Bibliography

[1] J. M. O. Ranola and B. H. Shirts, “AnalyzeMyVariant.org,” *AnalyzeMyVariant*, 2017. [Online]. Available: http://analyze.myvariant.org. [Accessed: 28-Sep-2017].

[2] G. P. Jarvik and B. L. Browning, “Consideration of Cosegregation in the Pathogenicity Classification of Genomic Variants,” *Am J Hum Genet*, vol. 98, no. 6, pp. 1077–81, Jun. 2016.

[3] J. M. O. Rañola, Q. Liu, E. A. Rosenthal, and B. H. Shirts, “A comparison of cosegregation analysis methods for the clinical setting,” *Fam. Cancer*, Jul. 2017.

[4] S. E. Plon *et al.*, “Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results,” *Hum Mutat*, vol. 29, no. 11, pp. 1282–91, 2008.

[5] M. J. Landrum *et al.*, “ClinVar: public archive of relationships among sequence variation and human phenotype,” *Nucleic Acids Res*, vol. 42, no. Database issue, pp. D980-5, Jan. 2014.

[6] B. H. Shirts, A. Jacobson, G. P. Jarvik, and B. L. Browning, “Large numbers of individuals are required to classify and define risk for rare variants in known cancer risk genes,” *Genet. Med.*, vol. 16, no. 7, pp. 529–534, Jul. 2014.

[7] E. A. Rosenthal, J. M. O. Ranola, and B. H. Shirts, “Power of pedigree likelihood analysis in extended pedigrees to classify rare variants of uncertain significance in cancer risk genes,” *Fam. Cancer*, vol. 16, no. 4, pp. 611–620, Oct. 2017.

[8] B. H. Shirts and E. A. Rosenthal, “Power of Single Extended Pedigrees to Classify Rare Variants of Uncertain Significance in BRCA1 and BRCA2,” presented at the American Society of Human Genetics Annual Meeting, 2015.

[9] D. Thompson, D. F. Easton, and D. E. Goldgar, “A Full-Likelihood Method for the Evaluation of Causality of Sequence Variants from Family Data,” *Am. J. Hum. Genet.*, vol. 73, no. 3, pp. 652–655, Sep. 2003.

[10] B. H. Shirts, C. C. Pritchard, and T. Walsh, “Family-Specific Variants and the Limits of Human Genetics,” *Trends Mol. Med.*, vol. 22, no. 11, pp. 925–934, Nov. 2016.

[11] L. Mohammadi *et al.*, “A simple method for co-segregation analysis to evaluate the pathogenicity of unclassified variants; BRCA1 and BRCA2 as an example,” *BMC Cancer*, vol. 9, no. 211, 2009.

[12] A. K. Jaiswal, R. Niraj, and P. Venugopal, “Context-general and Context-specific Determinants of Online Satisfaction and Loyalty for Commerce and Content Sites,” *J. Interact. Mark.*, vol. 24, no. 3, pp. 222–238, Aug. 2010.

[13] C. L. Corritore, B. Kracher, and S. Wiedenbeck, “On-line trust: concepts, evolving themes, a model,” *Int. J. Hum.-Comput. Stud.*, vol. 58, no. 6, pp. 737–758, Jun. 2003.

[14] M. Söllner, A. Hoffmann, H. Hoffmann, and J. M. Leimeister, “Towards a Theory of Explanation and Prediction for the Formation of Trust in IT Artifacts,” 2011.

[15] Z. Yang, M. Jun, and R. T. Peterson, “Measuring customer perceived online service quality: Scale development and managerial implications,” *Int. J. Oper. Prod. Manag.*, vol. 24, no. 11, pp. 1149–1174, Nov. 2004.

[16] Z. Yang, S. Cai, Z. Zhou, and N. Zhou, “Development and validation of an instrument to measure user perceived service quality of information presenting Web portals,” *Inf. Manage.*, vol. 42, no. 4, pp. 575–589, May 2005.

[17] J. Gummerus, V. Liljander, M. Pura, and A. van Riel, “Customer loyalty to content-based Web sites: the case of an online health-care service,” *J. Serv. Mark.*, vol. 18, no. 3, pp. 175–186, May 2004.

[18] Y. Lee and K. A. Kozar, “Understanding of website usability: Specifying and measuring constructs and their relationships,” *Decis. Support Syst.*, vol. 52, no. 2, pp. 450–463, Jan. 2012.

[19] A. Seffah, M. Donyaee, R. B. Kline, and H. K. Padda, “Usability measurement and metrics: A consolidated model,” *Softw. Qual. J.*, vol. 14, no. 2, pp. 159–178, Jun. 2006.

[20] W.-C. Chiou, C.-C. Lin, and C. Perng, “A strategic framework for website evaluation based on a review of the literature from 1995–2006,” *Inf. Manage.*, vol. 47, no. 5–6, pp. 282–290, Aug. 2010.

[21] L. L. Constantine and L. A. D. Lockwood, *Software for Use: A Practical Guide to the Models and Methods of Usage-Centered Design*. Pearson Education, 1999.

[22] B. Shneiderman, C. Plaisant, M. Cohen, and S. Jacobs, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson Education, 2017.

[23] J. Nielsen, *Usability Engineering*. Morgan Kaufmann, 1994.

[24] J. Preece, Y. Rogers, H. Sharp, D. Benyon, S. Holland, and T. Carey, *Human-Computer Interaction*. Essex, UK, UK: Addison-Wesley Longman Ltd., 1994.

[25] B. Shackel and S. J. Richardson, *Human Factors for Informatics Usability*. Cambridge University Press, 1991.

[26] B. J. Fogg and H. Tseng, “The elements of computer credibility,” in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 1999, pp. 80–87.

[27] M. J. Dutta-Bergman, “The Impact of Completeness and Web Use Motivation on the Credibility of e-Health Information,” *J. Commun.*, vol. 54, no. 2, pp. 253–269, Jun. 2004.

[28] S. A. Eroglu, K. A. Machleit, and L. M. Davis, “Atmospheric qualities of online retailing: A conceptual model and implications,” *J. Bus. Res.*, vol. 54, no. 2, pp. 177–184, 2001.

[29] V. McKinney, K. Yoon, and F. “Mariam” Zahedi, “The Measurement of Web-Customer Satisfaction: An Expectation and Disconfirmation Approach,” *Inf. Syst. Res.*, vol. 13, no. 3, pp. 296–315, 2002.

[30] E. Sillence, P. Briggs, P. Harris, and L. Fishwick, “A framework for understanding trust factors in web-based health advice,” *Int. J. Hum.-Comput. Stud.*, vol. 64, no. 8, pp. 697–713, Aug. 2006.

[31] W. F. Cohn *et al.*, “Health Heritage©, a Web-Based Tool for the Collection and Assessment of Family Health History: Initial User Experience and Analytic Validity,” *Public Health Genomics*, vol. 13, no. 7–8, pp. 477–491, 2010.

[32] M. B. Kinzie, W. F. Cohn, M. F. Julian, and W. A. Knaus, “A user-centered model for web site design: Needs Assessment, User Interface Design, and Rapid Prototyping,” *J. Am. Med. Inform. Assoc.*, vol. 9, no. 4, pp. 320–330, 2002.

[33] C. Shyr, A. Kushniruk, and W. W. Wasserman, “Usability study of clinical exome analysis software: Top lessons learned and recommendations,” *J. Biomed. Inform.*, vol. 51, pp. 129–136, Oct. 2014.

[34] H.-P. Lu and D. H. Gustafson, “An empirical study of perceived usefulness and perceived ease of use on computerized support system use over time,” *Int. J. Inf. Manag.*, vol. 14, no. 5, pp. 317–329, 1994.

[35] W. J. Doll and G. Torkzadeh, “The Measurement of End-User Computing Satisfaction,” *MIS Q.*, vol. 12, no. 2, pp. 259–274, 1988.

[36] S. Ha and L. Stoel, “Consumer e-shopping acceptance: Antecedents in a technology acceptance model,” *J. Bus. Res.*, vol. 62, no. 5, pp. 565–571, May 2009.

[37] L. J. Frewer, C. Howard, D. Hedderley, and R. Shepherd, “What determines trust in information about food-related risks? Underlying psychological constructs,” *Risk Anal.*, vol. 16, no. 4, pp. 473–486, 1996.

[38] P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, and J. G. Conde, “Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support,” *J. Biomed. Inform.*, vol. 42, no. 2, pp. 377–381, Apr. 2009.

[39] S. Jamieson, “Likert scales: how to (ab)use them,” *Med. Educ.*, vol. 38, no. 12, pp. 1217–1218, 2004.

[40] G. H. Lubke and B. O. Muthén, “Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons,” *Struct. Equ. Model. Multidiscip. J.*, vol. 11, no. 4, pp. 514–534, Oct. 2004.

[41] B. G. Glaser, “The Constant Comparative Method of Qualitative Analysis,” *Soc. Probl.*, vol. 12, no. 4, pp. 436–445, 1965.

[42] H. Boeije, “A Purposeful Approach to the Constant Comparative Method in the Analysis of Qualitative Interviews,” *Qual. Quant.*, vol. 36, p. 21, 2002.

[43] M. Maguire, “Context of Use within usability activities,” *Int. J. Hum.-Comput. Stud.*, vol. 55, no. 4, pp. 453–483, Oct. 2001.

[44] G. Joyce, M. Lilley, T. Barker, and A. Jefferies, “Mobile Application Usability Heuristics: Decoupling Context-of-Use,” in *Design, User Experience, and Usability: Theory, Methodology, and Management*, vol. 10288, A. Marcus and W. Wang, Eds. Cham: Springer International Publishing, 2017, pp. 410–423.

[45] S. Po, S. Howard, F. Vetere, and M. B. Skov, “Heuristic Evaluation and Mobile Usability: Bridging the Realism Gap,” in *Mobile Human-Computer Interaction - MobileHCI 2004*, 2004, pp. 49–60.

Appendices

Appendix A: Study Invitation - Surveys

This document shows sample text for invitations sent by email to prospective participants.

Dear [insert name],

My name is Aarti Swaminathan, and I am a graduate student in the Biomedical Informatics department at the University of Washington. I'm currently working on a thesis project with Dr. Annie Chen and Dr. Brian Shirts at UW Medicine, to assess user experience and trust in an online family cosegregation analysis tool.

I obtained your contact information from *[insert source]*, and I'm writing to invite you to participate in this study. You can do so by exploring the site and filling out a short online survey. If you have any colleagues who might be interested in the site or in participating in the survey, please share it with them as well.

AnalyzeMyVariant is a tool designed by Dr. Brian Shirts at the University of Washington for genetics experts to perform cosegregation analysis on pedigree data for several genes with known penetrance. The tool currently has models for eight hereditary cancer risk genes.

The aim of this study is to assess how well the site is suited to your educational, clinical, and research needs; how it can be improved to give you information you trust; and how we can give you a better holistic experience. By providing us with feedback, you will help us improve it for your future needs and those of the entire genetics community.

In addition, you can sign up to join a raffle draw, with a $100 Amazon gift card prize being given to the winner, at the end of the study.

Please note that whatever responses you submit will be anonymous. You also do not have to use any personal data when testing the calculator, but should you choose to do so, your data will not be logged or saved as part of the study.

If you would like to participate, please access the link below.

https://redcap.iths.org/surveys/?s=PAP7R3XFHE

Thanks, and I look forward to hearing from you!

Best,

Aarti Swaminathan

Graduate Student, Biomedical and Health Informatics

University of Washington

Please note: This study has been approved by the University Institutional Review Board (STUDY00003809).

Appendix B: Survey Splash Consent Page

This is the form that will be displayed when the prospective study participant clicks through to the survey. Completing this form is necessary in order to fill out the survey:

_____

Page 1 of 2

_____

Thank you for participating in our survey study to assess user experience. This is a study to investigate how online scientific tools are suited for users like you. We will ask how much you trust one specific website, AnalyzeMyVariant, including how easy it is to navigate, and to learn how to use the tool.

AnalyzeMyVariant

AnalyzeMyVariant is a tool designed for geneticists and genetics students to perform family cosegregation analyses to calculate the probability that a single genetic variant occurs with disease more often than expected by chance. This analysis generates a likelihood ratio or Bayes factor that a variant of interest is pathogenic with data from a single family. The tool currently uses built-in penetrance for several cancer risk genes relative to population risks.

Participation & Benefits

Your participation is completely voluntary.

There are ways in which we hope you will benefit from this study.

- You will have access to a new tool that we hope may be helpful to you in your work.

- You can choose to be enter a raffle draw for an Amazon gift card, with one winner being awarded a gift card worth $100 at the end of the study.

We will use information from the survey to understand how professionals interact with technical websites in general and specifically how AnalyzeMyVariant can improve. The only risk of participating in this research study is loss of confidential information, which is unlikely. Measures have been taken to protect your confidentiality. This survey is focused on your user experience of the tool. We collect limited information on demographic characteristics, which we will report on an aggregate level. Email addresses entered for the raffle draw will be stored separately from the study data.

It should take you approximately 15-20 minutes to explore and test the website if you use the example pedigree file that is provided. If you choose to use your own pedigree file, it may take longer. Completing the survey should take 15-20 minutes.

Contact & Consent

You may contact me with any questions about the survey at (425) 240-9943 or by email at aarti94@uw.edu, or the AnalyzeMyVariant developers with questions about the web tool at (617) 733-3930 or by email at findmyvariant@uw.edu. If you have questions or concerns about your rights as a research subject you may contact the Institutional Review Board at (206) 543-0098. If you would like to participate in this study, please follow the instructions below.

By clicking on the next page button below, you signify that you consent to participate in this study and allow your data to be used in analysis of the website AnalyzeMyVariant.

_____

Page 2 of 2

_____

The website tool is available here (http://analyze.myvariant.org/). Please open this link and take your time to explore any portion of the website that you wish. However, we do ask that, at some point, you test the Cosegregation Analysis tool (click the 'Get Started' button under Cosegregation Analysis on the main page or go here) so that you can complete the survey about your experience.

To run this tool, you will need a pedigree file in plain text (.txt) or Excel (.xlsx) format.

If you have a family for which you would like to calculate cosegregation likelihood, please do so. Refer to the "File formatting" and gene-specific information in the 'Penetrance Classes' tab on the website to find out how to code your pedigree file.

Alternatively, the "Example" tab on the site has an example pedigree. Read through the description on the Example tab to understand the features of the data in the file, then download the file in either of the formats linked at the bottom. You can use the example pedigree, a modified version of the example, or create your own realistic family that follows the formatting instructions provided. Download the example file, make any modifications if you wish to, re-upload it and hit 'Start Analysis' to see the results for the example file.

After uploading the file, select a gene you want to perform the analysis for, and click 'Start Analysis'.

After you have tested the Cosegregation Analysis website, please come back here and click the 'Next Page' button below to access the survey.

Appendix C: Final Survey Questions

USABILITY	
Information completeness	(I found the information provided on the website to be thorough and complete.)
Ease of use (website)	(I found the website clear and easy to use.)
Ease of use (results)	(I found the results clear and easy to understand.)
Output relevance	(I found the output relevant.)
QUALITY	
Accuracy	(I found the tool to be accurate.)
Efficiency	(I found the tool to be efficient.)
Reliability	(I found the tool to be reliable.)
Need fulfillment	(The tool told me exactly what I needed to know.)
SAFETY	
Privacy	(The website respects my right to control access to the personal information that I uploaded.)
Security	(I feel that this website is secure.)
SATISFACTION	
Satisfaction	(I was satisfied with my experience with the tool.)

Appendix D: Study Invitation – Interviews

Dear [insert name],

My name is Aarti Swaminathan, and I am a graduate student in the Biomedical Informatics department at the University of Washington. I'm currently working on a thesis project with Dr. Annie Chen and Dr. Brian Shirts at UW Medicine, to assess user experience and trust in an online family cosegregation analysis tool.

I obtained your contact information from [*insert source*], and I'm writing to invite you to participate in a short interview.

The interview would involve a general discussion of your work and how AnalyzeMyVariant may or may not fit into it, as well as exploring some of the usability issues that you might face with the tool. The goal is to understand how we can make this tool better for target users like you. You would not be asked to provide specific information about any of your clients in the interview, or any specific data that you work with. All of your responses would be stored and analyzed in a de-identified form.

The interview should not take more than an hour of your time, and I could either come to your office, or we could conduct the interview remotely. I will audio or video record, per your preference, transcribe and de-identify the interviews prior to analysis.

If you are interested in participating in this short interview, please read through and sign the consent form attached, and send it back in reply to this email at your earliest convenience, after which we can discuss when best to schedule the interview.

Thanks, and I look forward to hearing from you!

Best,

Aarti Swaminathan

Graduate Student, Biomedical and Health Informatics

University of Washington

Please note: This study has been approved by the University Institutional Review Board (STUDY00003809).

Appendix E: Consent Form – Interviews

Researchers' statement

We are asking you to be in a research study. The purpose of this consent form is to give you the information you will need to help you decide whether to be in the study or not. Please read the form carefully. You may ask questions about the purpose of the research, what we would ask you to do, the possible risks and benefits, your rights as a volunteer, and anything else about the research or this form that is not clear. When we have answered all your questions, you can decide if you want to be in the study or not. This process is called "informed consent." We will give you a copy of this form for your records.

PURPOSE OF THE STUDY

The purpose of this interview is to better understand how the tool (AnalyzeMyVariant) might fit into your work and accordingly can be improved to suit your needs, as well as to get a deeper sense of your opinions of various aspects of the site.

AnalyzeMyVariant is a tool designed for geneticists and genetics students to perform family cosegregation analyses to calculate the probability that a single genetic variant occurs with disease more often than expected by chance. This analysis generates a likelihood ratio or Bayes factor that a variant of interest is pathogenic with data from a single family. The tool currently uses built-in penetrance for several cancer risk genes relative to population risks.

STUDY PROCEDURES

The interview will consist of 4 parts.

1. A brief background of your experience and specialization
2. Understanding your work commitments, schedule, and the context of your work within which you may or may not find it useful to use this tool.
3. A think-aloud portion where you explore each section of the website, and provide your opinions and feedback of different features and the information provided.
4. A final wrap-up portion where you provide high-level opinions about the tool, as well as any thoughts, comments or suggestions you might have about improving it.

RISKS, STRESS, OR DISCOMFORT

The only potential risk of participating in this interview is loss of confidential information, which is unlikely. Measures have been taken to protect your confidentiality. Any and all artifacts (recordings, transcripts etc.) will be stored on a secure online server, in compliance with the University of Washington's data security guidelines.

The interview will be recorded. Ideally, a video recording will be made of the interview, but if you are uncomfortable with being video recorded, you may request an audio recording to be made instead. Some form of recording is necessary to be able to adequately transcribe the interview later.

The study researchers will personally transcribe the interviews from the recording, and will simultaneously de-identify any personal information mentioned during the interview, such as your name and place of work, and any information about your clients or colleagues. If there are any additional pieces of information you would like to be de-identified in the transcripts, please let the researchers know, at any point during or after the interview.

BENEFITS OF THE STUDY

By participating in this study, we hope you will benefit by gaining access to, and understanding of, a new tool that may be beneficial to your professional work. In addition, your feedback will help us to make improvements to the tool AnalyzeMyVariant for you and other users like you.

CONFIDENTIALITY OF RESEARCH INFORMATION

As mentioned previously, identifiable information will be present only in the audio/video recordings of this interview. All recordings will be stored on a secure server, in compliance with the University of Washington's data security guidelines.

De-identified transcripts will be stored with generic study ID numbers, and will also be stored on secure servers.

Government or university staff sometimes review studies such as this one to make sure they are being done safely and legally. If a review of this study takes place, your records may be examined. The reviewers will protect your privacy.

RESEARCH-RELATED INJURY

We do not anticipate that you will be harmed in any way or receive any injuries during this interview.

"The UW does not normally provide compensation for harm except through its discretionary program for medical injury. However, the law may allow you to seek other compensation if the harm is the fault of the researchers. You do not waive any right to seek payment by signing this consent form."

Subject's statement

This study has been explained to me. I volunteer to take part in this research. I have had a chance to ask questions. If I have questions later about the research, or if I have been harmed by participating in this study, I can contact one of the researchers. If I have questions about my rights as a research subject, I can call the Human Subjects Division at (206) 543-0098 or call collect at (206) 221-5940. I will receive a copy of this consent form.

Printed name of subject	Signature of subject	Date

Appendix F: Interview Transcript

Interview Protocol:

Assessing User Satisfaction and Trust with Online Tool AnalyzeMyVariant

Hi! Thank you for participating in this follow-up interview about AnalyzeMyVariant. The purpose of this interview is to better understand how this tool might fit into your work and accordingly can be improved to suit your needs, as well as to get a deeper sense of your opinions of various aspects of the site.

[if not done over email in case of remote interviews] First, please take a few minutes to read through the consent form, and sign. If you have any questions, please do ask.

With your permission, I'd like to video record this interview. If you're not comfortable with video recording it, we could also alternately audio record it. Some form of recording would make it much easier for me to remember everything you tell me later, when I'm transcribing the conversation.

The video recordings will be stored on a secure online server. I will later transcribe it, and everything you tell me in this interview will be de-identified to remove specifics of your name, job and any other identifiable pieces of information. All the data will be used only to guide the improvement of the site, and will not be used for any other purpose.

The full transcript of this interview will likewise be stored on a secure online server, and not be made available to anyone else.

Part 1: Background

What is your profession?

How many years of experience do you have in this field?

What is your area of specialization?

Part 2: Context of Use

This next part of the interview is to help me understand what your work is like, and how this tool may or may not fit into your workflow and help you carry out the responsibilities of your job.

What does your typical workday look like? What do you spend the majority of your day doing?

How often do you collect family data?

How often do you create pedigrees? How do you create them? [have them demonstrate the creation of a pedigree]

How often do you come across variants of unknown significance?

When you do come across families with a variant of unknown significance, what is the general procedure for assessing disease risk?

- What are all the pieces of information you need?
- How do you analyze them?
- In what form do you communicate the results of that analysis to clients?
- What are the key results that you need?

How do you see yourself using this tool?

Part 3: Using AnalyzeMyVariant

[Ask them to open the website, and go to each page in turn.]

For each page: [The following questions are meant to be a guide to the kinds of questions asked for each page; not all questions will be asked for all pages.]

- What do you think of this page?
- Is the information clear?
- Is there information missing?
- Is there anything you don't understand?
- Does it seem relevant to you?

[Ask them to try creating a pedigree according to the instructions. See if they can follow the instructions and make note of where they find it difficult.]

[Ask them to run through the task of uploading a pedigree, running the calculator, and analyzing the results. Ask them to think-aloud through the process.]

Potential questions to probe participants during the think-aloud portion:

- Do you understand the results?
- How much do you understand?
- Is there anything missing?
- What do you feel could be added to make it easier for you to understand the results?

What could be improved to help you use this information in working with your clients?

Would you trust this site to upload the personal data of your clients?

Part 4: Wrap-Up

Now that you've just gone through it, do you see yourself using this tool? Why or why not?

How often do you think you might use it?

What prevents you from using it more often?

Any other thoughts, comments or suggestions about potential improvements to the tool?
