# THE SYSTEMATIC ASCERTAINMENT OF STRUCTURED FAMILY HEALTH INFORMATION USING AN ONLINE PATIENT PORTAL

By

Jonathan Andrew Holt

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biomedical Informatics

August, 2012

Nashville, Tennessee

Approved:

Associate Professor Dario A Giuse

Professor Cindy S Gadd

Associate Professor Paul A Harris

*To my family*

ACKNOWLEDGEMENTS

I am grateful for the National Library of Medicine, and subsequently the American taxpayer, for supporting me financially to achieve this Master's degree and recognizing the importance of training physician leaders in the field of Biomedical Informatics. This thesis was funded thru a National Library of Medicine training grant, 5-T15-LM007450.

I would also like to acknowledge my wife, Mia Levy, and our daughter, Dana, for their continuing love and support.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AHRQ ..... Agency for Healthcare Research and Quality

AJAX ..... Asynchronous JavaScript and XML

CDC ..... Centers for Disease Control

CSS ..... Cascading Style Sheets

CUI ..... Concept Unique Identifier

DAG ..... Directed Acyclic Graph

EHR ..... Electronic Health Record

FHI ..... Family Health Information

FHQ ..... Family Health Questionnaire

GEDCOM ..... Genealogy Data Communication

HL7 ..... Health Level 7

HTML ..... HyperText Markup Language

JSON ..... JavaScript Object Notation

MRI ..... Magnetic Resonance Imaging

MyFaV ..... www.MyFamilyatVanderbilt.com

MyHaV ..... www.MyHealthatVanderbilt.com

OMIM ..... Online Mendelian Inheritance in Man

PHR ..... Patient Health Record

SUS ..... System Usability Scale

SVG ..... Scalable Vector Graphic

UMLS ..... Unified Medical Language System

XML ..... Extensible Markup Language

# INTRODUCTION

*"If a woman circumcised her first child and he died as a result of exsanguination and a second child died similarly, she must not circumcise her third child."*
The Jewish Talmud (1)

In current practice, clinicians obtain information regarding the health of a patient's family members during routine clinical encounters. This Family Health Information (FHI) helps identify individuals who are at increased risk for adverse health conditions due to inherited genetic or environmental predisposition, particularly cancer, heart disease and diabetes (2). Traditionally, clinicians ascertain FHI from patients in the form of an oral interview during the course of a clinical encounter. This patient-reported FHI is often the accumulation of verbal reports from other family members and is limited by the interpretation and health literacy of each of the family members. During the encounter, the clinician will often clarify these reports, updating the medical record, and then may use this information for clinical decision-making. Based on reported FHI, a patient may be stratified as being at increased risk for an adverse health condition, triggering specific preventative strategies to be implemented. For example, a family history of breast cancer may increase the risk of a patient developing the same depending on the number of affected relatives and the age of onset of this condition (3). Preventative-screening modalities, such as screening breast MRIs and earlier mammograms, are an option for individuals who are at increased risk of breast cancer (3). Stratifying risk based on family history and performing individualized screening measures may detect breast cancer early, thus decreasing the patient's morbidity and mortality (4). Appropriate stratification of patients based on familial risk relies on the clinician's ability to ascertain, and the patients' ability to report, complete and accurate FHI. Since the accuracy of FHI depends on the patients' understanding and interpretation of the information reported to them by other members of the family, secondhand information may lead

to incomplete and inaccurate reporting to the clinician. Complicating this factor, the collection of detailed FHI often requires more time than is available in the typical patient encounter in the primary care setting (5). As a result, FHI is often inconsistently and ineffectively communicated during clinical encounters, leading to FHI that is often incomplete, thus limiting its potential use for clinical decision-making. Yet, FHI epitomizes a cost effective strategy critical to the emerging practice of genome-informed and personalized medicine (6). Despite this, FHI is often overlooked and underutilized in the primary care setting.

Family Health Questionnaires (FHQ) have been use to ascertain FHI directly from patients, but are typically paper-based, in free-text form, and disease or specialty-specific (7,8). Most FHQs to date have focused on single diseases or a group of diseases, such as cancer, and are typically used for epidemiologic research rather than clinical care (7). The FHI obtained from a clinical medical geneticist or genetic counselor has classically served as the reference standard when comparing the accuracy of FHQs. These experts may spend up to 5.5 hours per patient reviewing patient-reported FHI and locating relevant documentation and medical records of family members (9).

In the primary care setting, few studies have validated the sensitivity and specificity of FHI, leading to the lack of its unequivocal clinical utility in routine medical decision making (10). Many epidemiological studies have represented the ascertainment of FHI as a dichotomized variable, only representing it as positive when FHI was documented in the medical record, by a patient survey, self-reported, discussed by the clinician, or confirmed by an observer of the clinical encounter (10). In other studies, a positive family history is defined as having one other affected relative with the same condition (11). This measure fails to adequately represent size of the family, number of affected relatives, inheritance pattern, or age of onset when assigning a risk score (12). The lack of a structured representation of FHI clearly limits the analysis of its significance and application into clinical practice. Moreover, simplifying FHI as a dichotomous

variable underrepresents its complexity and limits the interpretation of its sensitivity, specificity and clinical utility (13). Probabilistic risk scores based on reported family history using the number of affected family members have been attempted with limited success (12).

These gaps create an opportunity for the development of web-enabled tools to facilitate the ascertainment of structured FHI directly from patients and their family members. This approach encourages communication directly among the informants, and potentially provides more accurate information necessary for clinical decision-making by providers or future automated risk stratification tools.

The main contribution of this thesis will be to describe the development and evaluation of an online tool for the systematic ascertain of structured Family Health Information using a web-based patient portal, www.myfamilyatvanderbilt.com *(MyFaV)*. The overall goal of this thesis will be to determine if this patient portal is a usable tool for ascertaining structured FHI directly from patients. This thesis will describe previous attempts at building structured FHI and why they do not meet the needs of a scalable and generalizable solution. Next, the conceptual framework that led to the development of the information model utilized will be discussed, followed by the methodology employed to evaluate the usability of this site. The results that were obtained and the conclusions drawn from this study will be presented. Finally, this thesis will discuss the future directions that this research hopes to achieve, specifically a) improving the accuracy of FHI by reconciling agreement and disagreement between family members and b) the stratification of risk using decision support algorithms into the clinical workflow.

## BACKGROUND AND SIGNIFICANCE

*"To fail to take a good family history is bad medicine."*
Dr. Barton Childs (14)

Ninety-six percent of primary care providers surveyed by Fuller, et al. indicated that it is standard-of-care to review and update FHI (15). Yet, FHI is discussed in only 51% of new patient visits and a pedigree diagram (see figure 2 for an example) is present in only 11% of patient's medical records (16). This is often attributed to clinicians underestimating the value of reviewing FHI in clinical practice and to lack of sufficient time necessary to obtain it (6). The typical time required for primary care clinicians to build a 3-generation pedigree of FHI is 15-20 minutes (5). However, these clinicians report that on average they only spend 1-5 minutes discussing family history during a typical patient encounter (15). Complicating this problem, primary care clinicians report having limited confidence in their knowledge of genetics and their ability to properly evaluate FHI for genetic risk (17). Clinicians are not alone. The CDC analyzed data from the 2004 HealthStyles Survey to assess the practice, knowledge and attitudes of U.S. residents with regards to their FHI (18). This report indicated that 96.3% of the 4,345 survey respondents believed that FHI was an important factor in their health; however, only 29.8% reported actively collecting health information from their relatives to develop a family health history. Educated, previously or currently married women were found to be those most likely to actively collect and report their FHI (13,19).

The key elements important to risk stratification based on FHI include: number and gender of affected relatives, degree of relationship, ancestry (ethnicity or region of origin), lineage (maternal or paternal), and an unambiguous representation of health conditions along with the age of onset for each family member (10).

In one study using a 8-minute structured questionnaire and a 7 minute pedigree interview in a primary care clinic, approximately 20% of patient were identified as being at increased risk

for common adult onset diseases with known genetic component (20). Specifically, the prevalence of a family history of breast and ovarian cancer reported by patients via a Self-Administered Questionnaire found that 9.4% of women surveyed had a significant family history of cancer (21).

The U.S. Preventative Services Task Force recommends that women whose FHI is associated with an increased risk for deleterious mutations in the BRCA genes be referred for genetic counseling, and for possible genetic testing (4). These women, stratified as being at increased risk, can be counseled regarding risk-reducing options including personalized screening plans, chemoprevention, and prophylactic mastectomies (3).

Numerous reference standards to compare the accuracy of reported FHI have been proposed, with various degrees of practicality (22–24). One approach used medical records, death certificates and autopsy reports as the reference standard, and found that site specific accuracy for a cancer diagnosis was 84% among first-degree relatives (25). Another report used cancer family registries as the reference standard and found the probability of agreement for first-degree relatives was 95.4% (26). Validation by family members has also been used for comparison (23). Perhaps the most utilized reference standard for FHI is one ascertained by a Clinical Geneticist or Genetic Counselor, with levels of agreement reported to be over 77% (27).

While little is known about the accuracy of patient-generated FHI regarding common conditions and how it compares to that obtained by primary care clinicians (10), there is some evidence that patient-reported FHI regarding breast and colon cancer histories affecting first-degree relatives are accurate (28). For cancer, the specificity of patient-reported FHI is generally reported to be high (90-95%) but the sensitivity tends to be low (33-95% with a larger confidence interval (10)), indicating patients are typically better at reporting the absence of disease than the presence of a specific type of disease in their relatives. Sharing of a family history of a mental health condition has unique challenges. The sensitivity for a patient-reported family history of Affective Disorder was reported to be as low as 59% (23). It is no surprise that the accuracy of

FHI seems to be better the closer the degree of relationship is to one who is affected (10), and increasing the number of informants increases the sensitivity and decreases the false negative rate (23).

To determine if there was evidence that the routine ascertainment of FHI resulted in any psychological harm for the patients or relatives, the Agency for Healthcare Research and Quality (AHRQ) performed an extensive review of the available publications (29). Relevant research that met the inclusion criteria (randomized clinical trials) was limited to three studies (13,30,31). These findings suggest that structured collection of FHI had no major deleterious psychological harm in the term of 6-12 weeks after the study period. These findings were similar when the collection of FHI was augmented with the feedback of familial risk based on the reported FHI. During all three of these studies, the short form (six-item) of the Spielberger STAI (State-Trait Anxiety Inventory) was used to measure the psychological effect of ascertaining FHI. All three of these reports supported the conclusion that there appears to be minimal risk for potential psychological harm associated with ascertaining FHI from users.

The Family Health History Multi-Stakeholder Workgroup has previous defined the minimal core data set that should be used when representing structured FHI (32). Additionally, a well structured HL-7 messaging standard has been developed for the effective interoperable communication of FHI between electronic record systems (33).

A systematic approach for the ascertainment, representation and communication of FHI has previously been described as being crucial a national clinical decision support infrastructure and for the adoption of genome-informed and personalized medicine (34). To meet this requirement, polled primary care clinicians appear receptive to patient-generated FHI (15), but have also requested tools to help them in interpreting it (35). The following section reviews the previous attempts at building online tools for the ascertainment of FHI directly from patients, and why they do not solve the needs of the busy primary care clinician.

## EXISTING SOLUTIONS

The Surgeon General's My Family Health Portrait, developed through the US Centers for Disease Control and Prevention, http://www.hhs.gov/familyhistory/, is perhaps the most notable publically available web-enabled tool for collecting FHI. Created in 2004, it has over 18,000 unique visitors per month and collects FHI related to six common complex conditions: heart disease, stroke, diabetes and colon, breast, and ovarian cancers (36). When compared to the FHI ascertained by a genetic counselor, the validity of the FHI obtained via this tool has been supported for four of these conditions: diabetes, colon cancer, breast cancer and ovarian cancer (36). This tool is currently designed as an intermediary model for patient-entered FHI as it does not support storing or integration of the FHI into an individual's PHR. Collaboration and communication of entered FHI is also not an innate feature of the Surgeon General's tool, but could be accomplished by downloading an extensible markup language (XML) document and emailing it to family members. More recent improvements have enable storing the FHI obtained via the tool onto Microsoft HealthVault®. Owens, et al, performed an evaluation of the Surgeon General's My Family Health Portrait tool and found numerous suggested improvements (37). Users, most of whom were medical students, suggested enhancements to improve the usability of the site. For instance, users thought that the pre-select range for the age of onset for health conditions was too broad and there was no mechanism in place to represent a more precise age when known. Users also wanted an expanded list of medical conditions.

Zimmerman et al described their work towards building a Collaborative Medical History Portal (ChMP) (38). Their information schema was based on the Genealogy Data Communication (GEDCOM) plain text file format. Collaboration was accomplished by granting permissions to other users to perform direct editing of shared tree structures. ChMP does not

seem to have been developed further than proof-of-concept and is not actively maintained or integrated into a clinical workflow.

The Family Healthware Impact Trial showed that 34% of patients presenting to primary care clinics were at strong-to-moderate risk of breast colon or ovarian cancer using an interactive online 128-question survey that also captured FHI (39). However, the Family Healthware tool only collected FHI regarding six common diseases (stroke, diabetes, coronary artery disease, breast cancer, colon cancer, and ovarian cancer).

The Health Heritage® website built by the University of Virginia in Charlottesville covers 89 health conditions, but took respondents between 1 and 120 hours to complete (40,41).

The Genomedical Connection, a consortium between Duke University, the University of North Carolina at Greensboro, and the Moses Cone Health System in Greensboro, NC developed a FHI capture and clinical decision support analysis tool call meTree® (42). It collects FHI regarding 48 conditions while providing decision support to primary care clinicians for four conditions: breast cancer, colon cancer, ovarian cancer and thrombosis.

Given the disperse ways by which health events are communicated among families, and the diverse health literacy among family members, it is postulated that the current approaches are insufficiently robust to represent diverse health concepts and thus are not generalizable or scalable for the primary care setting. For instance, the existing solutions ask the user about a limited number of pre-determined lists of health conditions. If one were to select "Cancer" on the Surgeon General's My Family Health Portal, a sub-list of only 20 categories of cancer are presented to the user. The MyFaV website aims to solve this problem with a robust solution that is capable of representing all diseases, facilitating future reconciliation areas of agreement and disagreement among family members, and communicating it to the clinician.

## CONCEPTUAL FRAMEWORK

The central argument presented in this thesis is that FHI is fundamentally subjective and, in principle, hearsay information. Intrinsic in this notion is the idea that FHI is an aggregation of assertions made by informants, each with their own independent observations and interpretations of the health conditions affecting the family. These assertions are influenced by the health literacy, comprehension and communication of the involved family members. In order to properly represent both the information and the workflow, it is important to build a framework that takes this approach into consideration.

Based on this perspective, the information model uses the informant, or source of the information, as a key element associated with each assertion regarding a health condition, along with a date-time stamp and the representation of the subject, or relative, whom the assertion is referencing. The subject identifier (relative_id) may reference the informant, in which case the informant is making an assertion about himself or herself. If shared among a network of users, having an informant that references the unique identifier of a particular user allows for the controlled dissemination of information among users, yet retains the source of the original assertion. In future iterations of this tool, this will be used to facilitate agreement and disagreement among a family of users. This appears to be a unique approach among FHI acquisition tools and more closely represents the actual flow of information among family members.

In general, the CRUD (Create, Retrieve, Update and Delete) framework (43) was followed when building structured FHI and storing the data elements into persistent storage in MySQL tables. In order to adequately characterize all aspects of the data elements and the methods necessary to build a robust information model, a task decomposition analysis was performed. Representative tasks where created to better understand the data elements and

processes that contribute to FHI. The following list includes some of the representative tasks that involve generating structured FHI:

Create:

- Represent the canonical family members in a pedigree
- Add new family members with proper assertions regarding placement in the pedigree
- Add demographic information to a relative including name, date of birth, alive status
- Add health conditions for a given relative specifying an age of onset

Retrieve:

- Identify family members and location in the pedigree
- Identify the name of relatives in a family
- Determine how relatives are related to self
- Determine if a given relative is alive or not
- Identify demographic information including age/date of birth, age/date of death

Update

- Change the demographic information for a given relative
- Change the alive status of a given relative

Delete

- Remove a relative from the pedigree (presumably was added in error)
- Remove an assertion regarding a health condition for a given relative

In reviewing these representative tasks, it was decided that updating an assertion regarding the age of onset for a given relative was equivalent to removing the assertion and adding a new assertion, thus changing the age of onset was not included in the Update category. Instead, in order to accomplish this task, the user deletes the assertion and creates a new one with the modified specification regarding the age of onset.

Unlike the Surgeon General's My Family Portrait tool, which has a limited and poorly categorized list of health conditions, the MyFaV program aims to provide a robust representation of health conditions using a structured and hierarchical terminology, enabling FHI to be more generalizable and computable. Having a structured and semantic conceptual representation of the health conditions also allows for reconciling agreement and disagreement among family members, synonym matching, and the application of next-generation clinical decision support algorithms. To facilitate this aim, the user is presented with a textbox to enter a health condition affecting a specific relative. After the user types the first three characters into the textbox, she is presented with an auto-completion suggestion box. This tool exploits the parsimony of the user-interface for users to click or otherwise select one of the suggested concepts saving keystrokes. This action subsequently categorizes the entered condition with a unique identifier based on a controlled terminology. Free text representation of concepts is still accepted if the user chooses not to select one of suggested items. Subsequently, an attempt is made to map the entered text of the health condition to a unique concept identifier in a controlled terminology of diseases. Specifically, the items presented to the user in the autocompleting suggestion box are limited to the concepts found within the Human Disease Ontology (44). The Human Disease Ontology is an open-source aggregation of human disease containing over 8,500 unique health conditions with related hierarchical structure. The MyFaV server hosts a local version of the Human Disease Ontology provided by the National Center for Biomedical Ontology, *www.bioontology.org* (44). If the user fails to click on one of the suggested concepts in the autocompletion suggestion box, the entered free text is mapped to one of the following Unified Medical Language System (UMLS) version 2010AA terminologies (45): Systematized Nomenclature of Medicine--Clinical Terms (SNOMED-CT), Online Mendelian Inheritance in Man (OMIM), or MedlinePlus, in that order. Mapping is restricted to semantic types that are appropriate to represent health conditions. The full list of semantic types used to represent health conditions is shown in table 1. Should the entered string successfully map to a concept in the

UMLS, the text changes to green; if not, the text changes color to red. This gives the user a subtle feedback regarding the entered text.

| Semantic Types |
| --- |
| Disease or Syndrome |
| Finding |
| Neoplastic Process |
| Mental or Behavioral Dysfunction |
| Acquired Abnormality |
| Anatomical Abnormality |
| Therapeutic or Preventive Procedure |
| Gene or Genome |
| Congenital Abnormality |
| Sign or Symptom |
| Injury or Poisoning |
| Pathologic Function |

**Table 1. The semantic types of the Unified Medical Language System (UMLS) concepts that were used to represent health conditions**

## KEY DESIGN DECISIONS

A web-based solution was chosen as the platform for deploying this tool. This choice allowed the tool to be available to the largest number of users without installing specific software. It was hoped that the website would be viewable using all Internet browsers; however, inconsistences were found in the way Microsoft's Internet Explorer interprets the scalable vector graphic (SVG) extensible markup language (XML) that was use on this site. Thus, for the purpose of the study evaluation, users were restricted to Mozilla's Firefox, Google's Chrome or Apple's Safari, as the combination of these three browsers represent 80.2% of all Internet browser traffic (46).

Based on the Research-Based Web Design & Usability Guidelines published by the U.S. Department of Health and Human Services, the following is a partial list of additional design decisions (47):

- Optimized display for screen resolution of 1024 x 768 pixels. This will cover at least 75% of users.
- Most users do not have a fast Internet connection. The site should have a small data footprint to enable slow connection speeds. Most webpages on the site contain approximately 75kB of data and the largest page contains under 500kB. This keeps the page load time to under 4 seconds on a broadband connection and under 1 minute using a 56kps modem connection.
- Employ an elegantly simple design interface and navigation menu types.
- Enable quick data entry of important tasks to overcome the barrier of adoption.

FHI is presented to the user on the MyFaV site via a list of relatives (figure 1) as a

Hypertext Markup Language (HTML) table, or via a graphical pedigree (figure 2) drawn as a Scalable Vector Graphic (SVG). The table view consists of three main columns: names of relatives along with their age, how they are related to the user, and a list of health conditions asserted about each relative.

| NAME (age) | RELATIONSHIP | HEALTH INFORMATION |
|---|---|---|
| **Immediate Family Members ( and their offspring )** | | |
| John Appleseed<br>(57 years old) | self | ○ hypertension (age: 50 to 59)<br>(+) add |
| Junior<br>(20 years old) | son | (+) add |
| Jimmy<br>(19 years old) | son | (+) add |
| Jessica<br>(17 years old) | daughter | (+) add |
| Judy<br>(16 years old) | daughter | (+) add |
| Jack Appleseed<br>(81 years old) | father | ○ heart attack (age: 60 to 69)<br>○ coronary arteriosclerosis (age: 60 to 69)<br>○ hypertension (age: 50 to 59)<br>(+) add |
| Margaret Appleseed<br>(86 years old) | mother | ○ skin cancer (age: 70 or older)<br>(+) add |
| Jill<br>(48 years old) | sister | ○ hypertension (age: 40 to 49)<br>(+) add |
| James<br>(53 years old) | brother | ○ hypertension (age: 40 to 49)<br>(+) add |

**Figure 1. Representation of the Family Health Information in an HTML table view**

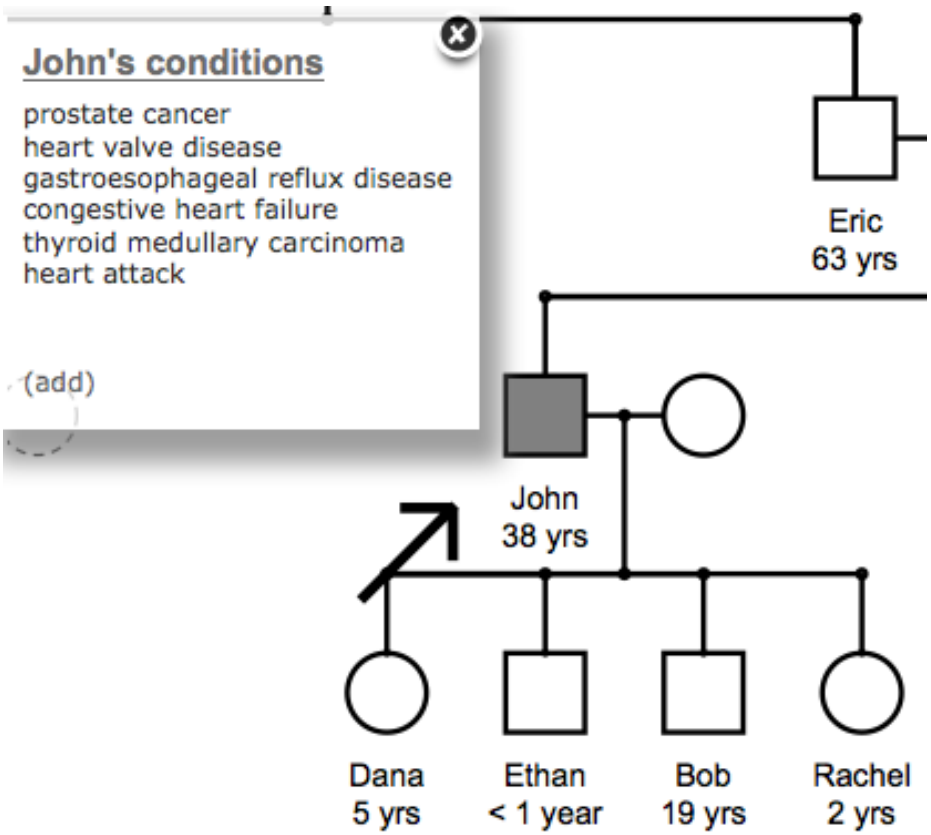


**Figure 24. Graphical representation of the Family Health Information in the pedigree view using Scalable Vector Graphics**

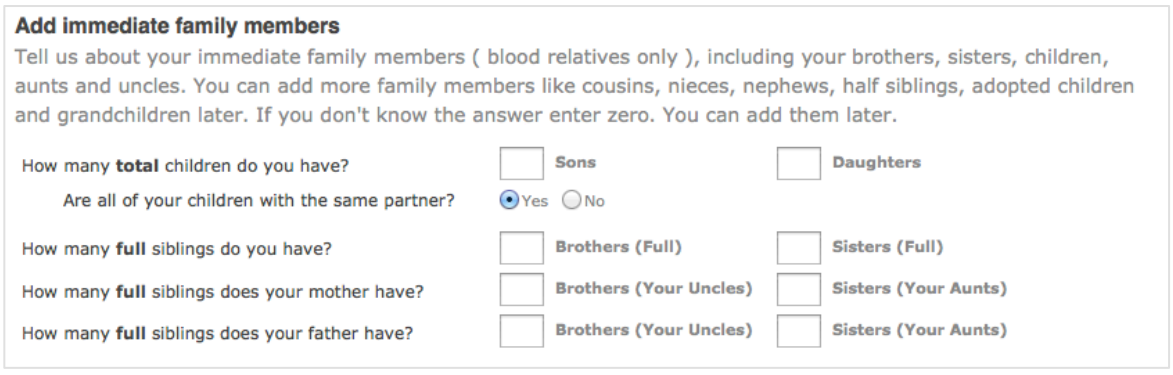


**Figure 3. Rapid building of the scaffolding of a family structure using 8 questions**

To facilitate rapid building of the family structure a strategy similar to the Surgeon General's My Family Portrait was used. After new users register for the MyFaV site and are

asked to build their own FHI, they are asked a series of questions aimed to build the foundation of their family structure. As shown in figure 3, a majority of the family structure can be ascertained by asking eight questions. This process creates scaffolding upon which the remaining relatives can be added later. Additional relatives, including half-siblings, are added to the pedigree structure by adding them as children, asserting the parent to whom they are related. This is accomplished by clicking the "(+) add son" or "(+) add daughter" button associated with the relative in the table view, figure 4. The user is then presented a modal window and must specify the complementary parent by choosing a name or creating a new partner of this relative. Users must assert at least a first name for this partner before proceeding. This step helps identify this partner in the model and aids the user when adding additional children. Additional values, such as first name, last name or date of birth, are optional.

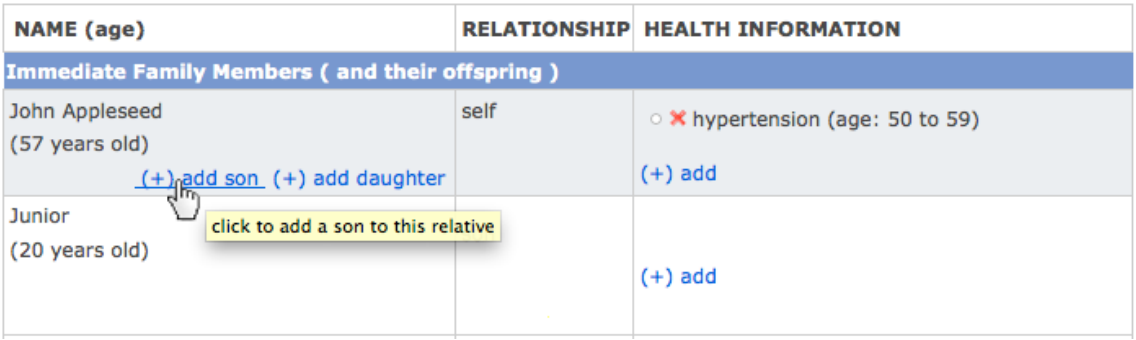


**Figure 4. Text Tool Description displays additional detail when hovering with mouse**

Hover text was used to display additional information regarding possible actions the user could perform. An example is given in figure 4, which shows additional details when the user hovers over the "(+) add son" button. This text tool description also used to convey additional information to the user, such as the parents of a relative in the list of family members as shown in figure 5 depicting Margaret as John's mother and Jack as John's father.


**Figure 5. Additional information is presented to the user using**

Another unique design decision was creating the ability to quickly assert the number of partners that a user shares offspring with and the number of sons and daughter with each partner. This is accomplished using the question "Are all of your children with the same partner?" when asking the total number of children the user has. If this is answered in the negative, then additional rows are shown allowing the user to assert the name of each partner and the number of sons and daughters shared with this person. Subsequently, the total number of sons and daughters is hidden. The name of each partner with whom the user shares offspring is explicitly asserted, ensuring a tight information model. This information will become necessary when reconciling relationships or sharing pedigree information in future iterations of this site.

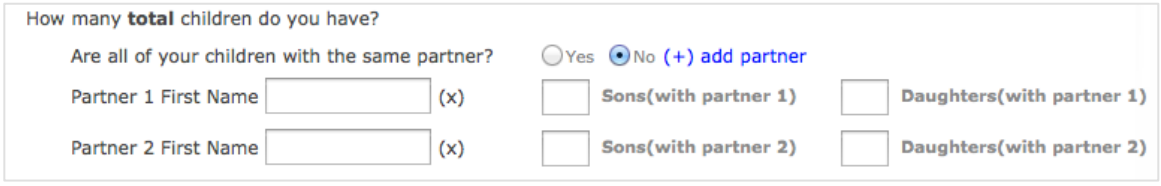


**Figure 6. Specifically identifying the number of children and with whom the user shares offspring**

To conserve screen space, icons and symbols were used to consolidate information and present possible actions to the user. These were carefully chosen to best represent the task that

the user was performing. For instance, when a user is adding a new relative to the pedigree, as in figure 4, the symbol "(+)", a plus sign enclosed in opening and closing parenthesis, was chosen to enable the association of the symbol with a Creation function. Additionally, in the same figure, a red shape in the form of an X was used to depict the task of Deletion. The symbol "(x)", an x enclosed in parenthesis, was used in a variety of places in the site to represent a Cancel function. An example is shown in figure 7 with the mouse pointer hovering over the "(x)" symbol, which when clicked allows the user to cancel the current option for selecting the precision of the asserted age for this relative. Additional details were provided as a text tool description, "cancel/change precision", when the user hovers over the symbol. This symbol was chosen because it is similar, but distinctly different from the close icon of the modal window (seen in the top right hand side of figure 7 as an x enclosed in a circled). These icons have the fundamentally similar function of escaping the current select. In the case of the modal window, the icon closes the window and in the case of changing the precision of the age attribute, it escapes the current selection.


**Figure 7. Modal window editing demographic information. "(x)" symbol used to escape from the current selection**

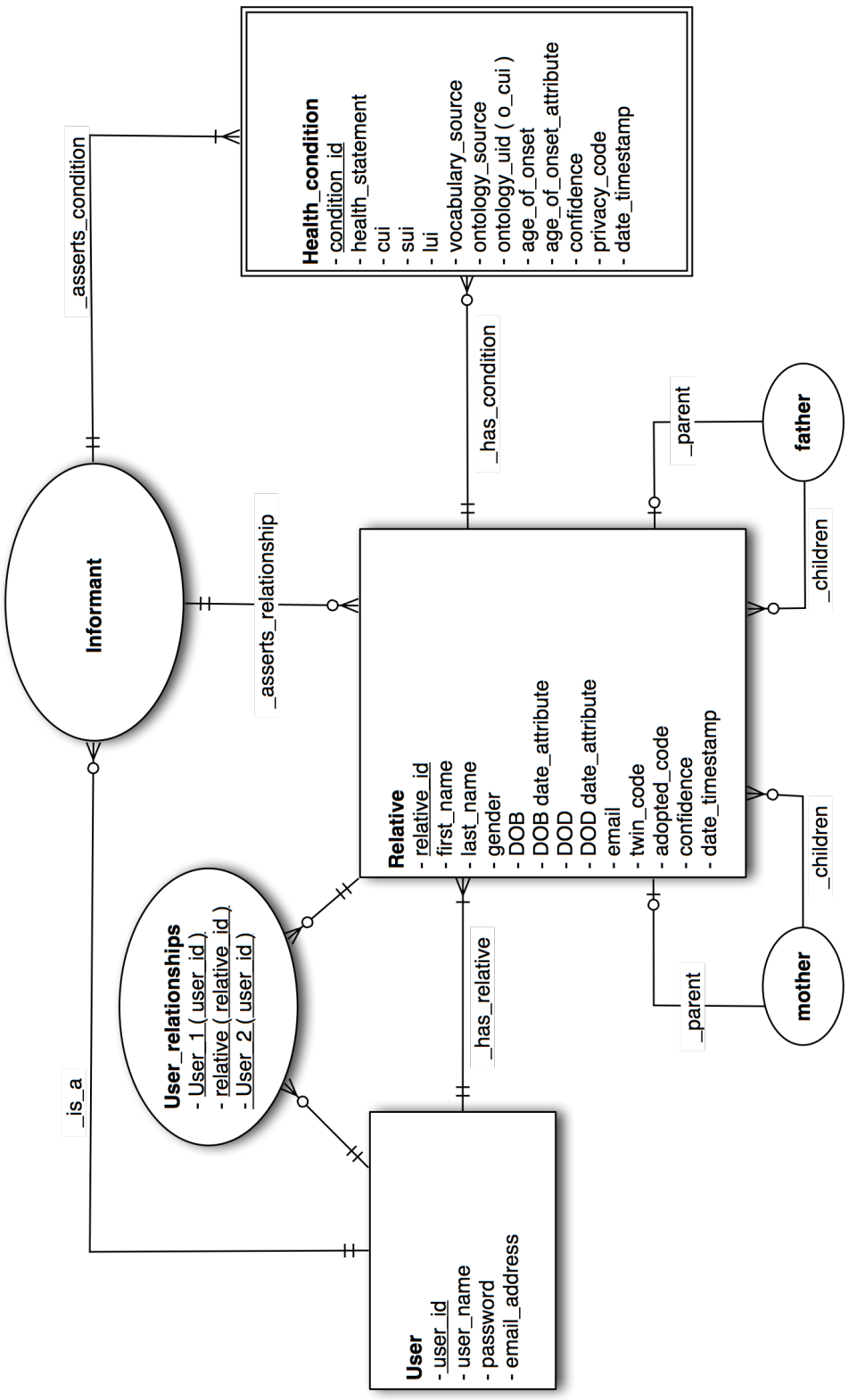


**Figure 8. Information model of the MyFaV tool**

# INFORMATION MODEL

The Genetics in Primary Care workgroup has previously described key characteristics of an ideal family history tool (9). And the American Health Information Community's Family Health History Multi-Stakeholder Workgroup has previously described the minimum core data set necessary for a standard communication of FHI (32). In addition, that report suggests the minimal data elements that are necessary to optimally represent FHI in the Electronic Health Record (EHR) and communicate this information between systems. These recommendations were taken into consideration when building the information model presented in this thesis. Some data elements, such as ethnicity, were excluded to simplify the evaluation, while others, such as consanguinity, are handled computationally in this model.

MyFaV uses a relational information model composed of a binary tree to represent the family structure (pedigree). Each relative is assigned a unique relative_id (rid). The assignment of a relative_id to the attribute mother or father in the model depicts the maternal and paternal lineage and is the backbone of the pedigree structure. The relative_id for self, parents and grandparents are canonically assigned and hard-coded into every pedigree model. As shown in figure 8, the relative_id of self (the user) is always assigned the value 1, increasing incrementally up to the maternal grandmother, with the value 7. Additional relatives are added incrementally to these initial values, and a zero value denotes a leaf node indicating the mother or father is not specified. For example, in this model, parents of a user's grandfather may not be specified (great-grandparents from the perspective of self) and the mother and father attributes for the grandfather is assigned the value of zero, to represent the null value. Likewise, if a relative does not have children then no relative in the pedigree has this relative's relative_id in the mother or father attribute.

In order to allow the appropriate amount of uncertainty when representing the date of birth, date of death and the age of onset for a health conditions, this information model allows for

various levels of precision when making a particular assertion. To facilitate this in the information model, a date_attribute field is utilized and is associated with the specified date value in the table. The possible values for this attribute associated with the date of birth and date of death are: 'exact_date', 'est_year', 'age_years'. The value 'exact_date' denotes that the value for the assertion is known to the exact date. For example, a date of birth of May 12th, 1992 with a date_attribute of 'exact_date' is meant to state that the precision of the date of birth is asserted to the level of the exact day. When a date of birth is estimated to the precision of a year, such as 1992, the date_attribute of 'est_year' and the value represented in the database is "07-01-1992". Here, July 1st is choosen as it is the mid-point of the year representing a median value, yet still allowing for date calcuations to be performed. Finally, for an estimated age using the date_attribute of 'age_years', the date represented in the database is the date the assertain was made minus the number of years being asserted. For instance, if today is June 1st, 2012 and if one were to assert that a given relative is 20 years old using the precision date_attribute of 'age_years' then the date inserted into the database as the date "06-01-1992". Additional levels of precision are allowed for the age of onset of a health condition for relatives. The age of onset date_attribute utilizes the exact_date, est_year, age_years as well as an attribute called 'decade', which loosely is associated with a general age category. Using this date_attribute the user can assert the age of onset to the precision of a decade. For example, a user may assert that a given relative had hypertension with an date_attribute of "decade" and a date associated with the assertion of 50 to 59, indicating the fifth decade of life. Additional possible values for the age of onset when the age of onset attribute has the value 'decade' are: prebirth, newborn, infancy, childhood, adolesence, 20-29, 30-39, 40-49, 50-59, 60-69 and 70 or older.

Health conditions are represented in this information model as a weak entity type. As such, they include the parent entity, in this case, the relative for which the assertion is being made. An auto-incrementing field, condition_id (cid), is used along with the relative_id (rid) and user_id(uid) to uniquely represent the primary key of this table. Ideally, an auto-incrementing

primary key should be avoided to enable maximal scalability. Initially an attempt was made using the user_id, relative_id, health_statement and age of onset as primary key for this table; however, this failed to adequately represent the case of multiple health events, such as multiple heart attacks, in an estimated year. This created a dilemma in representing conditions in both an appropriately semantic manner versus the greatest degree of accuracy. Since this evaluation was expected to be relatively small, a decision was made to use an auto-incrementing field as part of the primary key.

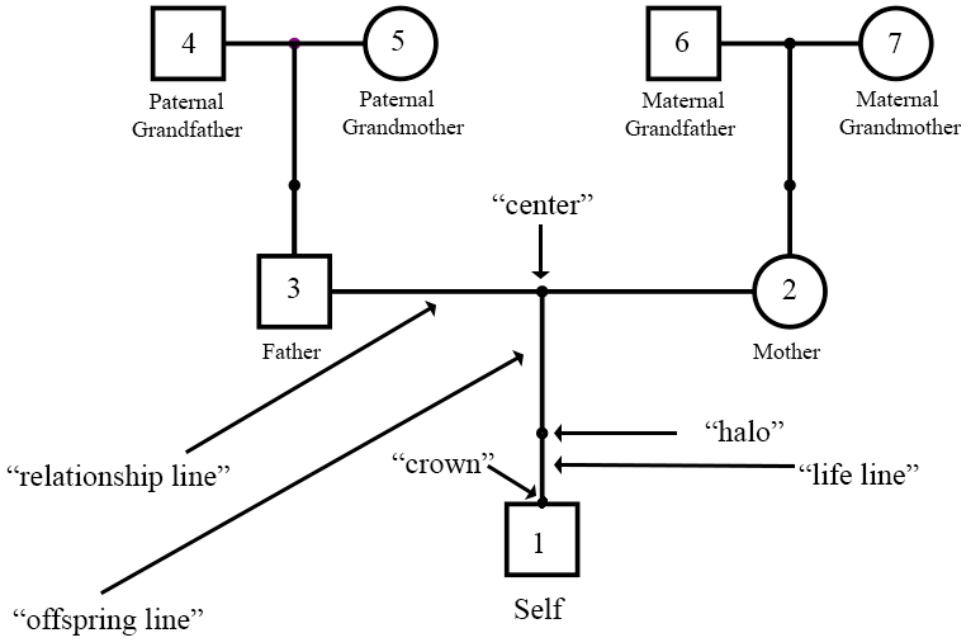


**Figure 9. Graphical representation of the pedigree structure indicating elements. Canonical relative are labeled**

Intrinsic in an accurate information model to represent the pedigree structure is the ability to graphically display that model in an unambiguous manner to the user. The Pedigree Standardization Task Force (PSTF) has published numerous guidelines for a standard representation schema for drawing pedigrees (48–50). An earnest attempt was made to adopt all

aspects of these standards and to adapt them to one suitable for viewing and interacting with via an Internet browser; however, not all elements of the guidelines could be implemented. For instance, due to time limitations, representing twin status and consanguinity were outside the scope of this initial development cycle, but will be prioritized in future iterations of the site. Graphically drawing the pedigree was accomplished using the HTML extension of the World Wide Web Consortium (W3C) standards for Scalable Vector Graphics (SVG). SVG is an application of the extensible markup language (XML) and is used to represent vector-based graphical structures. The graphical representation of the pedigree was translated to XML using the Raphael SVG jQuery library, version 1.5.2 (51).

JavaScript Object Notation (JSON) was used to exchange data between the server and client machines. JSON is a lightweight text-based model for representing data structures and is often used when passing data back and forth from servers via an Asynchronous JavaScript and XML (AJAX) requests (52). AJAX was used to transfer the standardized pedigree data and FHI to the client browser. The JSON object that contains all of the structured standardized pedigree data including all health conditions for each relative has a size of 9.5kB. To display the information in the browser window, the object was parsed and used to populate a specific html  tag of the MyFaV site. This approach enables maximal flexibility and adaptability when implementing the MyFaV onto future platforms. Formatting of the HTML content was managed using Cascading Style Sheets (CSS v.3) (53).

# ALGORITHMS

The information model presented collapses to a Directed Graph. When no consanguinity or inbreeding is present, the graph is a Directed Acyclic Graph (DAG). A simple test of consanguinity (inbreeding) could have been to check if the graph contained cycles. DAGs are amenable to recursive algorithms for data retrieval. This is particular useful for three aspects of constructing the FHI contained within the MySQL tables: populating the HTML table of family members, drawing the pedigree structure into a Scalable Vector Graphic, and determining relationships between relatives. Finally, an algorithm for determining the similarity of health conditions concepts within a family will be presented.

As mentioned above, all data was retrieved from the server as raw JavaScript Object Notation (JSON). On the page load, a JavaScript function parsed this data object, extracting the structure of the pedigree data and constructing properly formatted HTML elements. An HTML table row, or "<tr>" element, was constructed for each relative and populated with demographic information, relationship type and the list of health conditions.

The pseudo code for performing this function, called '_print_tr()' is shown :

*Print the table row (<tr>) for self*
*If self has a relative_id > 7 or = 1 then find children of self and*
    *For each child, _print_tr()*

This function is performed for each parent of self, sibling, half-siblings, mom's siblings and dad's siblings, mom's half-siblings and dad's half-siblings. This generates a complete three-generation pedigree for self and included all relevant family members and offspring.

A similar strategy was used when drawing the graphic representation of the pedigree into a Scalable Vector Graphic (SVG). First, the pedigree was grounded on a center point that was calculated from half the height and half the width of the browser window. This point, called 'center', serves as reference to the rest of the pedigree as it is being drawn. The focal point of the pedigree was set as a reference to 'self' identifying the current user; however, the algorithm is

dynamic and can focus the pedigree on any another relative. Each of the relative's SVG elements contained the following parts depicted in figure 9: halo, crown, line, name, and relationship name. These structures served as anchor points in the SVG and were handled as JavaScript objects and thus were called by reference. Based on published guidelines (50), the drawing algorithm represented each relative as either a square, if male, or circle, if female. If the 'self' object has children then a generic SVG element, depicted as a dashed line, was drawn representing this partner as a complementary gender. Children were then drawn recursively adjusting the sibling line in proportion to the number of children and returning an onset value to the calling function. Next, if self had any siblings, they were drawn to the right of the SVG element; if they had any children, then they were drawn recursively, where self referred to the sibling that was in focus.

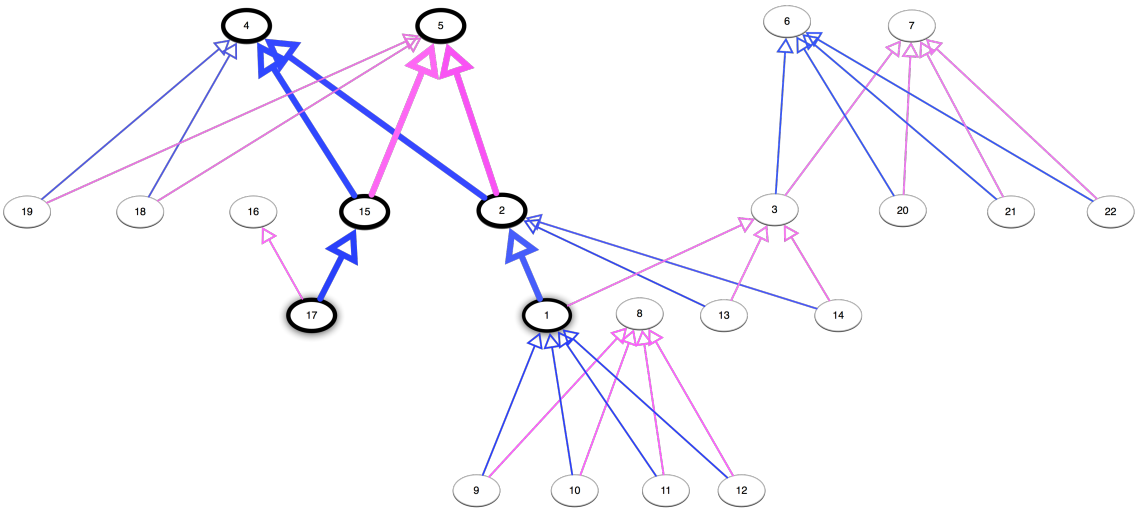


**Figure 10. Directed Acyclic Graph representation of pedigree from standardized profile**

Relationship between two relatives in the pedigree was calculated using a recursive algorithm that navigated the DAG representation of the pedigree as shown in figure 10 (54). First, using recursion, all ancestors of the two relatives being compared were retrieved while retaining the distance information from self. In the example shown in figure 10, relatives 1 and

17 (both shadowed) from the standardized pedigree, John and Jamie from figure 11, are being compared to determine their relationship. In this approach, each relative was also considered as ancestor of himself or herself, but with a distance measurement of zero. Comparisons were made to determine the common ancestors shared between the two relatives. Finally, each common ancestor was analyzed to determine the ancestor(s) with the smallest distance, called the Lowest Common Ancestor (LCA), and was used to calculate the relationship (55). In the example shown in figure 10, relative 1 (John) has the following ancestors (represented as relative_id): 1, 2, 3, 4, 5, 6 and 7 with distance measures of 0, 1, 1, 2, 2, 2, 2 respectively. Relative 17 (Jamie) had the following ancestors (represented as relative_id): 15, 16, 4 and 5 with distance measurements of 0, 1, 1, 2, 2 respectively. The common ancestors in this set are 4, and 5. In this case both have distance measurements of 2 from both relative 1 and relative 17. This information is used to determine that these relatives are first cousins. Consanguineous relatives, in this model, had more than one calculated relationship. For this evaluation, a method to ascertain consanguinity of family members was not explicitly presented to users. For the purposes the evaluation of the MyFaV website, only the lowest common ancestor and hence the closest relationship was used to calculate the assignment of a relationship between a given relative (relative_a) and self (relative_b); however, this information model supports determining this possibility.

When navigating the family tree, it is also useful to reconcile assertions regarding health conditions and determine if the asserted health condition are synonymous. For instance, if a relative were asserted to have the health condition ‘heart attack’ it was useful to reconcile this condition with other relatives who had the synonymous health condition ‘myocardial infarction’. As shown in the information model, health condition assertions have the attributes of cid, cui, lui, sui and o_cui which represents the auto-incrementing condition id, concept unique identifier, lexical unique identifier, string unique identifier and ontology concept unique identifier. These were assigned during a mapping process when the assertion was made and are derived from the UMLS and Human Disease Ontology sources. Reconciling synonymous matching between two

conditions (condition_1 and condition_2) was accomplished using the following approach written in pseudo-code.

*If the id of condition_1 equals the id of condition_2, return true.*
*Else if condition_1 and condition_2 are not the empty string AND the lowercase string of condition_1 is equal to the lowercase of condition_2, return true.*
*Else if the cui of condition_1 and condition_2 are not the empty string AND the cui for condition_1 is equal to the cui of condition_2, return true.*
*Else if the o_cui of condition_1 and condition_2 are not the empty string AND the o_cui for condition_1 is equal to the o_cui of condition_2, return true.*
*Else return false.*

## EVALUATION METHODS

The overall goal of the evaluation was to measure the usability and learnability of the MyFaV website for building structured FHI. This assessment will aid future efforts in focusing resources for iterative improvements.

Usability, in the context of evaluating the MyFaV portal, was defined as the effective and efficient ascertainment of structured FHI from users that corresponds with a high degree of user satisfaction (56). To evaluate the usability of the MyFaV portal, this study asked study participants to answer a series of questions related to a standardized pedigree with associated FHI. The standardized pedigree is shown in figure 10 and is the profile of a fictional person named John Appleseed. The full list of relatives and associated health conditions are listed in the Appendix B. Learnability, in this study, was considered a complementary component of Usability. For this evaluation, users did not have the benefit of training. The Learnability component of this evaluation was determined by the pragmatic test of whether the users were able to perform the given tasks without the aid of instruction; in essence, assessing the intuitiveness of this tool. This evaluation was broken up into three phases. Phase I had users perform representative tasks using a standardized pedigree. In Phase II, users built their FHI using the tool. Phase III consisted of an exit survey.

Using a standardized pedigree for all participants in this study allowed for normalized data and facilitated comparing results among users. Comparing the time required for each user to accomplish tasks associated with their own FHI would have been difficult, given the various size and complexity of individual users' pedigrees. Specific tasks where constructed that reflected the representative tasks users would perform when Creating, Retrieving and Updating their own FHI as explained in the Conceptual Framework section. The task of Deletion was omitted for the sake of brevity. The specific tasks shown in table 2 were tethered one at a time to the MyFaV portal

via an iframe that appeared at the bottom of the user's browser. This iframe contained JavaScript elements that enabled timing of the user's attempt at accomplishing each task. Tasks that had the user retrieve information also had an input field to document and assess the response. The user was not required to enter an answer; thus, this time value represents the time-on-task and not a time-to-task-completion. As the user advanced to the next task by clicking on the 'next' button, a JavaScript function determined if the task was accomplished correctly and submits the time-on-task value for the given step to the server via an Asynchronous JavaScript and extensible markup language (AJAX) post request, where it was stored in an anonymous MySQL table and was later used for analysis.

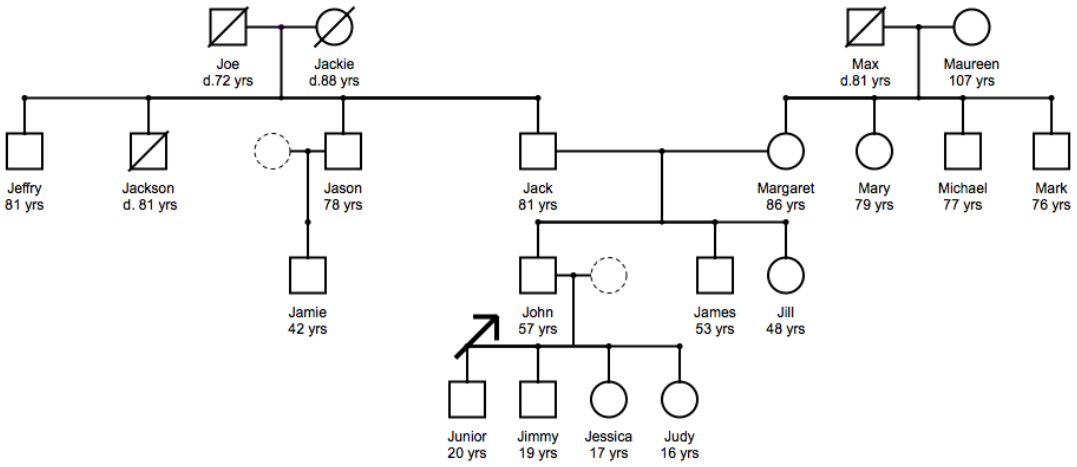


**Figure 11. Graphical view of the Standardized pedigree of fictional user, John Appleseed**

Table 2 shows the specific tasks users are asked to perform for each step in Phase I of the evaluation. Step 1 provided a brief introduction and instructions on how to perform the evaluation. Steps 2 thru 6 centered on asking the user to perform a specific task related to information retrieval of FHI including navigating the family tree and determining demographic information, the number of specific relatives with a certain relationship type, and the number of relatives that share a specific health conditions and the age of onset of a specific health condition

for an individual relative. Step 7 has the user add a new condition to a specific relative. Step 8 had the user update the demographic information of a specific user. In step 9, the user was asked to update FHI and indicate that a relative is now deceased. And finally, users were asked to create a new relative with specific demographic information in a appropriate position within the family tree. The participants were timed for the completion of these representative tasks and the results compared for correctness as a proxy of effectiveness.

| Step | Task |
|---|---|
| 1 | Introduction |
| 2 | How old is John's mother, Margaret? |
| 3 | How many sons does John have? |
| 4 | What is the name of the mother of John's son Jimmy? |
| 5 | How many relatives in John's entire family (including possibly John) have hypertension? |
| 6 | What was the age of onset of John's aunt Mary's Breast Cancer? |
| 7 | Add the condition 'heart attack' with the age of onset of age 50-59 years of age to the list of John's conditions. |
| 8 | John made a mistake regarding his sister's birthday. Change the date of birth for John's sister, Jill, to be the specific date of 5/29/1964 (not the 28th). |
| 9 | Change John's maternal grandmother's (Maureen) status to be deceased on the specific date of 6/10/2012. |
| 10 | John's son, Junior, has a new baby boy with wife, Jen. Add a child to Junior, who's first name is 'Joe' born on 6/10/12. |

**Table 2. Specific task users were asked to perform in Phase I**

Please check the box that reflects your immediate response to each statement. Don't think too long about each statement. Make sure you respond to every statement. If you don't know how to respond, simply check box "3".

| | Strongly Disagree | | | | Strongly Agree |
|---|---|---|---|---|---|
| 1.) I think that I would like to use this website frequently. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |
| 2.) I found this website unnecessarily complex. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |
| 3.) I thought this website was easy to use. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |
| 4.) I think that I would need the support of a technical person to be able to use this website. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |
| 5.) I found the various functions in this website were well integrated. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |
| 6.) I thought there was too much inconsistency in this website. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |
| 7.) I imagine that most people would learn to use this website very quickly. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |
| 8.) I found this website very awkward to use. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |
| 9.) I felt very confident using this website. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |
| 10.) I needed to learn a lot of things before I could get going with this website. | ☐ 1 | ☐ 2 | ☐ 3 | ☐ 4 | ☐ 5 |

11.) Overall, I would rate the user-friendliness of this website as:

| ☐ Worse Imaginable | ☐ Awful | ☐ Poor | ☐ OK | ☐ Good | ☐ Excellent | ☐ Best Imaginable |
|---|---|---|---|---|---|---|

**Figure 12. System Usability Scale survey**

In Phase II of the evaluation the participants were asked to generate their own FHI, as thoroughly as possible. This time-on-task was recorded and stored for later analysis. Phase III consisted of feedback regarding the usability of the MyFaV site to build structured FHI by asking the users to complete a modified 10-item System Usability Scale survey combined with an 11$^{th}$ item overall adjective assessment of the usability of the MyFaV portal, similar to that described by Bangor et al (57).

Subjective assessment of the generalized usability of the MyFaV portal was ascertained from users via an 11-item System Usability Scale (SUS) survey using a five point Likert scale for questions 1-10 (figure 11). Item 11 used a seven item adjective rating scale which asked the user to rate the overall user-friendliness of the website similar to that described by Bangor et al (56). A total calculated SUS score ranges from a value of zero to 100 in 2.5 increments. A SUS score of 100 indicates a perfect score. Ideal scores for each of the SUS questions alternate between each of the ten questions. Odd questions, items 1, 3, 5, 7 and 9, have an ideal response of "Strongly Agree", whereas even questions, items 2, 4, 6, 8, 10, have an ideal response of "Strongly Disagree". Numeric values are assigned to each of the Likert scale responses from 1 for "Strongly Disagree", to 5 for "Strongly Agree". To calculate the total SUS score, the score contribution of the odd items were determined by taking the response Likert value minus 1, and the even items were determined using 5 minus the response Likert value. The overall SUS score was determined by multiplying the sum of the individual items' contribution by 2.5 (56). The adjective rating score (item 11) was determined independently of the 10-item SUS score using numeric values ranging from 1 to 7, to represent to range from "Worse imaginable" to "Best Imaginable". While Bangor et al (58) have shown that there is a great deal of correlation between the 10-item SUS survey and the 11-item adjective assessment score, including the 11-item adjective assessment aids in communicating the significance of the results by giving a *gestalt* representation of the responses in common language.

As suggested by the U.S. Department of Health and Human Services' Research-Based Web Design & Usability Guidelines, effectiveness was defined as the number of correct responses by users over the total number of attempts (47). A goal was set to have 90% of users able to perform each of the tasks outlined above correctly. A reasonable goal for efficiency was to have the average time it takes for users to build their own structured FHI be less than 16 minutes, which is the average time it takes in current clinical practice in the primary care setting (5). The goal for the Usability survey was a SUS score of over 70 with an adjective assessment of at least 'Good'. As Bangor pointed out, this is in the "Acceptable" range (58).

During the pilot phase of this evaluation, a selected cohort of self-identified super-users from the *www.myhealthatvanderbilt.com (MHaV)* patient portal were used as proxy users to provide initial feedback of the MyFaV website. The goal was to have at least twelve users provide feedback and test the site for gross errors. This cohort of users has previously participated in focus groups and evaluation studies for the *MHaV* patient portal. Additionally, users from the Department of Biomedical Informatics were solicited directly via email. All users contacted directly via email were provided a link to the MyFaV portal to register as new users or were given credentials to access the site. In order to register for the site, each new user was presented with an electronic consent to participate in this research study and users were required to agree before proceeding. This study was approved by the Institutional Review Board (IRB); study #120093. Feedback was solicited from these users regarding the general usability of the *MyFaV* site provided direct feedback via email and telephone. In total, 22 users piloted the site before the actual evaluation. Based on this feedback modifications were made to the site and the evaluation process was optimized.

Soliciting of research participants was from the campus-wide *research.notifications@vanderbilt.edu* list-serve email announcements associated with the Research Notifications distribution list. To encourage participation in the research study, a raffle

for an iPod touch (estimated value $200) was performed at the conclusion of the study period. Those users who completed all aspects of the exit survey were eligible to enter the lottery for the iPod touch by entering their name and email address into a form that followed the 11-item adjective assessment score. The goal was to have 100 users participate in the study. The evaluation period lasted 1 week, from June 21st, 2012 at 5pm till June 28th, 2012 at 5 pm. Potential study participants followed the URL in the body of the email and were directed to the *MyFaV* portal. New users registered for the site by creating username and password and entering basic demographic information. Participants were required to consent to participate in this research study in order to proceed to further areas of the portal (see Appendix A). Documentation of consent was stored as part of the registration process. Research participants were informed that their reported FHI would NOT be utilized to make clinical decisions and would NOT be communicated to their clinicians or other family members. Additionally, research participants were informed that the evaluation would ONLY focus on the usability of the *MyFaV* portal and its ability to ascertain and represent family relationships and the asserted health conditions of these family members made by the user. Once authenticated, users started with Phase I of the evaluation based on the standardized pedigree. The users were also able to enter the email addresses of each of their family members when building their FHI and send an email inviting relatives to also participate in this research study evaluating the *MyFaV* portal. Relatives of users who were invited to the site received email invitations with the structure, "Your relative, <first_name_of_inviter>, has sent you an invitation to participate in the My Family at Vanderbilt portal." During the present evaluation process, individual users were NOT able to see the information entered by other family members. Aggregated data of the family clusters created during this viral expansion of invited users, however, was available to the researchers for the purpose of comparing individual asserted FHI to the aggregated FHI of the entire family unit that agreed to participate in this study. The database was queried in a manner that hid any identifiable fields associating Protected Health Information (PHI) with individual users.

All information, including Protected Health Information (PHI), was stored on a secure database associated with the MyFaV portal in a locked server room on the 6th floor of the 2525 West End Building of the Vanderbilt Campus. The server utilized a SSL secure certificate for communication and a robust firewall. All hard drives on the server were encrypted. At the conclusion of the evaluation period, all PHI stored on the server was erased. The winner of the lottery was determined by selecting a random row from the MySQL table and was notified via email. In order for the winner to claim the prize, he or she was required to provide identifying information including full name, residential address (for 1099 reporting) and tax identification number (Social Security Number), pursuant to the Vanderbilt University reporting guidelines.

At any time during the evaluation period, participants were able to withdraw from the research study. This was accomplished by users editing their account profile and clicking on the delete account via the online portal or by sending an email to the Principal Investigator. If a user deleted his/her account, all information related to that individual was dropped from the MySQL tables. The number of users who withdraw from this study was kept and reported. User-initiated relatives that became new users of the MyFaV site were retained.

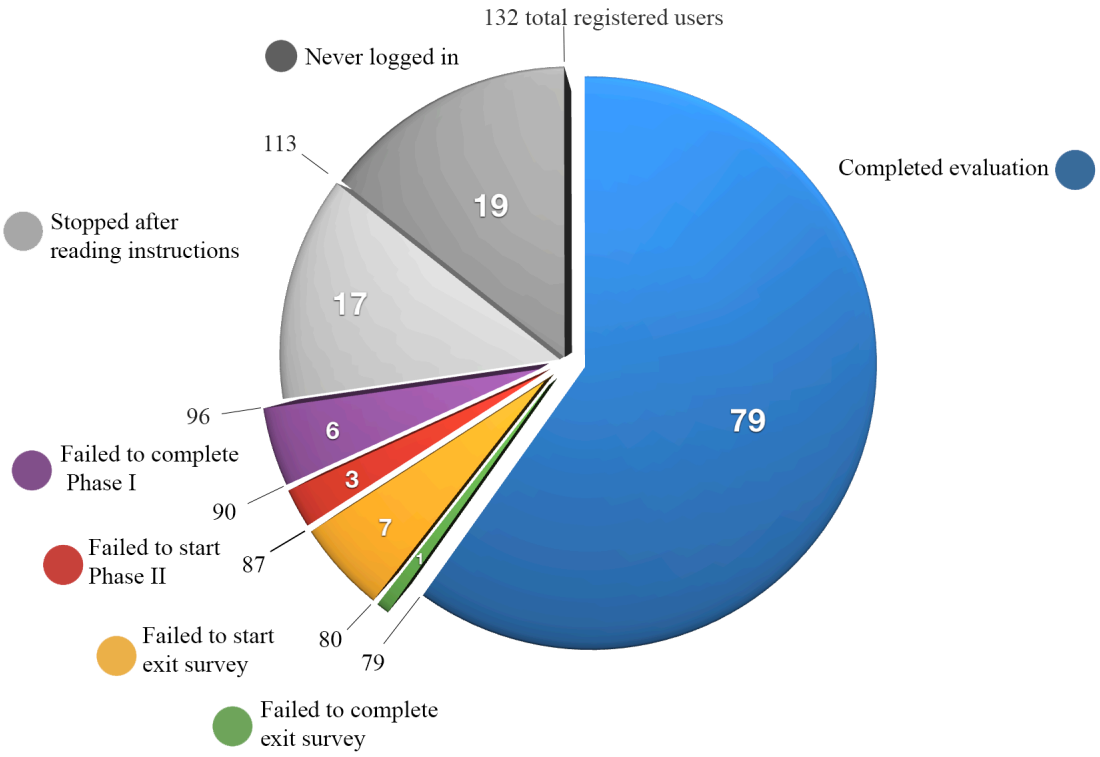


**Figure 13. Number of total registered users and how many completed each phase**

In total, during the evaluation period, 132 new users registered for the site; of these, 97 (73.5%) were female and 35 (26.5%) were male. The age distribution of the users is shown in figure 13, with a mean of 40 years of age. The minimal user age of user was 20 and the maximum was 69. The data shows outliers with age of zero, which is likely an artifact of the user not entering a proper age during the registration process. Of the 132 new registered users, 19 never logged into the site, leaving 113 active users who started the evaluation. After reading the introduction and instructions, 17 users failed to proceed, leaving 96 users that started Phase I of the evaluation (performing representative tasks using the standardized family of John Appleseed).

There was a small attrition of 6 users during the ten steps that comprised Phase I of the evaluation, leaving 90 users to start Phase II of the evaluation (building their own FHI using this site). In total, 87 users created a pedigree in this phase of the evaluation. Seven users failed to complete the exit surveys and one user started the first part of the exit survey completing the 10 item System Usability Survey but failed to complete the 11th item, adjective rating. This left a total of 79 users who completed all aspects of the user evaluation of the site. A total of 7 invitations to relatives were sent from 3 users of the MyFaV site to participate of this evaluation, but none of the invitations were accepted.

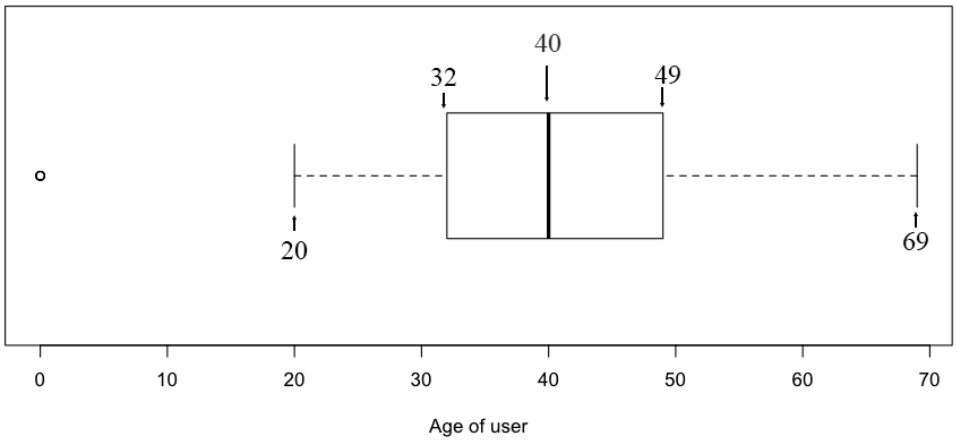


**Figure 14. Boxplot diagram depicting the distribution of the age of users**

Figure 14 shows the distribution of the time-on-task of users for each step of Phase I as a measure of the efficiency of this tool. These results include data from the 22 pilot users. The specific tasks associated with each step can be found on table 2. As noted above, 96 users started Phase I of the evaluation, comprised of the ten steps shown. Of these, 90 users completed all ten steps with an attrition of 6 users.

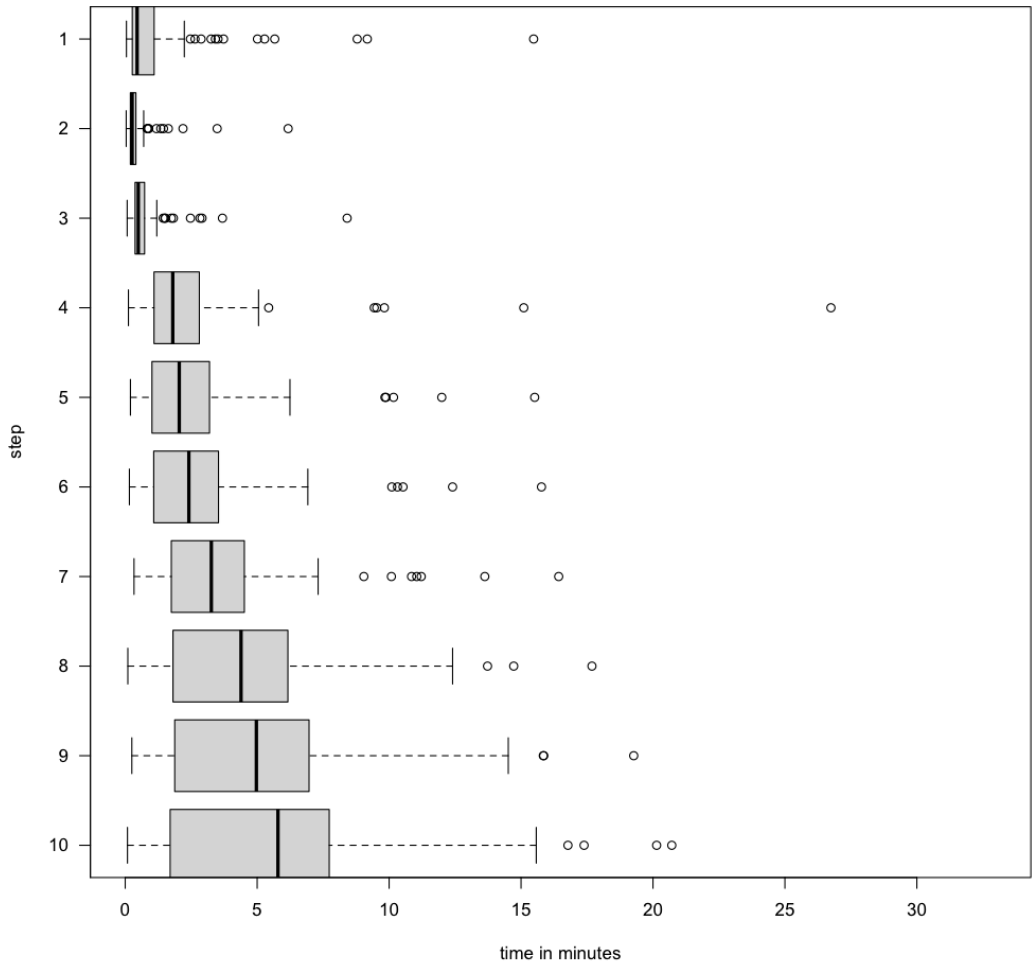


**Figure 15. Efficiency results for each step of Phase I measured in minutes**

Figure 15 shows the results of the effectiveness of the MyFaV tool when analyzing the accuracy of users performing the specific tasks in Phase I on the evaluation and includes data from the 22 pilot users. Alongside the bar graph of each of the steps performed by the users is a percent correct value. The lowest percent correct is for step 4, which asked the user to retrieve information regarding the name of the mother of one of John Appleseed's children, Jimmy. Possible reasons for this result are discussed later. The next lowest percent correct is step 8, which asked the user to change the date of birth of John's sister to a specific date, requiring the user to change the precision of the assertion.

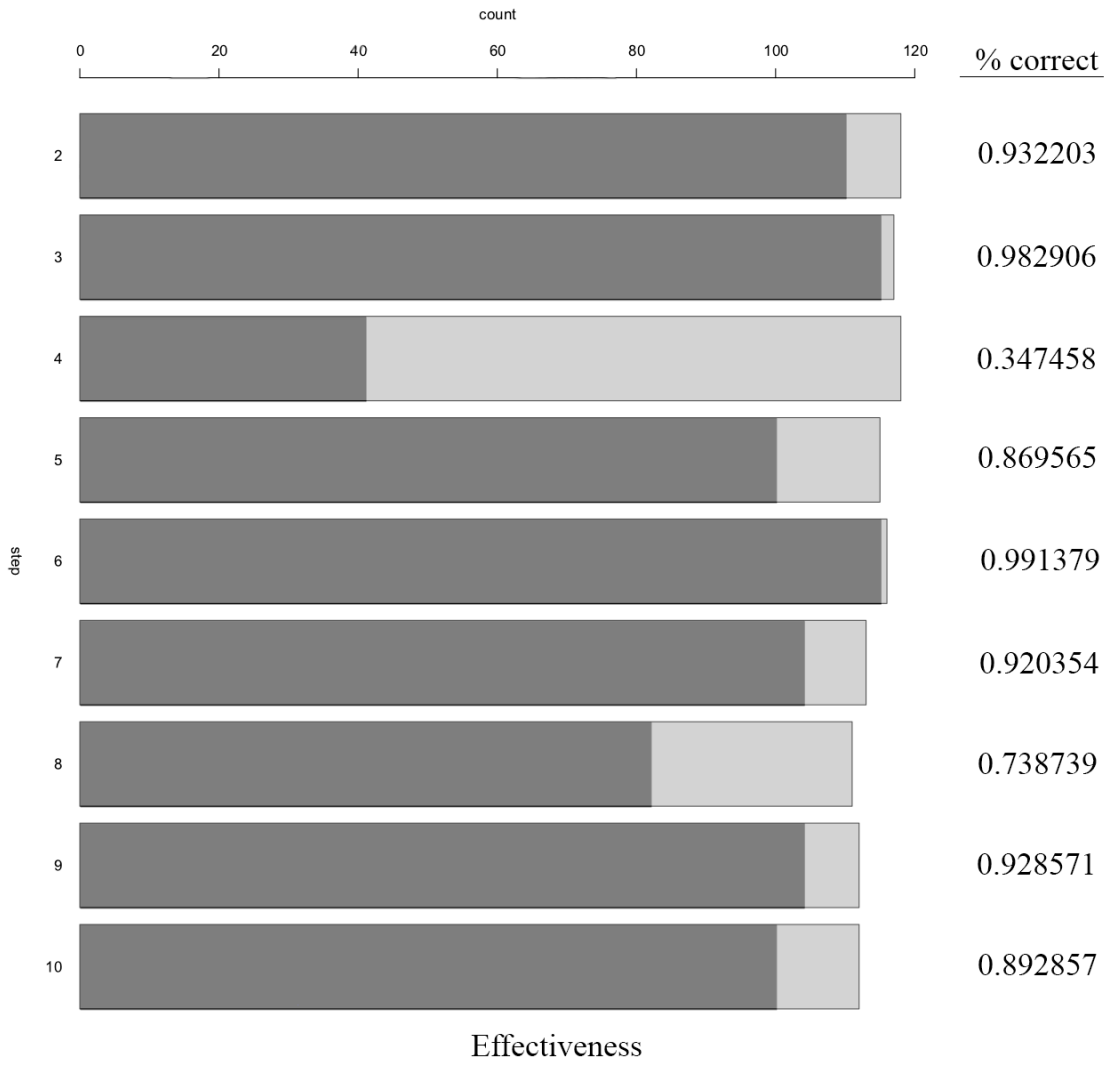


**Figure 16. Effectiveness results of Phase I showing percent correct for each step**

Figure 16 shows the individual Likert scale scores for each of the ten questions involved in the System Usability Scale (SUS) survey found in figure 12. A score of 5 denotes 'strongly agree', while a score of 1 denotes 'strongly disagree'. Perfect scores in the SUS survey alternate between 'strongly agree' (5) and 'strongly disagree (1)', where the odd questions are in the affirmative and the even scores are in the negative (56). Results are shown as boxplot diagrams with the median of the distribution depicted as a dark bar. Outliers are noted as open circles and the minimum and maximum scores as whiskers.

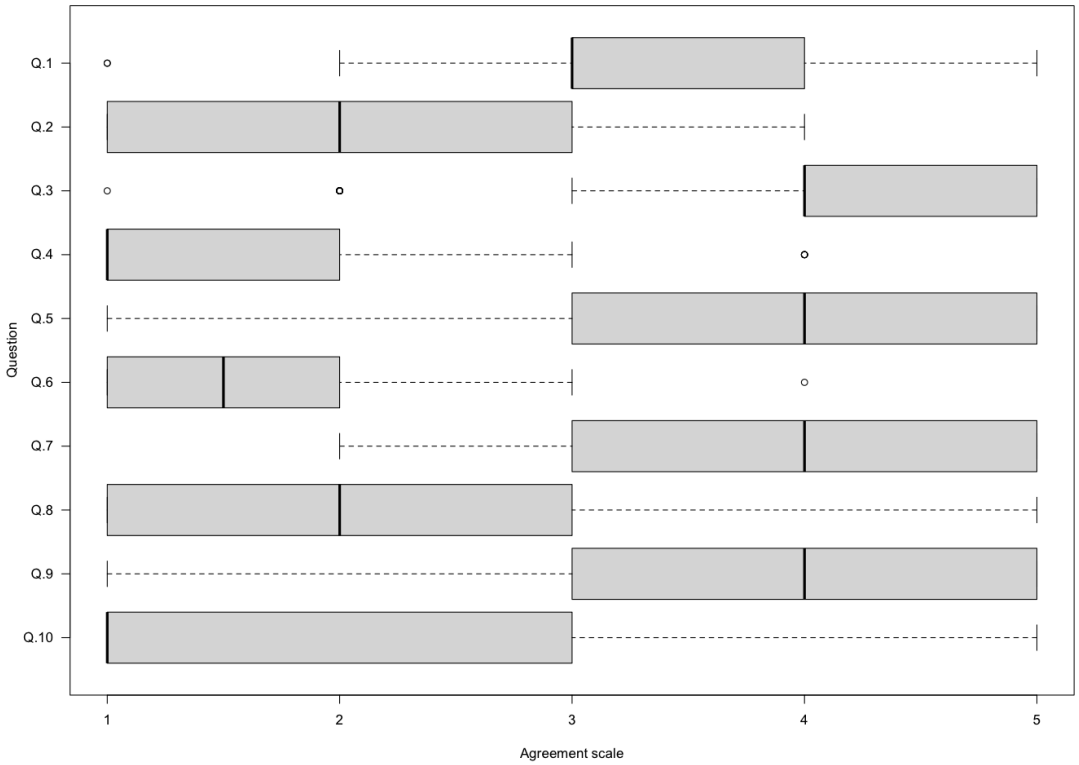


**Figure 17. Distribution of System Usability Scale responses plotted as boxplot diagram**

Figure 17 displays the composite results as a calculated System Usability Scale score of the first 10 items of the System Usability Scale survey. The mean of the scores was 74.55 (depicted as a red square), the median was 76.25 (depicted with a dark bar) with a minimum score of 37.5 a maximum score of 100 calculated from a total of 80 respondents.

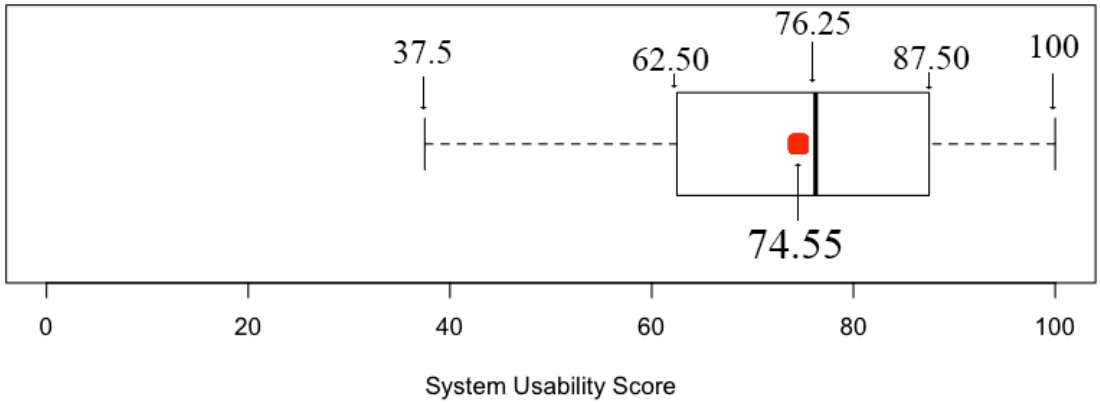


**Figure 18. Distribution of the calculated System Usability Scale score as boxplot diagram**

The result of the Adjective Rating Score, 11th item of the modified System Usability Scale survey, which was also used as a gestalt measure of usability in this study, is shown in figure 18. This part of the exit survey asked the user, "Overall, I would rate the user-friendliness of this website as:" with the following seven options: Worse Imaginable (1), Awful (2), Poor (3), OK (4), Good (5), Excellent (6), or Best Imaginable (7). The median score for this evaluation was found to be 6 (Excellent) with an average score of 5.338 (between Good and Excellent) from a total of 79 users who completed this part of the exit survey. The minimal score was three (an outlier) and the maximum score was seven. The correlation of the numeric score to the adjective scale is also shown.

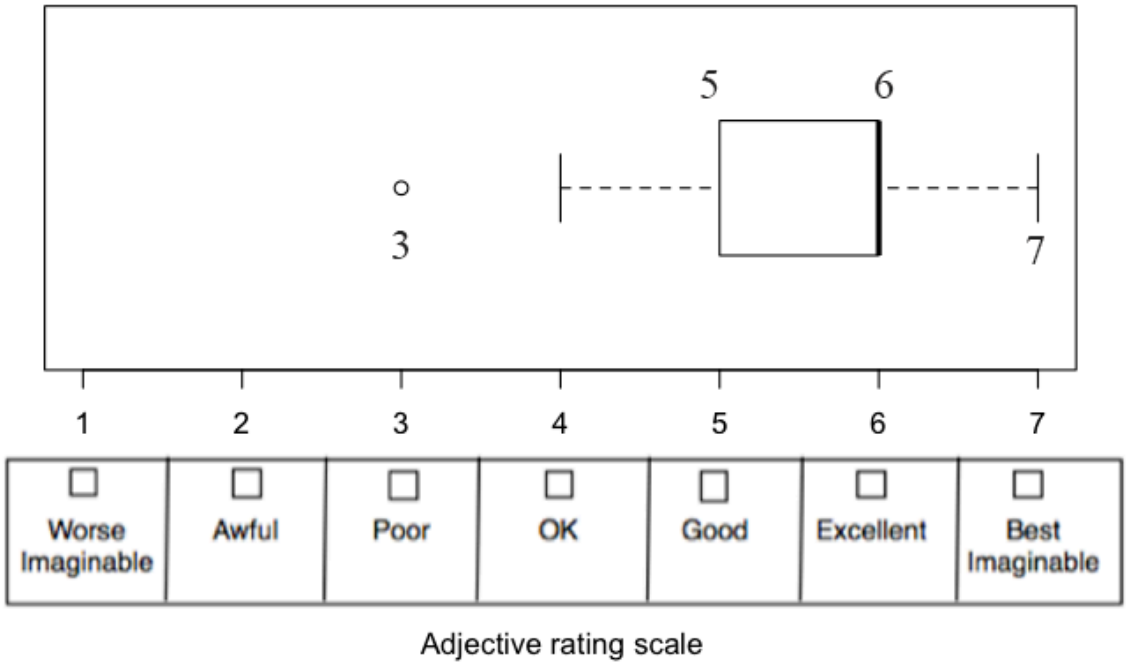


**Figure 19. Distribution of Adjective Rating score plotted as boxplot diagram**

Phase II of the evaluation (time spent in building a personal FHI) is shown in figure 19. Users spent an average of 8.35 minutes (depicted as a red square) creating there own FHI during Phase II of the evaluation with a median time of 6.46 minutes. The minimal time was 0.0418 minutes, or 2.5 seconds and the maximal time was 35.0206 minutes. As seen in the diagram, there are a handful of outliers in the range of 25-35 minutes. It seems likely that those who spent 2.5 second did not build a thorough amount of FHI. The graph shows that it took the majority of users less than 10 minutes to create their FHI using the MyFaV website.

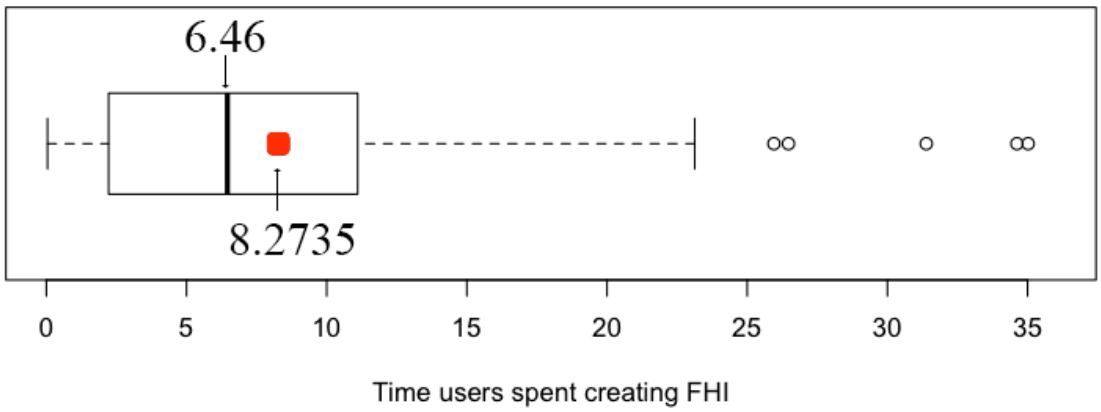


**Figure 20. Distribution of the time users spent creating their own FHI plotted as boxplot diagram**

The distribution of the total number of relatives added by each user is shown in figure 20. The mean number was 17.94 with a minimum of 9 and maximum of 66. The least number of relatives possible in the system is 7. This would include the user as self, two parents and two sets of grandparents.

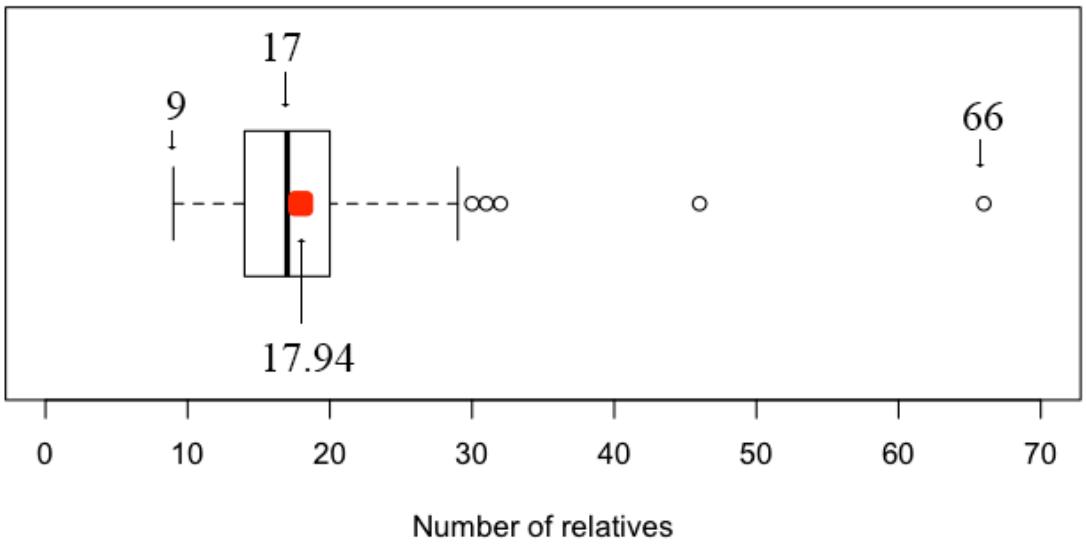


**Figure 21. Distribution of the number of relatives created to their own pedigree as boxplot diagram**

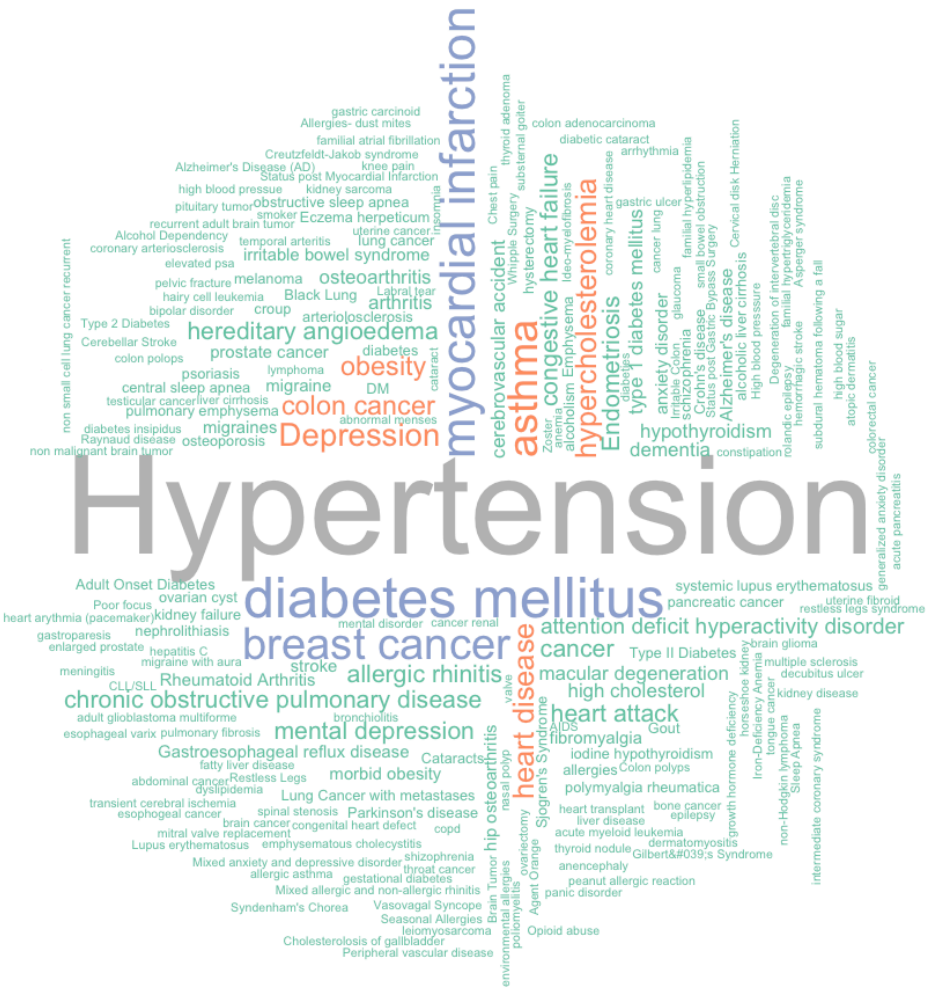


**Figure 21. Wordcloud representation of the health conditions entered by users. Font size is representative of relative reporting frequency.**

Figure 21 depicts a 'Wordcloud' of all of the health conditions that where created by users of the MyFaV website. The font size of the word represents the relative reporting frequency of the concept. As shown in this diagram, Hypertension is clearly the most frequent condition entered by users. There were 140 distinct conditions that where represented less frequent and are shown in green. Some editing of the health conditions was necessary in three health conditions entered by users to remove potentially identifiable health information, such as, the hospital or region that a procedure was performed. The distribution of the number of health

conditions entered by each user is shown in figure 22. The mean number of conditions was 9.64 with a minimum number of conditions was 1 and a maximum number of conditions was 34.

| Count | Health condition | CUI |
|---|---|---|
| **52** | **Hypertension** | C0020538 |
| **19** | **Diabetes Mellitus** | C0011849 |
| **16** | **Myocardial Infarction** | C0027051 |
| **14** | **Breast Cancer** | C0006142 |
| **13** | **Asthma** | C0004096 |
| **9** | **Depresssion** | C0011581 |
| **8** | **Hypercholesterolemia** | C0020443 |
| **8** | **Heart Disease** | C0018799 |
| **7** | **Colon Cancer** | C0699790 |
| **7** | **Obesity** | C0028754 |

**Table 3. Top ten conditions entered by users showing frequency and corresponding concept unique identifier (cui)**

Table 3 shows a list of the aggregated top 10 conditions added by all users for family members along with the respective counts for each and the Concept Unique Identifier (CUI) from the UMLS. In aggregate, users entered 'hypertension' as a health condition for themselves and relatives 52 times. The frequency was not determined for individual users.

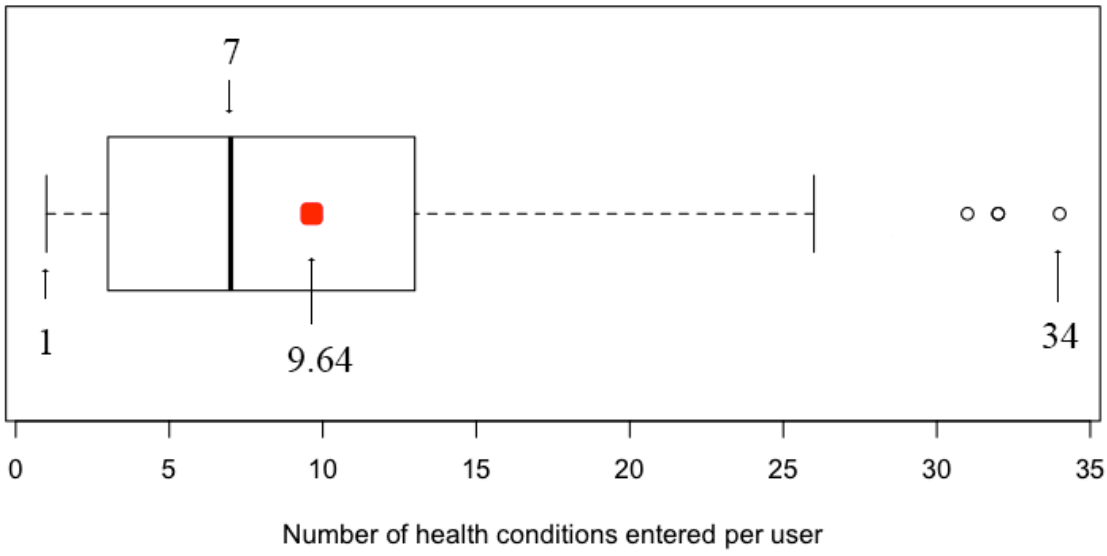


**Figure 23. Distribution of the number of health conditions entered by each user plotted as boxplot diagram**

There were 29 health conditions out of a total of 241 distinct health conditions that were entered by users that failed to map to concepts in either the UMLS or the Human Disease ontology. In total, 94% of the total 482 health conditions entered by users of this tool mapped to a unique concepts. The conditions that did not map to either the Human Disease Ontology or UMLS concepts are listed in table 4. Some of these entered health conditions failed to map to concepts due to minor spelling errors, such as “Diabeties”, “esophogeal cancer”, “colon polops”, “high blood pressue”.

| Health condition |
| --- |
| Sjogren's Syndrome |
| Iron-Deficiency Anemia |
| Allergies- dust mites |
| Alcohol Dependency |
| Status post Gastric Bypass Surgery |
| Lung Cancer with metastases |
| Black Lung |
| Agent Orange |
| Syndenham's Chorea |
| High blood pressue |
| Heart arythmia (pacemaker) |
| High blood sugar |
| Valve |
| Diabeties |
| Labral tear |
| Abnormal menses |
| Migraines |
| Environmental allergies |
| Subdural hematoma following a fall |
| Alzheimer's Disease (AD) |
| Ideo-myelofibrosis |
| Esophogeal cancer |
| Crohn's disease |
| High cholesterol |
| Non malignant brain tumor |
| Horseshoe kidney |
| Colon polops |
| Seasonal Allergies |
| Cancer lung |

**Table 4. List of 29 conditions that did not map to concepts using this approach**

# DISCUSSION

This thesis developed and evaluated the MyFaV website, an online portal for the ascertainment of structured FHI directly from users. Seventy-nine users completed all aspects of the evaluation. The majority of users were female, with a median age of 40, which supports previous findings regarding those most likely to maintain their FHI (59). This section summarizes the key findings of the evaluation, the informatics contributions and limitations, as well as the potential for clinical applications and knowledge discovery.

Efficiency results of this evaluation are difficult to interpret and are likely skewed due to the composition of the research participants and the unforeseen environmental conditions in which it was performed. Users were emailed an announcement to participate in this evaluation throught the Vanderbilt University research notification listserv, which was in most cases a work email address. Based on a review of server access logs, most participants appear to have been performing the evaluation while at work and were possibly multi-tasking. To maximize enrollment, users were not specifically told to only focus on this evaluation, nor were they explicitly told they would be timed while attempting to complete it. This was specifically done to prevent a Hawthorne effect (60), but may have adversely prolonged the time-on-task results.

The correctness of users' responses to tasks was used as a metric to evaluate the tool's effectiveness. These results highlight specific areas where this tool could be improved. As stated previously, the effectiveness goal was a correctness score of at least 90% for each of the specified tasks. The result for Task 4, 34.8% fell significantly short of this goal. This task, which had users retrieve information regarding the mother of one of John's children, was challenging for most. The task was purposely included in the evaluation as pilot users also experienced difficulties with this task. After receiving feedback from pilot users, improvements were made to the design of the user interface. Specifically, to provide a larger target area that provided this information, the title attribute that contained this text was changed from the  tag to the

entire <td> tag under the relationship description column. To correctly complete this task, users needed to mouse-over the relationship column and wait for half a second to get the text from the title attribute of the <td> tag to display. Alternatively the user could have clicked on the "(edit)" button where the information was displayed at the top of the modal window. Even with the modifications made after receiving feedback from pilot users the issue remained and further enhancements to the user interface are obviously needed. The result of Task 8 also illuminated an opportunity for improvement. The score of 74.9% fell slightly below the goal of 90%. This task had users update the status of the date of birth for John's sister, Jill. This task required users to change the level of precision by clicking on the "(x)" symbol. This action allowed users to cancel the current selection and then modify the level of precision to be an exact date of birth, effectively changing the data attribute to 'exact_date'. Users seemed to have difficulty understanding the role of the "(x)" symbol in canceling and changing the precision of this field despite the use of a text tool description. The text tool displayed "cancel/change precision" when the user hovered over the "(x)" symbol with the mouse (see figure 7). This issue seems identical to the difficulty experienced with Task 4 where users also failed to adequate recognize the text description that provided additional details contained within the title attribute of the <td> tag and displayed on mouse-over. These results highlight areas for future improvements.

Despite these focal areas for future improvement, the results of the user-satisfaction aspect of the evaluation were very favorable. While individual SUS responses are not meaningful on their own (61), they generally trended towards the ideal, with the exception of question 1, which had a median response of 3. This question has the users respond to the statement, "I think that I would like to use this site frequently." In retrospect, this seems to be the most appropriate response from the standpoint of a patient. To put this perspective in context, pilot users gave the following feedback: "Why would I want to see the doctor regularly?" and "Isn't the purpose of this tool so I don't have to enter [FHI] frequently?" Given that the SUS survey is a well-tested and validated tool (61), the text of this statement was not modified for this evaluation.

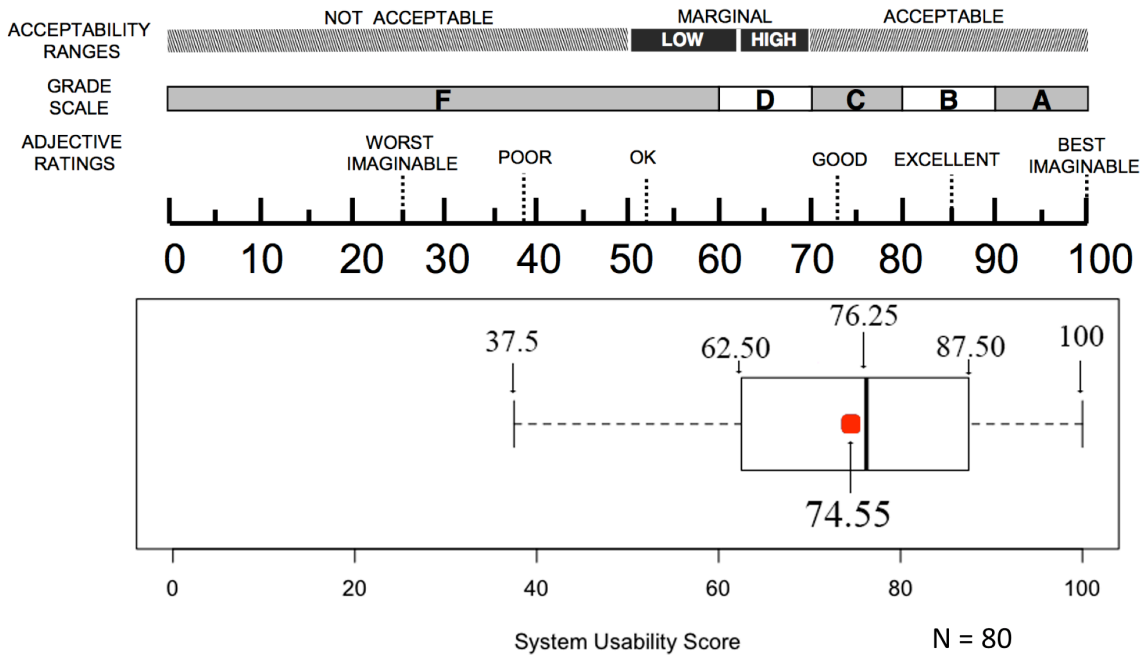


**Figure 24. Distribution of System Usability Scale results plotted beneath scale for interpretation. Image adapted and used with permission (58)**

Using Bangor's interpretation and comparison regarding the significance of the SUS score (58), the distribution of the calculated SUS score, including the median (76.25) and mean (74.55), are plotted beneath a scale of the letter grade score, the adjective ratings and acceptability range in Figure 23. Overall, the mean score seems acceptable and this correlates with the distribution of the results for the adjective assessment score from figure 18 with a median score of 6, "Excellent" and a mean score of 5.33, "Good" to "Excellent". In general, this is interpreted as an acceptable score given that this project represents the first iteration of a new tool. As Bangor pointed out, the use of an adjective assessment score is complementary to the SUS score and provides a common language label regarding its significance and should be used together to create a clearer picture of this tools overall usability. (58). In this context, the MyFaV site appears to have scored a solid letter grade of 'C' when assessing its usability. This correlates with the "Good to Excellent" adjective score with both results found to be in the acceptable range.

The average amount of time required for users to enter their FHI using the MyFaV site was 8.274 minutes. This was well within the goal of 16 minutes. More impressive is the richness of the conditions and the average number of relatives that were added by users. Figure 21 shows a graphical representation of the health conditions entered and their frequency. The top ten conditions, also listed in table 3, correlate with the top chronic conditions seen in the primary care setting (62). Moreover, of the 482 total health conditions entered by users of this tool, 94% mapped to unique concepts. A few of the conditions listed in table 4 are obvious misspellings that could be included in future attempts to correct the user's input and better categorize erroneous entries. Other conditions listed seem to represent complex concepts, such as "subdural hematoma following a fall", or lay terms for health conditions such as "high cholesterol" and "black lung". The goal of the present study was not to find or develop the most robust terminology for representing health conditions for lay-users, but the data from this tool could be used to refine future consumer health vocabularies.

The work presented in this thesis makes several novel informatics contributions in the area of ascertaining FHI directly from patients. First, this appears to be the first patient-facing tool for capturing structured FHI using an autocompleting search box. Moreover, the implementation of this approach appears sufficiently expressive in representing a range of health condition concepts. Next, the combination of a stricter identification of the parental lineage with subtle User-Interface design features enables a tighter information model compared to existing solutions. It also uses a concept-based representation of health conditions enabling the reconciliation of synonyms and presenting these to users via the suggestion box and Graphic User Interface. The combination of these innovations allowed for the capture of structured FHI that took users an average of only 8½ minutes to complete. The novelty of these approaches has the potential to offset the workload of ascertaining FHI away from the busy clinician while maintaining acceptable levels of usability and structured representation.

The implementation of this approach, however, does have several informatics limitations. First, the gradual attrition of users is concerning as there is no clear understanding of the underlying cause. It is postulated that given the broad distribution of the time-on-task results, that users were multi-tasking and distracted during the evaluation period; however, this could not have been determined without directly observing the users during the evaluation. The fact that it took some users over ten minutes to read the instructions suggests that some users might not have been prepared to perform the evaluation or tolerate the time required to complete it. On the other hand, these results are offset by the average time required for users to build their FHI using this tool, 8.274 minutes. This finding combined with an average number of relatives entered, 17.94 and the average number of conditions entered, 9.64, appears to be moderately comprehensive.

This thesis does make important potential clinical contributions. First, being able to offset the workload of busy clinician and automate the task of ascertaining FHI in a structured manner directly from patients is the first step in making FHI more accessible for clinical decision-making. Next, having the FHI in a structured format allows it to be used when building next-generation clinical decision-making algorithms. This has the potential to reduce morbidity and mortality by enabling informed preventative strategies to those patients identified as being at increased risk. It also enables the building of future methods to reconcile areas of agreement and disagreement among a network of family members using this system and between the patient and clinician potentially improving the accuracy of FHI.

This approach has important potential clinical limitations as well. Despite its historical and widespread use, the clinical utility of FHI has not been adequately studied in the primary care setting (10). Moreover, the amount of information generated by this approach may be, in fact, overwhelming to busy clinicians who are already overwhelmed by the genetic contributions of disease (20). Clinicians may ultimately choose to reject FHI information as they may feel a sense of liability to interpret the clinical significance of all entered data (35). These arguments,

however, underscore the need for the development of automated interpretive algorithms using the structured FHI (63,64).

This tool also has significant research contributions. Research studies that use FHI have decreased largely due to the amount of time required to ascertain it structurally (65). Having a tool, such as this one, that offsets the task to individual users yet represents health conditions in the family using a structured and systematic manner may rekindle the use of FHI for knowledge discovery (66).

In summary, the MyFaV website provides several important informatics, clinical and research contributions including an improved framework for representing the health conditions entered by users and tightens the information model of the pedigree structure. This tool facilitates offloading the workload from busy clinicians and may ultimately enable FHI to be better utilized for clinical decision-making and future risk assessment algorithms.

# CONCLUSION

The problem that the MyFaV tool attempted to solve in this thesis was to determine feasibility of ascertainment of FHI directly from patients in a systematic and structured manner. This will enable the more frequent use of FHI for clinical decision-making, both by clinicians in the primary care setting and future risk assessment algorithms. The philosophical priority inherent in this approach is that, before we can reason using FHI, we must have adequate data that is unambiguous and well represented. The MyFaV tool appears to solve this problem, yet, as this evaluation highlights, there are some specific areas where it could be improved. These shortcomings may be an artifact of this project being performed in the scope of a goal-driven Master's thesis rather than following a user-centered iterative design philosophy. The MyFaV tool does appear to solve an important problem facing the perceived clinical utility of FHI. Previous research looking for evidence supporting the clinical utility of FHI in the primary care setting depended upon the determination of what constitutes a "positive family history". In many studies it was analyzed as a dichotomized variable, only representing as positive when FHI was documented in the medical record or if the same condition was found in another family member. Using MyFaV, FHI can be represented systematically and in a structured manner, allowing future analysis that can take its true complexity into consideration, not just answering whether or not it was "positive", but how it was found to be significant. Instead the family history will be reported as "positive" when it was found to represent a significant increased risk of adverse health based on the pattern described using the model presented in this thesis.

The MyFaV approach also allows for the development of clinical decision support tools that are concept-based and more scalable and generalizable than the current tools that focus on a small number of conditions. While this thesis did not specifically evaluate the application of a concept-based representation of FHI to improve the health literacy of family members, the framework does allow linking to additional knowledge repositories and may aid the interpretation

of health conditions. Moreover, this information model facilitates concepts to be reconciled both from the standpoint of synonym matching and by potentially facilitating agreement and disagreement among a network of family members. Finally, beyond the clear clinical application, there is also a role of using the structured FHI ascertained using this model for new knowledge discovery in the context of family-based research.

In summary, this thesis developed and evaluated the usability of a novel approach for the ascertainment of structured FHI directly from patients using an improved information model, user-interface design and representation of health conditions.

# FUTURE WORK

Following a user-centered design approach and using the findings of this evaluation, this research will be extended by making iterative improvements to the MyFaV site, including enabling compatibility with Internet Explorer. Next, this tool will be deployed into clinical practice as part of the MyHealthatVanderbilt (www.myfamilyatvanderbilt.com) website, which is the patient portal of the Electronic Health Record affiliated with Vanderbilt University Medical Center. Presenting the patient-generated FHI to primary care clinicians creates an opportunity to compare how the patient-generated FHI ascertained using this tool compares to that obtained from a clinician and how to best reconcile the two. Ideally, this will involve a randomized intervention that compares use of this tool to that of a control group to determine if having FHI changed clinical decisions. This will involve building scalable, semantically derived clinical decision-support algorithms using the structurally ascertained FHI. Next, enabling the invitation of family members to the MyFaV site will allow sharing of FHI among relatives. This will provide a method to measure the joint probability of agreement of reported FHI among users. The goal would be to reconcile potential conflicting assertions while improving the accuracy and completion of shared FHI. This will also help determine the attitudes these users have about sharing health conditions with their relatives. Finally, the structured FHI will be introduced into research databases and used for knowledge discovery including analyzing the clinical utility of FHI.

REFERENCES


1. Hemophilia A. N Engl J Med. 1994;330(22):1617.

2. Yoon PW, Scheuner MT, Peterson-Oehlke KL, Gwinn M, Faucett A, Khoury MJ. Can family history be used as a tool for public health and preventive medicine? Genet Med. 2002;4:304–10.

3. Nelson HD, Huffman LH, Fu R, Harris EL. Genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility: systematic evidence review for the U.S. Preventive Services Task Force. Ann Intern Med. 2005;143:362–79.

4. Murphy CD, Lee JM, Drohan B, Euhus DM, Kopans DB, Gadd MA, et al. The American Cancer Society guidelines for breast screening with magnetic resonance imaging: an argument for genetic testing. Cancer. 2008;113:3116–20.

5. Rogers J, Durkin M. The semi-structured genogram interview. I: Protocol. II: Evaluation. Family Systems Medicine. 1984;2:176–87.

6. Guttmacher AE, Collins FS, Carmona RH. The family history--more important than ever. N Engl J Med. 2004;351:2333–6.

7. Reid GT, Walter FM, Brisbane JM, Emery JD. Family history questionnaires designed for clinical use: a systematic review. Public Health Genomics. 2009;12:73–83.

8. Scheuner MT, de Vries H, Kim B, Meili RC, Olmstead SH, Teleki S. Are electronic health records ready for genomic medicine? Genetics in Medicine. 2009 May 6;11(7):510–7.

9. Rich EC, Burke W, Heaton CJ, Haga S, Pinsky L, Short MP, et al. Reconsidering the family history in primary care. J Gen Intern Med. 2004;19:273–80.

10. Berg AO, Baird MA, Botkin JR, Driscoll DA, Fishman PA, Guarino PD, et al. National Institutes of Health State-of-the-Science Conference Statement: Family History and Improving Health. Ann Intern Med. 2009;151:872–7.

11. Silberberg J, Fryer J, Wlodarczyk J, Robertson R, Dear K. Comparison of family history measures used to identify high risk of coronary heart disease. Genetic Epidemiology. 1999;16(4):344–55.

12. Murad H, Kalter-Leibovici O, Chetrit A, Freedman LS. A statistical comparison of different family history scores. Statistics in Medicine. 2007;26(14):2785–98.

13. Qureshi N, Wilson B, Santaguida P, Little J, Allanson J, Raina P. NIH State-of-the- Science Conference: Family History and Improving Health. Evidence Report/Technology Assessment No. 186. (Prepared by the McMaster University Evidence-based Practice Center, under Contract No. 290-2007-10060-I.). Agency for Healthcare Research and Quality. 2009;

14. Nussbaum R, McInnes R, Willard H. Thompson & Thompson Genetics in Medicine. Seventh ed. Elsevier; 2007.

15. Fuller M, Myers M, Webb T, Tabangin M, Prows C. Primary Care Providers' Responses to Patient-Generated Family History. J Genet Couns [Internet]. 2009.

16. Acheson LS, Wiesner GL, Zyzanski SJ, Goodwin MA, Stange KC. Family history-taking in community family practice: implications for genetic screening. Genet Med. 2000;2:180–5.

17. Suchard MA, Yudkin P, Sinsheimer JS. Are General Practitioners Willing and Able to Provide Genetic Services for Common Diseases? Journal of Genetic Counseling. 1999;8(5):301–11.

18. Yoon P, Scheuner MT, Gwinn M, Khoury MJ, Jorgensen C, Hariri S, et al. Awareness of family health history as a risk factor for disease--United States, 2004. MMWR Morb Mortal Wkly Rep. 2004;53:1044–7.

19. Murff HJ, Byrne D, Syngal S. Cancer risk assessment: quality and impact of the family history interview. Am J Prev Med. 2004;27:239–45.

20. Frezzo TM, Rubinstein WS, Dunham D, Ormond KE. The genetic family history as a risk assessment tool in internal medicine. Genet Med. 2003;5:84–91.

21. Hughes KS, Roche C, Campbell CT, Siegel N, Salisbury L, Chekos A, et al. Prevalence of family history of breast and ovarian cancer in a single primary care practice using a self-administered questionnaire. Breast J. 2003;9:19–25.

22. Zimmerman M, Coryell W, Pfohl B, Stangl D. The reliability of the family history method for psychiatric diagnoses. Arch Gen Psychiatry. 1988;45:320–2.

23. Andreasen NC, Endicott J, Spitzer RL, Winokur G. The family history method using diagnostic criteria. Reliability and validity. Arch Gen Psychiatry. 1977;34:1229–35.

24. Scheuner MT, Wang SJ, Raffel LJ, Larabell SK, Rotter JI. Family history: a comprehensive genetic risk assessment method for the chronic conditions of adulthood. Am J Med Genet. 1997;71:315–24.

25. Love RR, Evans AM, Josten DM. The accuracy of patient reports of a family history of cancer. J Chronic Dis. 1985;38:289–93.

26. Ziogas A, Anton-Culver H. Validation of family history data in cancer family registries. Am J Prev Med. 2003 Feb;24(2):190–8.

27. Acheson LS, Zyzanski SJ, Stange KC, Deptowicz A, Wiesner GL. Validation of a self-administered, computerized tool for collecting and displaying the family history of cancer. J Clin Oncol. 2006;24:5395–402.

28. Murff HJ, Spigel DR, Syngal S. Does this patient have a family history of cancer? An evidence-based analysis of the accuracy of family cancer history. JAMA. 2004;292:1480–9.

29. Qureshi N, Carroll JC, Wilson B, Santaguida P, Allanson J, Brouwers M, et al. The current state of cancer family history collection tools in primary care: a systematic review. Genet Med. 2009;11:495–506.

30. Leggatt V, Mackay J, Marteau TM, Yates JR. The psychological impact of a cancer family history questionnaire completed in general practice. Journal of medical genetics. 2000;37:470–2.

31. Rose P, Humm E, Hey K, Jones L, Huson SM. Family history taking and genetic counselling in primary care. Family practice. 1999;16:78–83.

32. Feero WG, Bigley MB, Brinner KM. New standards and enhanced utility for family health history information in the electronic health record: an update from the American Health Information Community's Family Health History Multi-Stakeholder Workgroup. J Am Med Inform Assoc. 2008;15:723–8.

33. Shabo A, Hughes KS. Family History Information Exchange Services Using HL7 Clinical Genomics Standard Specifications. Int'ernational Journal on Semantic Web and Information Systems. 2005;1:44–67.

34. Kawamoto K, Lobach DF, Willard HF, Ginsburg GS. A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine. BMC Med Inform Decis Mak. 2009;9:17.

35. Christianson CA, Powell KP, Hahn SE, Blanton SH, Bogacik J, Henrich VC. The Use of a Family History Risk Assessment Tool within a Community Health Care System: Views of Primary Care Providers. Journal of genetic counseling [Internet]. 2012 Jan 21 [cited 2012 Jul 8]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/22271379

36. Facio FM, Feero WG, Linn A, Oden N, Manickam K, Biesecker LG. Validation of My Family Health Portrait for six common heritable conditions. Genetics in Medicine. 2010 May 14;12(6):370–5.

37. Owens K, Marvin M, Gelehrter T, Ruffin M, Uhlmann W. Clinical Use of the Surgeon General's "My Family Health Portrait" (MFHP) Tool: Opinions of Future Health Care Providers. Journal of Genetic Counseling. 2011;20(5):510–25.

38. Zimmerman NH, Patel C, Chen DP. ChMP: A Collaborative Medical History Portal. AMIA Annual Symposium Proceedings. 2008;2008:859.

39. Rubinstein WS, Acheson LS, O'Neill SM, Ruffin MT 4th, Wang C, Beaumont JL, et al. Clinical utility of family history for cancer screening and referral in primary care: a report from the Family Healthware Impact Trial. Genet. Med. 2011 Nov;13(11):956–65.

40. Cohn WF, Ropka ME, Pelletier SL, Barrett JR, Kinzie MB, Harrison MB, 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. 2010;13(7-8):477–91.

41. Kinzie MB. A User-centered Model for Web Site Design: Needs Assessment, User Interface Design, and Rapid Prototyping. Journal of the American Medical Informatics Association. 2002 Jul 1;9(4):320–30.

42. Orlando LA, Hauser ER, Christianson C, Powell KP, Buchanan AH, Chesnut B, et al. Protocol for implementation of family health history collection and decision support into primary care using a computerized family health history system. BMC Health Services Research. 2011 Oct 11;11(1):264.

43. Hoffer JA, Venkataraman R, Topi H. Modern Database Management. 10th ed. Prentice Hall; 2010.

44. Du P, Feng G, Flatow J, Song J, Holko M, Kibbe WA, et al. From disease ontology to disease-ontology lite: statistical methods to adapt a general-purpose ontology for the test of gene-ontology associations. Bioinformatics. 2009;25:i63–8.

45. Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. Journal of the American Medical Informatics Association : JAMIA. 1998;5:1–11.

46. w3schools.com. Available from: http://www.w3schools.com/browsers/browsers_stats.asp

47. www.usability.gov USD of H and HS. Available from: file:///Users/jonathan/Dropbox/BMI_research_project/references/family_history_project_references.Data/PDF/guidelines_book-3172003331/guidelines_book.pdf

48. Steinhaus KA, Bennett RL, Resta RG, Uhrich SB, Doyle DL, Markel DS, et al. Inconsistencies in pedigree symbols in human genetics publications: a need for standardization. Am J Med Genet. 1995;56:291–5.

49. Bennett RL, Steinhaus KA, Uhrich SB, O'Sullivan CK, Resta RG, Lochner-Doyle D, et al. Recommendations for standardized human pedigree

nomenclature. Pedigree Standardization Task Force of the National Society of Genetic Counselors. Am J Hum Genet. 1995;56:745–52.

50. Bennett RL, French KS, Resta RG, Doyle DL. Standardized human pedigree nomenclature: update and assessment of the recommendations of the National Society of Genetic Counselors. J Genet Couns. 2008;17:424–33.

51. http://raphaeljs.com/. Available from: http://raphaeljs.com/

52. JSON [Internet]. [cited 2012 Jul 6]. Available from: http://www.json.org/

53. Cascading Style Sheets [Internet]. [cited 2012 Jul 9]. Available from: http://www.w3.org/Style/CSS/Overview.en.html

54. Bender MA, Pemmasani G, Skiena S, Sumazin P. Finding least common ancestors in directed acyclic graphs. Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms [Internet]. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics; 2001 [cited 2012 Jun 30]. p. 845–54. Available from: http://dl.acm.org/citation.cfm?id=365411.365795

55. Fineout D. graph theory - Calculate Family Relationship from Genealogical Data - Stack Overflow [Internet]. [cited 2011 Jul 6]. Available from: http://stackoverflow.com/questions/1063666/calculate-family-relationship-from-genealogical-data

56. Brooke J. SUS - A quick and dirty usability scale. 1996.

57. Bangor A KP. An Empirical Evaluation of the System Usability Scale. International Journal of Human Computer Interaction. 2008;24:574–94.

58. Bangor A KP. Determining What Individual SUS Scores Mean: Adding an Adjetive Rating Scale. Journal of Usability Studies. 2009;4:114–23.

59. Murff HJ, Greevy RA, Syngal S. The comprehensiveness of family cancer history assessments in primary care. Community Genet. 2007;10:174–80.

60. Ampt A, Westbrook J, Creswick N, Mallock N. A comparison of self-reported and observational work sampling techniques for measuring time in nursing tasks. J Health Serv Res Policy. 2007 Jan;12(1):18–24.

61. Lewis JR, Sauro J. The factor structure of the System Usability Scale. Lecture Notes in Computer Science. 2009;5619:94–103.

62. national health statistics reports number 27, november 3, 2010 - nhsr027.pdf [Internet]. [cited 2012 Jul 8]. Available from: http://www.cdc.gov/nchs/data/nhsr/nhsr027.pdf

63. Emery J, Morris H, Goodchild R, Fanshawe T, Prevost AT, Bobrow M, et al. The GRAIDS Trial: a cluster randomised controlled trial of computer decision support for the management of familial cancer risk in primary care. Br J Cancer. 2007;97:486–93.

64. Yoon PW, Scheuner MT, Jorgensen C, Khoury MJ. Developing Family Healthware, a family history screening tool to prevent common chronic diseases. Prev Chronic Dis. 2009;6:A33.

65. Cupples LA. Family study designs in the age of genome-wide association studies: experience from the Framingham Heart Study. Curr. Opin. Lipidol. 2008 Apr;19(2):144–50.

66. Valdez R, Yoon PW, Qureshi N, Green RF, Khoury MJ. Family History in Public Health Practice: A Genomic Tool for Disease Prevention and Health Promotion. Annu Rev Public Health [Internet]. 2010.

# Vanderbilt University Institutional Review Board Informed Consent Document for Research

**The following information is provided to inform you about this research study and your participation in it. Please read this content carefully.**

Your participation in this research study is voluntary. You are also free to withdraw from this study at any time.

1. **Purpose of the study:** The purpose of the study is to evaluate the usability of this patient portal (www.myfamilyatvanderbilt.com), which is designed to aid a user in building structured Family Health Information using online tools. You are being asked to participate in this research study because you may help provide valuable feedback to improve this website's usability and design. Ascertaining structured Family Health Information directly from patients via an online portal such as this one facilitates unambiguous communication and could ease the burden of collecting it away from busy clinicians. The Family Health Information that will be ascertained from you will **NOT** be used for medical decision-making and will **NOT** be shared with your doctor.

2. **Procedures to be followed and approximate duration of the study:** We plan to implement and evaluate a web-based portal for the ascertainment of structured Family Health Information directly from users. Participants in this study will be asked to:

    a. Register for the www.myfamilyatvanderbilt.com website.
    b. Electronically sign this consent to participate in a research study.
    c. Complete tasks related to a standardized family tree with associated health conditions that all participants will be presented.
    d. Construct their own family tree and input health conditions regarding their own family members.
    e. If you choose, invite family members to also evaluate the website (shared Family Health Information will NOT be shared with your family members).
    f. Complete a survey evaluating their experience and provide feedback.

    The study period is one week, with the goal of recruiting 100 participants.

3. **Expected costs:** There is no cost to participate in this study. Participants will need to commit to the time required to construct and enter Family Health Information of their relatives.

4. **Description of the discomforts, inconveniences, and/or risks that can be reasonably expected as a result of participation in this study:** There are no significant discomforts or risks to participating in the study. Depending on the enthusiasm of the participant, the expected time commitment may vary. On average, it may take 30-45 minutes to complete this study.

5. **Unforeseeable risks:** As with all online health portals, there is a minimal risk for the loss of Protected Health Information.

6. **Compensation in case of study-related injury:** If you are injured because you are in this study, you can get reasonable, immediate, and necessary medical care for your injury at

Vanderbilt without charge to you. There are no plans for Vanderbilt to pay for the costs of care beyond your injury, or to give you money for such injury.

**7. Good effects that might result from this study:**

**a.) The benefits to science and humankind that <u>might</u> result from this study:** By providing valuable feedback that will be used to improve the usability and design of this patient portal, participants will be contributing to the advancement of science and all of humankind.

**b.) The benefits you might get from being in this study:** None.

**8. Alternative treatments available:** This is not a treatment study.

**9. Compensation for participation:** The first 100 Participants who complete the study will be eligible to enter a raffle to win an iPod touch. If you enter the raffle, you have a 1 in 100 chance to win the iPod Touch. The winning participant will need to provide identifying information and their Social Security number in order to claim their prize.

**10. Circumstances under which the Principal Investigator may withdraw you from study participation:** Participants who do not complete the evaluation survey will be withdrawn from the study.

**11. What happens if you choose to withdraw from study participation:** There are no consequences if a participant chooses to withdraw from the study, however they will not be eligible to win the raffle for the iPod Touch. Otherwise, participants may withdraw from this research study at any time without penalty.

**12. Contact Information.** If you should have any questions about this research study, please feel free to **contact the study's Principal Investigator (PI): Jonathan Holt, DO at 1-615-936-3720 or via email at jonathan.holt@vanderbilt.edu.**

For additional information about giving consent or your rights as a participant in this study, please feel free to contact the Vanderbilt University Institutional Review Board Office at (615) 322-2918 or toll free at (866) 224- 8273.

**13. Confidentiality:**

All efforts, within reason, will be made to keep your personal information in your research record confidential, however total confidentiality cannot be guaranteed. All information will be kept in a password-protected computer behind locked doors in a secure server room located on campus. Only aggregated data will be use for presentations or publications. No information will be linked to an individual or institution. All information will be kept confidentially. No study participants' names will be released to anyone. No evaluations or identifying information will be released. All Protected Health Information and associated Family Health Information will be erased at the conclusion of this study.

14. **Privacy:**

All efforts, within reason, will be made to keep your protected health information (PHI) private. PHI is your health information that is, or has been gathered or kept by Vanderbilt as a result of your healthcare. This includes data gathered for research studies that can be traced back to you. Using or sharing ("disclosure") such data must follow federal privacy rules. By signing the consent for this study, you are agreeing ("authorization") to the uses and likely sharing of your PHI. If you decide to be in this research study, you are also agreeing to let the study team use and share your PHI as described below.

Your information may be shared with Vanderbilt or the government, such as the Vanderbilt University Institutional Review Board, Federal Government Office for Human Research Protections, if you or someone else is in danger or if we are required to do so by law. Vanderbilt may give or sell your data without identifiers for other research projects not listed in this form. There are no plans to pay you for the use or transfer of this de-identified information.

**STATEMENT BY PERSON AGREEING TO PARTICIPATE IN THIS STUDY**

**By clicking 'agree', you state:**

- **You are 18 years of age or older**
- **You have read and understand this informed consent document**
- **You freely and voluntarily choose to participate in this research study**

**After you click to submit this form, documentation of your consent will be kept by the study investigator, you can download or print a copy of this form by clicking on the 'pdf' or 'print' icon above.**

# APPENDIX B. Standardized Family Health Information

| NAME (age) | RELATIONSHIP | HEALTH INFORMATION |
|---|---|---|
| **Immediate Family Members ( and their offspring )** | | |
| John Appleseed<br>(57 years old) | self | ○ hypertension (age: 50 to 59) |
| Junior<br>(20 years old) | son | |
| Jimmy<br>(19 years old) | son | |
| Jessica<br>(17 years old) | daughter | |
| Judy<br>(16 years old) | daughter | |
| Jack Appleseed<br>(81 years old) | father | ○ heart attack (age: 60 to 69)<br>○ coronary arteriosclerosis (age: 60 to 69)<br>○ hypertension (age: 50 to 59) |
| Margaret Appleseed<br>(86 years old) | mother | ○ skin cancer (age: 70 or older) |
| Jill<br>(48 years old) | sister | ○ hypertension (age: 40 to 49) |
| James<br>(53 years old) | brother | ○ hypertension (age: 40 to 49) |
| **Fathers's side of the family** | | |
| Jackie<br>***(died 88 years old)***<br>(edit) (+) add son (+) add daughter | grandmother | |
| Joe<br>***(died 72 years old)*** | grandfather | ○ hypertension (age: 60 to 69) |
| Jeffry<br>(81 years old) | uncle | |
| Jackson<br>***(died 81 years old)*** | uncle | |
| Jason<br>(78 years old) | uncle | |
| Jamie<br>(42 years old) | male cousin | |

| Mother's side of the family | | |
|---|---|---|
| Maureen<br>(107 years old)<br>(edit) (+) add son (+) add daughter | grandmother | |
| Max<br>***(died 81 years old)*** | grandfather | |
| Mark<br>(76 years old) | uncle | |
| Michael<br>(77 years old) | uncle | ○ hypertension (age: 50 to 59) |
| Mary<br>(79 years old) | aunt | ○ breast cancer (age: 60 to 69)<br>○ arthritis (age: 70 or older) |

# APPENDIX C. Technical Specifications

The software that underlies the *www.MyFamilyatVanderbilt.com* MyFaV webserver is written in PHP and MySQL utilizing a well-known collaborative framework called Drupal (version 6). Drupal (www.drupal.org) is an open-source, popular PHP based Content Management System (CMS) and web application framework. Drupal is the CMS of choice for notable domains such as *www.whitehouse.gov*. Drupal is composed of a core code base and by creating modules to make enhancements and modifications. The MyFaV website was constructed by building a Drupal module. MyFaV is running on an Apple Mac Mini server 2.66 GHz Intel Core 2 Duo with 8 GB of memory and 1 TB Hard drive running Mac OS X Lion 10.7 under Apache, MySQL, PHP and the Drupal Framework.

The programming approach utilized by the MyFaV site distilled data into essential elements for transmission as JSON objects sent to the clients' Internet browsers where it was used to populate a target  tag, which serves as a container. This approach allows for the rapid deployment into existing patient portals and Electronic Health Records.