

Characterizing the Use and Contents of Free-Text Family History Comments in the Electronic Health Record

Elizabeth S. Chen, PhD^{1,2}, Genevieve B. Melton, MD, MA^{7,8},
Timothy E. Burdick, MD, MS^{3,6}, Paul T. Rosenau, MD, MS^{4,6}, Indra Neil Sarkar, PhD, MLIS^{1,5}

¹Center for Clinical & Translational Science, Department of ²Medicine, ³Family Medicine, ⁴Pediatrics, ⁵Microbiology & Molecular Genetics, University of Vermont; ⁶Fletcher Allen Health Care, Burlington, VT
⁷Institute for Health Informatics, ⁸Department of Surgery, University of Minnesota, Minneapolis, MN

Abstract

The detailed collection of family history information is becoming increasingly important for patient care and biomedical research. Recent reports have highlighted the need for efforts to better understand collection and use of this information in resources such as the Electronic Health Record (EHR). This two-part study involved characterizing the use and contents of free-text comments within the family history section of an EHR. Based on a manual review of a subset of 11,456 cancer-related family history entries, 20 “reasons for use” were identified and the distribution across these reasons determined. A semi-automated analysis of the 3,358 unique comments associated with these entries was then performed to identify and quantify key categories of information. Implications of this study include guiding efforts for the improved use, collection, and subsequent analysis of family history information in the EHR.

INTRODUCTION

The understanding of a patient’s medical family history is an essential component for patient management, risk assessment, personalized medicine, and clinical genomic studies¹⁻⁴. Medical pedigrees, or “genograms,” can be used as a powerful clinical tool when linked to relevant phenotypic and genotypic data. The development of technologies and resources to collect, represent, integrate, and generate pedigrees based on information captured within disparate electronic sources could be valuable for enriching existing knowledge, enabling better patient care, and facilitating research studies.

The arrival of the era of personalized medicine has led to renewed interest and emphasis on the importance of medical family history. Family history has been described as a valuable personalized genomic tool for individualized disease prevention, diagnosis, and treatment¹. Many studies have demonstrated the use of family history to help predict the risks of health concerns such as heart disease and cancer^{1,2}. Despite the clear value of family history, obstacles to optimal use include lack of awareness of its relevance and potential impact, poor recall and limited knowledge about illnesses within the family by the patient, and limited time of clinicians². To address these barriers, numerous resources and computer-based tools have emerged to provide education and facilitate the collection, maintenance, and analysis of detailed family history (e.g., My Family Health Portrait⁵).

While the 2009 final statement of the National Institutes of Health (NIH) “State of the Science Conference: Family History and Improving Health” recognized the importance of family history information for personalized healthcare and risk assessment tools, it concluded that there is limited evidence regarding the effective collection and use of this information for common diseases⁶. This NIH statement calls for efforts to better understand collection and analysis of family history information (e.g., in Electronic Health Records [EHR]) where specific research priorities include studying the: (1) structure or characteristics of family history, (2) process of acquiring family history, and (3) outcomes of family history acquisition, interpretation, and application. Other efforts such as the Centers for Disease Control and Prevention Family History Public Health Initiative⁷ further emphasize the importance of family history and the need for more effective use (e.g., in pediatric primary care and public health⁸⁻¹⁰).

In recent years, there have been some efforts specifically focused on representing family history information¹¹ and extracting this information from “unstructured” clinical notes in the EHR (e.g., admission notes, discharge summaries, and outpatient clinic notes) using natural language processing^{12,13}. The present study is focused on studying another source of family history information in the EHR: the “structured” family history section and free-text comments within this section. A two-part approach was used to gain a better understanding of how the comments field is used and what is contained within this field. As part of this work, a semi-automated process was developed to facilitate extraction and analysis of information from the free-text comments. The findings from this study may have implications for improving use of the family history section and guiding user training locally as well as contribute to enhancing the design and implementation of this section in EHR systems more broadly.

METHODS & RESULTS

This study involved analyzing the structured family history section of an EHR system for a comprehensive healthcare system with a focus on the free-text comments (representing unstructured data) within this section. The overall approach involved two major parts for characterizing the use and contents of comments: (1) manual review of the structured family history entries to identify reasons for use of the comments field and (2) using an automated approach to identify and structure information captured within the narrative comments (Figure 1).

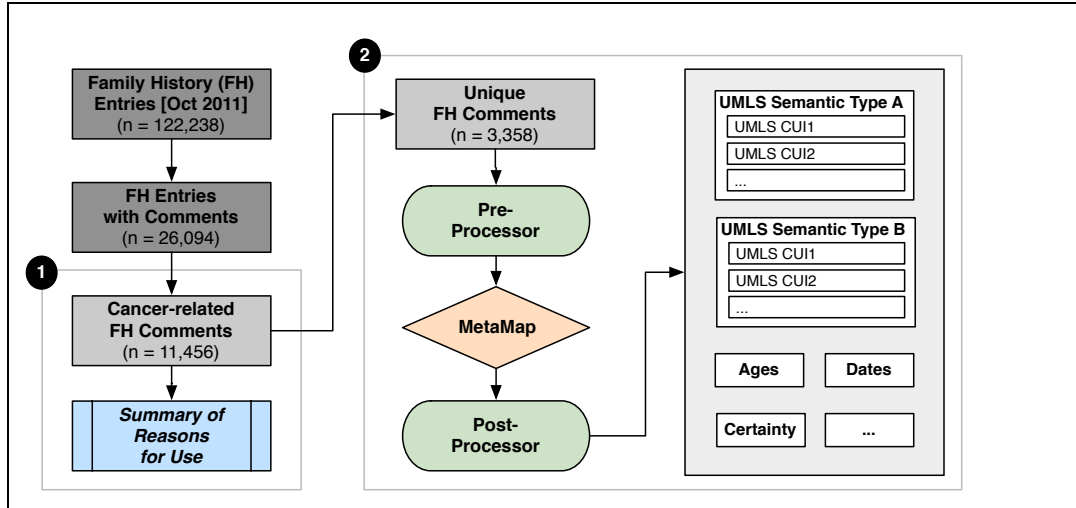


Figure 1: Overview of Methods

Dataset of Family History Entries

Fletcher Allen Health Care is the tertiary care academic medical center affiliated with the University of Vermont that provides care for over 60% of the state’s population¹⁴. The Epic EHR (Verona, WI)¹⁵ has been in use at Fletcher Allen since 2009 and includes a family history section for collecting information about the medical history and living status of family members in the inpatient and outpatient settings. The medical history portion allows for structured entry of problems (selected from a locally customized list of 210 values such as Cancer, Diabetes, or “*” for Other), familial relations (selected from a list of 21 values such as Mother, Brother, Other, or Neg Hx for absence of a relative with a specific problem), and age of onset (expressed in years as a numeric value such as 68.0). The status portion of the section includes structured fields for relation and status (e.g., Alive or Deceased). Both portions include free-text fields for specifying the family member’s name and providing comments. This study focuses on the comments associated with the medical history portion (Table 1).

Table 1: Example Family History Entries

Problem	Relation	Age of Onset	Comments
Cancer	-	-	bone cancer - both parents
Cancer	Mother	-	uterine cancer dx at age 62 died at age 68
Cancer	Other	55.00	Cousin AML
Cancer	Neg Hx	-	No known colon or prostate cancer.
Breast Cancer	Daughter	46.00	and colon cancer at 47
Colon Cancer	Paternal Uncle	-	age 50s, deceased
Ovarian Cancer	Other	42.00	pat 1st cousin

Family medical history entries entered during a one-month period (October 1, 2011 to October 31, 2011) were obtained, providing a total of 122,238 entries for 16,995 patients. Of these entries, 21.3% (26,094 entries for 9,057 patients) included comments. Since “Cancer” was found to be the most frequent problem that these comments qualified (37.2%; 9,707 entries), we decided to focus our initial analysis on cancer-related problems. All entries with a cancer-related problem were extracted for inclusion in the study dataset; this included those for “Cancer” as well as 14 specific types (Breast, Colon, Prostate, Ovarian, Thyroid, Liver, Stomach, Kidney, Esophageal, Cervical, Pancreatic, Endocrine, Endometrial, and Intestinal). The resulting dataset included 11,456 cancer-related entries

with comments (for 5,466 patients). This dataset provided a total of 3,358 unique comments where about one-third of these represent duplicates (e.g., “lung CA”, “deceased”, “unknown”, “Uncle”, and “70s”) with the remaining two-thirds occurring only once.

Part 1: Characterizing Use of the Family History Comments Field

A manual review of entries in the dataset was performed to characterize the use of the family medical history comments field in order to gain a better understanding of how this field is used. A random sample of 50 entries was analyzed to create an initial coding scheme representing a list of “reasons for use” (where an entry may have more than one reason). For example, the reason “multiple problems” indicates that multiple problems are mentioned in the comments (since only one problem can be selected in an entry), the reason “missing relation” indicates that the member is not in the list of 21 values available, and the reason “onset date” represents a case where a date is specified (rather than a specific age). Two reviewers used this coding scheme to analyze another random sample of 50 entries for determining inter-rater reliability and enhancing the coding scheme if needed. The main analysis then involved coding a random sample of 500 entries (250 each) by the two reviewers. Collectively, the number of comments reviewed covered about 5% of the entries in the dataset and over 15% of the unique comments.

Table 2: Reasons for Use – Description, Examples, and Distribution Across Entries

Reason	Description	Examples	Frequency
A. Problem Reasons			
Problem in list	Includes a problem that is in the list of values (should use problem field)	“breast cancer at 70” “colon, age 75”	187 (37.4%)
Missing problem	Includes a more specific or general problem, or problem that is not in the list of values	“brain cancer” “throat, deceased 1963 age 60’s”	200 (40.0%)
Multiple problems	Lists multiple problems	“skin, ovarian” “uterine, breast, kidney cancers”	71 (14.2%)
B. Relation Reasons			
Relation in list	Includes relation that is in the list of values (should use relation field)	“Prostate, father” “skin cancer, grandmother”	34 (6.8%)
Missing relation	Includes a more specific or general relation, or relation that is not in the list of values	“great grandmother ovarian” “lymphoma, cousin”	41 (8.2%)
Multiple relations	Lists multiple relations	“SISTERS, AUNT”	21 (4.2%)
C. Onset Age Reasons			
Onset age – exact	Includes a specific onset age (should use age of onset field)	“dx’d age 65” “onset age 60”	87 (17.4%)
Onset age – fuzzy	Includes an estimated onset age	“diagnosed in his 70s” “elderly onset”	45 (9.0%)
Onset date	Includes onset or diagnosis date	“diagnosed 10/2011”	1 (0.2%)
Multiple onset ages or dates	Lists multiple onset dates	“bilateral, age 37 then 45 in opposite breast”	2 (0.4%)
D. Other Reasons			
Living status	Indicates living status of family member (should use status field)	“diagnosed at 62 Still living” “died of stomach cancer”	80 (16.0%)
Deceased age	Includes exact or estimated age of death	“aunt (died @ 90 years)”	60 (12.0%)
Deceased date	Includes exact or estimated date of death	“throat, deceased 1963 age 60’s”	5 (1.0%)
Ambiguous age	Includes an age but unclear if it is age of onset or death	“9/2004, sister age 46” “8/6/99, brother 60”	2 (0.4%)
Ambiguous date	Includes a date but unclear if it is for onset or death	“10/31/97”, “4/2002”	16 (3.2%)
Negative history	Specifies negative history for specific problem	“no breast or colon cancer”	0 (0.0%)*
Certainty	Indicates uncertainty about a particular problem, relation, or age	“? bone or thyroid cancer” “uncertain what kind”	40 (8.0%)
Other conditions/ findings	Includes conditions or findings other than type of cancer	“laryngeal, heavy smoker” “cervical cancer; obesity”	24 (4.8%)
Procedures/ therapies/tests	Includes reference to procedures, therapeutics, and/or tests	“breast cancer – lumpectomy” “pat 1st cousin, BRCA neg”	20 (4.0%)
Other	Any other information included	“prostate metastasized”, “x 3”, “lymphoma remission”	31 (6.2%)

* No occurrences in the sample of 500 entries

A total of 20 reasons for use (including “Other”) was identified based on the two samples of 50 entries (16 reasons were initially identified with the first sample and an additional 4 reasons were added based on the second sample). Inter-rater reliability between the two reviewers in the assignment of reasons for the 50 entries yielded κ (0.948) and proportion agreement (99.2%). Table 2 lists each reason (along with a brief description and examples) and the distribution of each for the main sample of 500 entries.

Part 2: Characterizing the Contents of Family History Comments

To facilitate the analysis of contents in the comments and demonstrate the feasibility of semi-automating this analysis, MetaMap¹² from the National Library of Medicine was used to extract information from the comments. As part of this process, a pre-processor and post-processor were developed for generating the input for MetaMap and formatting the results for subsequent use, respectively.

A pre-processor (implemented as a set of Ruby scripts) was developed to perform various pre-processing tasks such as removing extra whitespaces, lowercasing the text, reformatting dates, and fixing misspellings. Several date formats were found across the set of comments (e.g., “1/2003”, “5/99”, and “3/23/01”) and were standardized to “YYYY-MM-DD” (e.g., “2003-01”, “1999-05”, and “2001-03-23”). A list of misspellings was created that included a mapping of misspelled words to their correct spellings; this list was subsequently used to fix misspellings in the comments (e.g., “decerased” → “deceased”, “larygeal” → “laryngeal”, “lukemia” → “leukemia”, “melinoma” → “melanoma”). Additional word transformations were performed in order to improve MetaMap performance (e.g., “mom” → “mother”, “passed away” → “died”, and “hodgkin’s” → “hodgkin lymphoma”).

The 2011 version of MetaMap¹³ was applied to the pre-processed comments. Based on iterative testing of the various MetaMap options, the following configuration was used in this study: *-z* for processing the comments as terms rather than full text, *-R NCI* for restricting the use of sources to the NCI Thesaurus, *-N* for printing the results as fielded output, and *--UDA <file>* for specifying a list of user-defined acronyms and abbreviations (UDAs) and their expansions. This UDA file was created based on acronyms and abbreviations found throughout the comments (e.g., “ca” → “cancer”, “mgm” → maternal grandmother, “nhl” → “non-hodgkin lymphoma”).

Table 3: Example Comments and Extracted Information

Pre-Processed Comment (Original)	MetaMap Results (CUI Name [Semantic Type])	Other Results
<i>gf esophageal</i> , gm breast cancer (<i>GF esophogeal</i> , GM breast cancer)	C1522619 Esophageal [spco] C0678222 Breast Carcinoma [neop] C0006142 Malignant Neoplasm of Breast [neop] C0337475 Grandfather [famg] C0337474 Grandmother [famg]	N/A
2009-12, bladder cancer, sister age 39 (<u>12/2009</u> , bladder cancer, sister age 39)	C0337514 Sister [famg] C0699885 Carcinoma of bladder [neop] C0005684 Malignant neoplasm of urinary bladder [neop] C0001779 Age [orga]	Date = 2009-12 Age = 39
died of colon ca ? 50's or 60's	C0699790 Carcinoma of colon [neop] C0007102 Malignant tumor of colon [neop] C0011065 Death [orgf]	Certainty = ? Age = 50's^60's
lung-smoker, hodgkin's <i>disease</i> (lung-smoker, hodgkin's <i>disease</i>)	C0024109 Lung [bpoc] C0019829 Hodgkin disease [neop] C0337664 Smoker [fndg]	N/A
maternal cousin; leukemia	C0023418 leukemia [neop] C0337580 Cousin [famg] C2347083 Maternal Relative [famg]	N/A

italics = misspelling; underline = date transformation; [bpoc] = Body Part, Organ, or Organ Component; [famg] = Family Group; [fndg] = Finding; [neop] = Neoplastic Process; [orga] = Organism Attribute; [orgf] = Organism Function; [spco] = Spatial Concept

A post-processor (implemented as a Ruby script) was created to extract the UMLS Concept Unique Identifiers (CUIs), names, and semantic types for each comment from the MetaMap output and transform them into a tabular format to facilitate subsequent analysis and use¹⁴. For example, for each comment, concepts with a semantic type of “Neoplastic Process” were combined into a single field and concepts with a semantic type of “Family Group” were combined into a separate field. Other post-processing tasks included those for extracting additional information that

was not detected by MetaMap such as ages (e.g., “52”, “70s”, “@~75”, and “@ 93 y/o”), dates, and certainty (e.g., use of “?” in the comment). Since MetaMap identified concepts for “diagnoses” (C0011900), “onset” (C0332162), “death” (C0011065), and “age” (C0001779), these were used to indicate that a particular comment included information about onset, living status, and age that could be linked to the specific age and date information. Table 3 includes several examples depicting the pre-processed comment, original comment (if different than the pre-processed version), and post-processed results of MetaMap output and other information.

For the 3,358 unique comments, MetaMap identified a total of 8,384 concepts (830 unique concepts) representing 77 semantic types in 3,217 of them (95.8%). Table 4 lists the top 10 semantic types (Table 4A) and top 10 concepts for the 3 most frequent types: “Neoplastic Process” (Table 4B), “Body Part, Organ, or Organ Component” (Table 4C), and “Family Group” (Table 4D). The concept for “death” occurred in 500 (14.8%) comments, concepts for “diagnoses” and “onset” occurred in 108 (3.2%) comments, and concept for “age” occurred in 483 (14.4%) comments. In addition to the MetaMap findings, 159 (4.7%) of the comments were found to include date information and 878 (26.1%) included age information.

Table 4: Top 10 Semantic Types and Concepts for Specific Types

(A) Semantic Types		(B) Neoplastic Process	
<i>Semantic Type</i>	<i>Frequency</i> [*]	<i>Concept</i>	<i>Frequency</i> [‡]
Neoplastic Process	2851 (34.0%)	C0006142 Malignant Neoplasm of Breast C0678222 Breast Carcinoma	163 (5.1%) 160 (5.0%)
Body Part, Organ, or Organ Component	1244 (14.8%)	C0242379 Malignant Neoplasm of Lung C0684249 Carcinoma of Lung	154 (4.8%) 152 (4.7%)
Family Group	660 (7.9%)	C1306459 Primary Malignant Neoplasm C0006826 Malignant Neoplasm	121 (3.8%) 121 (3.8%)
Organism Function (e.g., C0011065 Death)	512 (6.1%)	C0007102 Malignant Tumor of Colon C0699790 Colon Carcinoma	120 (3.7%) 119 (3.7%)
Organism Attribute (e.g., C0001811 Age)	486 (5.8%)	C0025202 melanoma	79 (2.5)
Qualitative Concept (e.g., C0439673 Unknown)	352 (4.2%)	C0023418 leukemia	61 (1.9%)
Quantitative Concept (e.g., C0439064 Multiple)	327 (3.9%)	C0007114 Malignant Neoplasm of Skin	54 (1.7%)
Finding (e.g., C0011900 Diagnosis)	300 (3.6%)	C0153567 Uterine Cancer	53 (1.6%)
Spatial Concept (e.g., C0011900 Cervical)	239 (2.9%)	C0376358 Malignant Neoplasm of Prostate C0600139 Prostate Carcinoma	53 (1.6%) 52 (1.6%)
Temporal Concept (e.g., C0205087 Late)	193 (2.3%)	C0024299 Lymphoma	51 (1.6%)
(C) Body Part, Organ, or Organ Component		(D) Family Group	
<i>Concept</i>	<i>Frequency</i> [‡]	<i>Concept</i>	<i>Frequency</i> [‡]
C0006141 Breast	189 (5.9%)	C0337576 Aunt	96 (3.0%)
C0024109 Lung	176 (5.5%)	C2347083 Maternal Relative	92 (2.9%)
C0009368 Colon	147 (4.6%)	C0337580 Cousin	82 (2.5%)
C0033572 Prostate	91 (2.8%)	C0337474 Grandmother	58 (1.8%)
C0042149 Uterus	67 (2.1%)	C0026591 Mother	46 (1.4%)
C0030274 Pancreas	60 (1.9%)	C0337514 Sister	45 (1.4%)
C0205065 Ovarian	57 (1.8%)	C2347452 Paternal Relative	42 (1.3%)
C0031354 Pharynx	48 (1.5%)	C0337577 Uncle	37 (1.2%)
C0005682 Urinary Bladder	47 (1.5%)	C0015671 Father	27 (0.8%)
C0006104 Brain	44 (1.4%)	C0337475 Grandfather	24 (0.7%)

^{*} Percentage out of the total number of concepts (n=8,384); [‡] Percentage out of the total number of comments with concepts (n=3,217)

DISCUSSION

In this paper, we have described an approach and early results for characterizing the use and contents of free-text family history comments in the EHR. A manual review was conducted to identify and summarize reasons for use of the comments field. In addition, a semi-automated process was developed to identify and quantify key categories of information within a set of comments.

As reflected in Table 2, “Problem in list”, “Onset age – exact”, and “Living status” are among the top 5 reasons for comment use, which conveys that the comments field is being used to collect information that should be entered into available structured fields (i.e., “Problem” and “Age of Onset” in the family medical history portion and “Status” in the family status portion). These reasons along with the reasons “Multiple problems” and “Multiple relations” may

be addressed by training or user interface modifications to enable more flexible entry of information (currently, there are two modes for entering family medical history and status, the efficiencies of which vary with respect to the reasons for documentation inferred from this study). Other frequent reasons such as “Missing problem” and “Missing relation” suggest that the locally customized list of values for problems and relations could be enhanced to include additional types of cancer and family members (guided by results from both parts of this study). For example, as shown in Table 4, “leukemia”, “uterine cancer”, and “brain” are among the top 10 concepts but are not in the list of values provided for problems; similarly, “cousin” is not currently in the list of values for relation. The aforementioned findings are similar to those described by previous efforts focused on the study of structured “data-entry exit strategies” for understanding reasons for using free-text rather than standardized codes for problems, diagnoses, and medications in the EHR^{22,29}. The frequency of concepts for “maternal relative” and “paternal relative” also suggests that there may be a need for more flexible specification of side of family (i.e., maternal and paternal). While the list of relations includes some “pre-coordinated” values such as “Maternal Grandmother” and “Paternal Uncle”, there may be value in being able to “post-coordinate” side of family (e.g., separately specifying “Maternal” and “Grandmother”) rather than attempting to anticipate all possible combinations in the list.

The initial pipeline implemented in this study consisted of a pre-processor, MetaMap, and a post-processor. Challenges encountered included misspellings, acronyms, and abbreviations that were found throughout the comments as well as variations in age and date formats. A manual process was used to address each of these challenges to some extent in this study where future work will involve developing more robust and automated methods for handling each of these issues. Next steps also include performing a formal evaluation to characterize false positives and false negatives, and determining what adjustments can be made to the MetaMap configuration used in this study to improve performance. For this study, use of all source vocabularies, SNOMED CT only, and NCI Thesaurus only were tested and found to produce similar results with the former two configurations providing additional concepts, particularly for body parts, organs, or organ components (e.g., Entire Lung [C1278908] in addition to Lung [C0024109] for “lung”). Given the noise introduced by these two configurations, we chose to use NCI Thesaurus only in order to demonstrate the feasibility of using MetaMap to study the contents of free-text family history contents; however, future work would involve incorporating additional sources or potentially all sources to enhance the results, and exploring strategies for filtering concepts as appropriate. For example, SNOMED CT²⁰ and HL7 Version 3.0²¹ could be included as other source vocabularies to detect additional concepts such as “great grandmother”, which was not found when restricting to use of the NCI Thesaurus.

In order to limit the scope, this study focused on cancer-related comments found within the medical history portion of the family history section for a specific time period. Next steps include applying the approach to all comments for the medical history portion (that are associated with a range of conditions as well as a non-specific “Other” value) as well as the status portion. In addition, the techniques could be extended to clinical notes and build upon previous efforts to extract family history information from notes^{12,13}. A comparison of the various structured and unstructured sources of family history information in the EHR (e.g., free-text comments, clinical notes, and problem list) could then be performed to quantify the distribution of information across these sources and determine if the information is complementary, redundant, or potentially conflicting. Other comparisons include studying the differences in use and contents of comments based on provider characteristics (e.g., role, specialty, or practice) and patient characteristics (e.g., age, gender, or problem). These characteristics or contexts may have significant influence in how and what family history information is documented and contribute to guiding EHR customization. For example, top concepts for Family Group (Table 4D) indicate that aside from the gender-neutral concepts, the occurrence of female relatives is more frequent than male relatives, which may be due to the occurrence of breast cancer related entries in the dataset and supports the potential value of having context-specific functionality (e.g., customized or ranked lists for familial relations based on the selected problem). A broader goal will be to test the generalizability of the approach by applying the methods to other sources of free-text comments in the EHR (e.g., for problems²²) as well as to EHR systems at other institutions.

There have been several initiatives focused on the representation and standardization of information related to family history (e.g., American Health Information Community’s Family Health History Workgroup²³ and HL7 Clinical Genomics Family History Model^{24,25}). In previous work¹¹, we assessed the adequacy of the HL7 Clinical Genomics Family History Model and HL7 Clinical Statement Model^{26,27} for representing family history information in a set of clinical notes. While these existing models were found to be able to represent most information, the results indicated that several enhancements are needed including ability to represent paternal/maternal side of family and flexibility in handling age information such as different age events (e.g., current age, age of onset/diagnosis, and

age of death), non-specific ages (e.g., elderly), and age ranges (e.g., 50-60). The findings from the present study further support the need for such enhancements and will be used to extend the Merged Family History Model that was created in this previous study. In addition to contributing to these modeling efforts, the results of this work may also be used to supplement relevant vocabularies or code systems (e.g., the HL7 V3 Vocabulary for RoleCode²⁸ that defines a list of relatives) with additional values found in the comments (e.g., great aunt).

Collectively, the results from both parts of this study provide valuable insights into clinician thought-processes and specifically how the comments field for family history has been used. These findings could help inform recommendations for enhancing system functionality and user training for improved use and collection of family history information. In addition, the ability to automate the extraction, structuring, and encoding of information captured within family history comments may further improve use of this information by making it more accessible for patient care, decision support, and research. Complementing the approach described in this study with qualitative methods (e.g., interviews and focus groups with clinicians and researchers) could provide further insights to the needs and uses of family history for guiding enhancements and customizations in the EHR.

CONCLUSION

There has been increasing emphasis on the importance of family history and the need to improve its collection and use. The goal of this study was to characterize the use and contents of free-text family history comments in the electronic health record. Through use of manual and automated approaches, insights were gained about how comments have been used and what types of information are contained within them. The preliminary findings have the potential to guide system enhancements and training for improved collection and use of family history information.

References

1. Guttmacher AE, Collins FS, Carmona RH. The family history--more important than ever. *N Engl J Med*. 2004 Nov 25;351(22):2333-6.
2. Hinton RB, Jr. The family history: reemergence of an established tool. *Crit Care Nurs Clin North Am*. 2008 Jun;20(2):149-58, v.
3. 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 Mar;19(3):273-80.
4. Wattendorf DJ, Hadley DW. Family history: the three-generation pedigree. *Am Fam Physician*. 2005 Aug 1;72(3):441-8.
5. <https://familyhistory.hhs.gov>.
6. 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 Dec 15;151(12):872-7.
7. <http://www.cdc.gov/genomics/famhistory/famhist.htm>.
8. Olney RS, Yoon PW. Role of family medical history information in pediatric primary care and public health: introduction. *Pediatrics*. 2007 Sep;120 Suppl 2:S57-9.
9. Green RF. Summary of workgroup meeting on use of family history information in pediatric primary care and public health. *Pediatrics*. 2007 Sep;120 Suppl 2:S87-100.
10. Trotter TL, Martin HM. Family history in pediatric primary care. *Pediatrics*. 2007 Sep;120 Suppl 2:S60-5.
11. Melton GB, Raman N, Chen ES, Sarkar IN, Pakhomov S, Madoff RD. Evaluation of family history information within clinical documents and adequacy of HL7 clinical statement and clinical genomics family history models for its representation: a case report. *Journal of the American Medical Informatics Association : JAMIA*. 2010 May-Jun;17(3):337-40.
12. Friedlin J, McDonald CJ. Using a natural language processing system to extract and code family history data from admission reports. *AMIA Annu Symp Proc*. 2006:925.
13. Goryachev S, Kim H, Zeng-Treitler Q. Identification and extraction of family history information from clinical reports. *AMIA Annu Symp Proc*. 2008:247-51.
14. McDowell SW, Wahl R, Michelson J. Herding cats: the challenges of EMR vendor selection. *J Healthc Inf Manag*. 2003 Summer;17(3):63-71.
15. <http://www.epic.com/>.

16. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. *Journal of the American Medical Informatics Association : JAMIA*. 2010 May-Jun;17(3):229-36.
17. <http://metamap.nlm.nih.gov/>.
18. <http://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NCI/>.
19. Chen ES, Hripcsak G, Friedman C. Disseminating natural language processed clinical narratives. *AMIA Annual Symposium proceedings / AMIA Symposium*. 2006:126-30.
20. <http://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/SNOMEDCT/>.
21. <http://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/HL7V30/>.
22. Wang SJ, Bates DW, Chueh HC, Karson AS, Maviglia SM, Greim JA, et al. Automated coded ambulatory problem lists: evaluation of a vocabulary and a data entry tool. *Int J Med Inform*. 2003 Dec;72(1-3):17-28.
23. 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. *Journal of the American Medical Informatics Association : JAMIA*. 2008 Nov-Dec;15(6):723-8.
24. <http://www.hl7.org/Special/committees/clingenomics/>.
25. <http://www.hl7.org/v3ballot/html/domains/uvcg/uvcg.html>.
26. <http://www.hl7.org/v3ballot/html/domains/uvcs/uvcs.html>.
27. <http://www.hl7.org/special/Committees/clinicalstatements/>.
28. <http://www.hl7.org/v3ballot/html/infrastructure/vocabulary/RoleCode.html>.
29. Zheng K, Hanauer DA, Padman R, Johnson MP, Hussain AA, Ye W, Zhou X, Diamond HS. Handling anticipated exceptions in clinical care: investigating clinician use of 'exit strategies' in an electronic health records system. *J Am Med Inform Assoc*. 2011 Nov-Dec;18(6):883-9.