Examining the Use, Contents, and Quality of Free-Text Tobacco Use Documentation in the Electronic Health Record

Elizabeth S. Chen, PhD[1,2], Elizabeth W. Carter, MS[1], Indra Neil Sarkar, PhD, MLIS[1,3], Tamara J. Winden, MBA[4,6], Genevieve B. Melton, MD, MA[4,5]
[1]Center for Clinical & Translational Science, [2]Medicine,
[3]Microbiology & Molecular Genetics, University of Vermont, Burlington, VT;
[4]Institute for Health Informatics, [5]Surgery, University of Minnesota, Minneapolis, MN;
[6]Division of Applied Research, Allina Health, Minneapolis, MN

Abstract

Recent initiatives have emphasized the potential role of Electronic Health Record (EHR) systems for improving tobacco use assessment and cessation. In support of these efforts, the goal of the present study was to examine tobacco use documentation in the EHR with an emphasis on free-text. Three coding schemes were developed and applied to analyze 525 tobacco use entries, including structured fields and a free-text comment field, from the social history module of an EHR system to characterize: (1) potential reasons for using free-text, (2) contents within the free-text, and (3) data quality issues. Free-text was most commonly used due to limitations for describing tobacco use amount (23.2%), frequency (26.9%), and start or quit dates (28.2%) as well as secondhand smoke exposure (17.9%) using a variety of words and phrases. The collective results provide insights for informing system enhancements, user training, natural language processing, and standards for tobacco use documentation.

Introduction

Tobacco use continues to be the leading preventable cause of morbidity and mortality in the United States[1, 2]. Worldwide, direct tobacco use is responsible for more than 5 million deaths each year while exposure to secondhand smoke is responsible for over 600,000[3, 4]. Public health initiatives such as Healthy People 2020 Tobacco Use[5, 6], the Centers for Disease Control and Prevention's National Tobacco Control Program[7], the U.S. Preventive Services Task Force[8, 9], and the World Health Organization's Tobacco Free Initiative[10] involve efforts aimed at ending the tobacco epidemic through targeted prevention and treatment strategies for children, adolescents, and adults.

In the last five years, there has been increasing emphasis on the potential role of Electronic Health Record (EHR) systems for identification and treatment of tobacco use. Currently among the Centers for Medicare & Medicaid Services Meaningful Use Objectives[12] is a core measure focused on the recording of smoking status as structured data in the EHR using a specified set of SNOMED CT codes (e.g., for "Current every day smoker," "Former smoker," and "Never smoker")[13] and clinical quality measures for tobacco use screening and cessation intervention[14]. A recent Institute of Medicine report further highlighted the importance of capturing behavioral determinants of health in the EHR and specified nicotine use and exposure among the domains to consider for Stage 3 Meaningful Use[11]. A Cochrane Review of 11 studies that involved using the EHR to improve documentation or treatment of tobacco use found that there were modest improvements and concluded that additional research is needed to understand the role of EHRs in this context[15]. Among these studies was a demonstration project where workflow modifications included incorporating evidence-based prompts in the Epic EHR at Dean Health Systems for guiding the identification of current tobacco users, determining their willingness to quit, and offering a set of tobacco cessation interventions[16]. Recent studies have also described workflow changes such as incorporating medical assistants in the documentation and referral process[17] as well as decision support functionality such as alerts and pre-defined order sets[18].

Within the EHR, tobacco use, secondhand smoke exposure, and related interventions may be documented in various parts as structured data or free-text (e.g., problem list[19, 20], social history[21], medications[19, 22], clinical notes, or patient instructions). A number of efforts have focused on developing natural language processing (NLP) techniques to extract smoking status[23, 24] and tobacco cessation interventions (e.g., searching for the "5 A's" for tobacco treatment and prevention)[25, 26] from clinical notes such as discharge summaries. In a recent study, supplementing structured fields with information from free-text fields was found to substantially improve smoking status data in the EHR[27].

While previous efforts have focused on enhancing the EHR for smoking status and extracting this information from free-text clinical notes, there has been limited discussion on improving the collection of details about tobacco use

(e.g., amount and frequency) and exploring free-text tobacco use documentation throughout the EHR. To this end, the objective of the present study was to examine the use, contents, and quality of free-text comments for tobacco use in the primarily structured social history module of an EHR system. Potential implications of the findings include informing system enhancements, user training, NLP, and standards for tobacco use documentation that may ultimately contribute to improving tobacco use assessment and cessation interventions using the EHR.

Methods

Setting and Study Design

This study involved the retrospective analysis of information collected in the social history module of the Epic EHR (Epic Systems Corporation, Verona, WI)[28] at Fletcher Allen Health Care, the academic health center affiliated with the University of Vermont. This module can be used for primarily structured documentation of tobacco use, alcohol use, illicit drug use, and a range of other social history-related information, which may subsequently be used to pre-populate the social history section in clinical notes. At the time of this study, each tobacco use entry included a set of structured fields associated with smoking, another set of structured fields associated with smokeless tobacco use, and a free-text field for comments (Table 1). Of the 158,608 patients with information documented using the social history module in 2013, this free-text field was used for 18,221 (11.5%) patients where the average length of the comments was 24±19 characters (minimum = 1 and maximum = 255).

Table 1. Example Tobacco Use Entries.

Field	Example 1	Example 2	Example 3
Smoking status	Current Everyday Smoker	Former Smoker	Passive Smoker
Start date	-	-	-
Quit date	-	2/12/02 0:00	-
Types (*Cigarettes*, *Pipe*, or *Cigars*)	Cigarettes	Cigarettes	-
Packs/day	0.5	1	-
Years	15	11	-
Pack years*	7.5	11	-
Smokeless tobacco	Never Used	Current User	Unknown
Quit date	-	-	-
Types (*Snuff* or *Chew*)	-	Chew	-
Comment	Started smoking again in 2010 after quitting a few years	2 cans weekly	Parents smoke outside

** Calculated based on Packs/day and Years*

Three coding schemes were used to manually analyze tobacco use entries in order to characterize: (1) reasons for using the free-text comment field, (2) contents within this free-text field, and (3) data quality issues. The general approach for developing and applying each of these coding schemes (further described below) involved three phases: (1) generating initial coding schemes based on analysis of 100 tobacco use entries from September 2013 and enhancing the schemes using an iterative, consensus-based process involving individuals with expertise in clinical care and biomedical informatics (ESC, EWC, INS, TJW, and GMM); (2) calculating inter-rater reliability using the kappa statistic to ensure consistency in coding between two reviewers (ESC and EWC) using the final versions of each coding scheme for 50 entries from October 2013; and, (3) performing the main analysis on a random sample of 525 tobacco use entries from November 2013 by one reviewer (ESC) where this sample size was based on a total of 4,056 most recent entries for patients during this time, confidence level of 95%, and estimated precision of 4%.

Analysis of Potential Reasons for Using Free-Text Tobacco Use Comments

The first coding scheme for "reasons for use" was developed for identifying potential explanations for why the free-text comment field was used for each patient. In the initial version of the coding scheme, 16 different reasons were identified, which was expanded to 18 reasons (including one for *Other*) in the final version that were grouped into four major categories: (1) Misplaced or redundant information in free-text, (2) Missing values for available structured fields, (3) Limited capabilities of available structured fields, and (4) Other (Table 2). Comments could be associated with one or more potential explanations. For example, the comment "Occasional cigar" would be coded with two reasons: (1) *Misplaced – use Type field* and (2) *Limited ability to describe frequency*. One reviewer then analyzed the set of 525 entries to determine the most frequent reasons for using the free-text tobacco use comment

field. Inter-rater reliability between two reviewers for the set of 50 entries (almost 10%) was calculated using Cohen's kappa, achieving κ of 0.91 for coding reasons.

Table 2. Coding Scheme for Reasons.

#	Potential Reason	Brief Description	Example Comments
Misplaced or Redundant Information in Free-Text			
1	Misplaced – use Smoking status field	Smoking status field includes 10 values, including "Heavy Tobacco Smoker" and "Passive Smoker"	• hx of heavy tobacco use • exposed to second hand smoke
2	Misplaced – use Packs/day field	Could be entered using Packs/day field	• 50 years 2ppd = 100+ pack-years • 30yr x 0.5 ppd
3	Misplaced – use Years field	Could be entered using Years field	• 50 years 2ppd = 100+ pack-years • 30yr x 0.5 ppd
4	Misplaced – use Type field	Could be entered using options for Types (e.g., cigarettes, pipe, or cigars)	• Occasional cigar • Pipe
5	Misplaced – use Start or Quit date field	Could be entered using Start or Quit date fields	• quit 4/3/2011 • Quit just recently. (8/16/13)
6	Redundant Text	Same/synonymous information in comment also entered into structured fields	• Former smoker • nonsmoker
Missing Values in Available Structured Fields			
7	Missing value for Type field	Comment includes type that is not among available values for Type field	• Switched to an inhaler • electronic cigarette
Limited Capabilities of Available Structured Fields			
8	Limited ability to describe amount	Comment includes amount that cannot be described using available fields (Packs/day and Pack years fields)	• 1-2 cigars a day • 2 cans weekly • a few a day
9	Limited ability to describe frequency	Comment includes frequency that cannot be described using available fields (Packs/day)	• Smoked cigars sporadically • occasional pipe • a few a day
10	Limited ability to describe start or quit date	Comment includes a date that cannot be described using Start date or Quit date fields that require mm/dd/yy	• Quit April 2010 • Quit one year ago • Quit 1971
11	Limited ability to describe start or quit age	Comment includes an age related to starting or quitting	• smoked until age 16 • from age 18-26
12	Limited ability to describe duration or timepoint	Comment includes tobacco use or quit duration or timepoint	• Many years • Quit for 10 yr
13	Limited ability to describe situation	Comment includes information about situation or context of tobacco use	• Social • occasional in college
14	Limited ability to describe cessation attempts	Comment includes information about quit attempts, interventions, etc.	• Would like to quit • Working on quitting
15	Limited ability to describe passive smoke exposure	Comment includes information about passive smoke exposure that cannot be described in available fields	• Parents smoke in the home • No second hand smoke
16	Limited ability to specify multiple values	Comment includes additional status, age/date, etc.	• Quit 04/12/2008, restarted in 07/2008, quit 2/2010
Other			
17	Multiple statements	Comment includes multiple pieces of information	• 1-2 cigarettes/day. Quit on /1/13.
18	Other	Any other reason for use	• Smokes marijuana daily

Analysis of Contents within Free-Text Tobacco Use Comments

The second coding scheme for analyzing the "contents" of free-text was focused on categorizing words and phrases within the free-text comments into separate elements. The initial coding scheme included a combination of 10 elements identified in previous work involving the analysis of tobacco use information in clinical notes from multiple EHR systems as well as public health surveys[29, 30]. These elements included: (1) *Status* – current or past tobacco use, (2) *Temporal* – age or date when patient started or quit using tobacco (may be exact or estimated), (3) *Method* – how tobacco is/was used, (4) *Type* – what type of tobacco is/was used, (5) *Subtype* – additional details about type such as brand or filtered/unfiltered, (6) *Amount* – amount of tobacco the patient uses/used, (7) *Frequency* – how often tobacco is/was used, (8) *Certainty* – conviction of source (e.g., patient) regarding tobacco use, (9)

Experiencer – who uses/used tobacco, and (10) *Location* – where tobacco is/was used. An additional four elements were incorporated in the final version of the coding scheme for: (1) *Negation* – absence of tobacco use or exposure, (2) *Duration* – length of time a patient has used or quit using tobacco (explicitly separated out from the Temporal element), (3) *Situation* – context in which tobacco is/was used, and (4) *Cessation* – details about cessation such as attempts, interventions, or treatments.

Each comment was analyzed according to these 14 elements plus an element for *Other* where there could be multiple words or phrases associated with a particular element. For example, the comment "1-2 cigarettes/day" would be coded as Type = "cigarettes," Amount = "1-2," and Frequency = "/day" while the comment "no second hand smoke exposure" would be coded as Negation = "no," Method = "exposure," and Type = "second hand smoke". A κ of 0.94 was obtained for coding contents and main analysis involved determining more commonly used words and phrases for each element where groupings were created to combine those with similar meaning or pattern. For example, the words "chews," "chewed," and "chewing" were grouped together for the Method element while the pattern "*n* years ago" covered specific number of years such as "30 years ago" as well as estimated numbers such as "about 10 years ago" and "over 40 years ago".

Analysis of Data Quality Issues in Tobacco Use Entries

The third coding scheme for "issues" was designed to highlight potential data quality issues based on review of both the structured fields and free-text comment field. From review of the initial 100 tobacco use entries, seven issues were identified that were expanded to a total of 12 data quality issues where an entry could be associated with one or more issues (Table 3). For example, an entry where the value "Never Smoker" is specified in the structured smoking status field while the comment states "OCCASSIONAL CIGAR" would be coded with an issue of *Inconsistent smoking status*. As another example, an entry where the quit date is specified as "11/11/2011" and comment is "quit 2 years ago" would be coded with an issue of *Different temporal references*. Similar to the analysis of reasons for use, a κ of 0.91 was achieved for coding issues and main analysis involved determining the most frequent data quality issues across the 525 entries.

Table 3. Coding Scheme for Issues.

#	Issue	Brief Description	Example
1	Inconsistent smoking status	Contents of comment inconsistent with Smoking status field	Smoking status = Never Smoker Comment = OCCASIONAL CIGAR
2	Inconsistent packs/day	Contents of comment inconsistent with Packs/day field	Packs/day = 1.5 Comment = 01/01/2012 1 pack/day
3	Inconsistent years	Contents of comment inconsistent with Years field	Years = 15 Comment = 10 year smoking hx
4	Inconsistent pack years	Contents of comment inconsistent with calculated Pack years	Packs/day = 1 Years = 50 Comment = 50 years 2 ppd = 100+ pack-years
5	Inconsistent type	Contents of comment inconsistent with Type fields	Types = (not specified) Comment = cigars on occasion
6	Inconsistent start or quit date	Contents of comment inconsistent with Start or Quit date field	Quit date = 8/12/75 0:00 Comment = quit 1980's
7	Different levels of granularity	Contents of comment at different granularity level	Packs/day = 0.5 Comment = 5-10 cigarettes daily
8	Different temporal references	Comment includes relative time rather than absolute time	Quit date = 2/2/05 0:00 Comment = Quit smoking 8 years ago
9	Acronym or abbreviation	Comment includes acronym or abbreviation	• occ. cigar • 2 pks a week
10	Misspelling	Comment includes misspelling	• No smiking x3 days per pt • 3-4 cigarrettes a day
11	Ambiguous	Comment includes ambiguous information	• 2-3/week (*# of times, cigarettes, or packs?*) • 2005 (*start or quit year?*)
12	Not tobacco use	Contents of comment not related to tobacco use	• Smokes marijuana daily • Does not consume alcohol

Results

Based on analysis of the 525 tobacco use entries, Figure 1 depicts the distribution of potential reasons for using the free-text comment field. This field was most often used due to limited ability to describe amount (23.2%), frequency (26.9%), dates associated with starting or quitting (28.2%), and passive smoke exposure (17.9%). In addition, 26.9% of the comments included information considered redundant to what was captured in the structured fields such as smoking status and type.

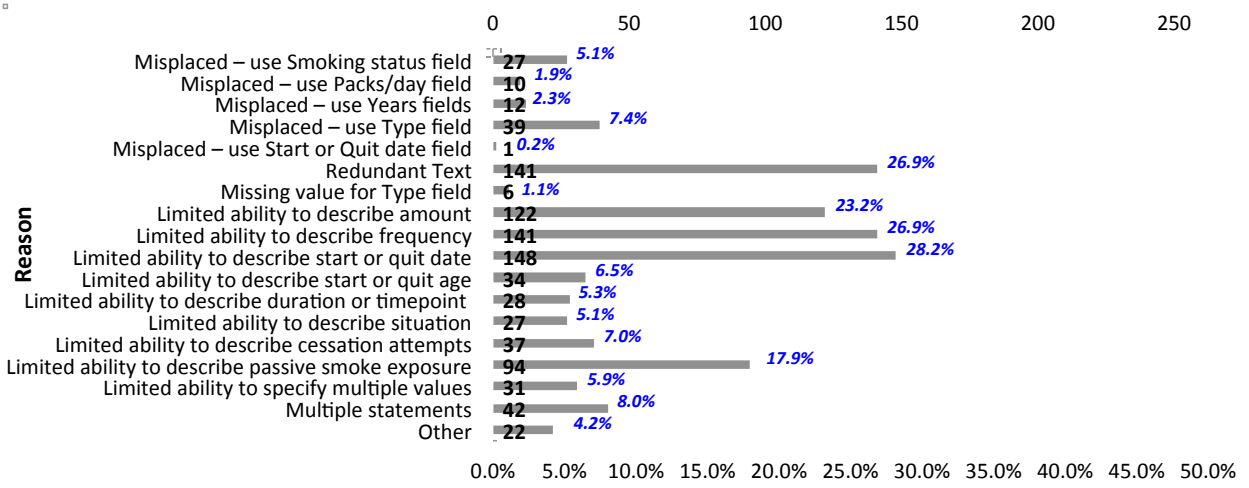

Figure 1. Distribution of Reasons for Use.

With respect to contents, Figure 2 shows the distribution of elements within the free-text comments where words or phrases most frequently described temporal information (38.3%), method (36.6%), type (33.5%), status (31.0%), frequency (29.1%), and amount (28.2%). For each of these elements, Table 4 includes the total number of values (i.e., words or phrases), number of unique values and groups, and the top 3 groups of values along with some examples. For example, of the 220 total values (154 unique values) for the Temporal element that were categorized into 16 groups, 30.0% reflected a specific or estimated year related to use of tobacco or quitting, 27.3% described a specific or estimated number of years ago, and 15.5% provided a specific or estimated age of use or quitting. For the Frequency element, the most frequent words or phrases were related to daily use or use every *n* days where *n* is a specific number or range (50.9%), occasional use (20.0%), and weekly use or use every *n* weeks (13.9%). While occurring less frequently, the majority of phrases categorized as Other were related to decreases in tobacco use (e.g., "cutting back," "down to," and "weaned down") suggesting the need to extend the coding scheme to include an additional element for Change.

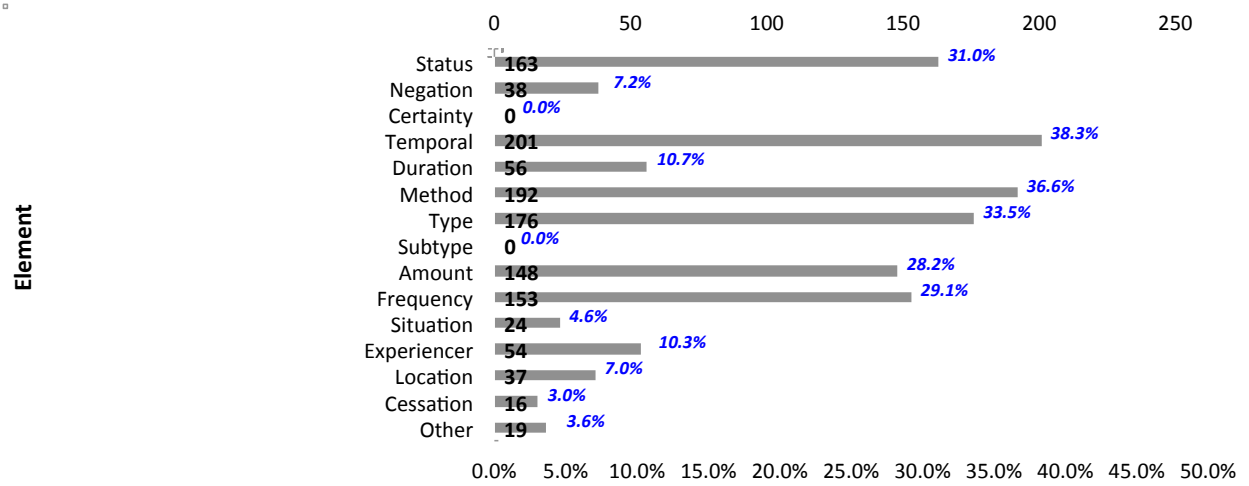

Figure 2. Distribution of Contents.

Table 4. Distribution of Values and Groups of Values for Top 6 Elements.

Element	Total # Values	# Unique Values [# Groups]	Top 3 Groups of Values (*Examples*)	Frequency	
Status	171	24 [8]	• quit (*quit, quitting, former smoker*)	120	(70.2%)
			• smoker (*smoker, smokers*)	18	(10.5%)
			• quit attempt (*trying to quit, process of quitting*)	14	(8.2%)
Temporal	220	154 [16]	• specific or estimated year (*1956, 1970s, early 2000s*)	66	(30.0%)
			• specific or estimated number of years ago (*10 years ago, about 8-10 years ago, over 40 years ago*)	60	(27.3%)
			• specific or estimated age (*age 18, early twenties, teenager*)	24	(15.5%)
Method	195	22 [8]	• exposure (*exposed, exposure*)	110	(56.4%)
			• smoke (*smokes, smoked, smoking*)	61	(31.3%)
			• chew (*chews, chewed, chewing*)	11	(5.6%)
Type	178	36 [13]	• secondhand smoke (*2nd hand, passive smoke, second hand tobacco*)	65	(36.5%)
			• cigarette (*cig., cigs, cigarettes*)	56	(31.5%)
			• cigar (*cigar, cigars*)	25	(14.0%)
Amount	155	85 [17]	• *n* (*3, 7-10, ~8, about 5*)	82	(52.9%)
			• *n* packs (*1/2 pk, 2-3 packs, less than 1 pack, half a pack*)	35	(22.6%)
			• *n* ppd (*1/2-1 PPD, 2-3 ppd, over 1ppd*) *	15	(9.7%)
Frequency	165	60 [22]	• per day or *n* days (*daily, /day, qd, every couple days, 8-10 x/day*)	84	(50.9%)
			• occasional (*occ., now and then, periodically, rarely*)	33	(20.0%)
			• per week or *n* weeks (*weekly, /week, every 2 weeks*)	23	(13.9%)

** addresses both amount and frequency*

Figure 3 reflects the distribution of potential data quality issues associated with the set of tobacco use entries. The most frequent issue was use of acronyms and abbreviations in the comments (18.1%) such as for cigarette or cigarettes where there were 3 different abbreviated forms ("cig,," "cig.," and "cigs") and 2 types of misspellings ("cigarrettes" and "cigarretts"). Other more frequent issues were related to granularity differences (14.1%) such as specifying only the quit year in the comments as opposed to an exact date as provided in the structured quit date field and use of relative rather than absolute temporal references (7.6%) in the comments such as *n* years ago instead of a specific date. Finally, there were several cases of inconsistent number of packs/day (6.7%) that may be due to changes in tobacco use and inconsistent type (5.0%) where the type of tobacco use was specified in the comment but not in the relevant structured fields.

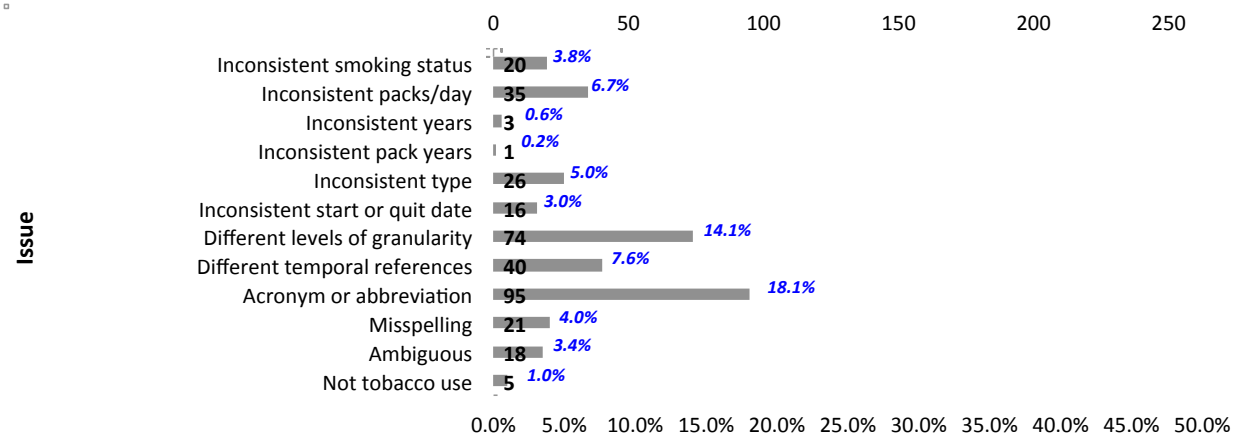

Figure 3. Distribution of Data Quality Issues.

Discussion

The findings of this study provide insights to the current use, contents, and data quality issues associated with free-text tobacco use documentation in the social history module of an EHR at an academic health center. The collective results highlight limitations in capturing details related to tobacco use and secondhand smoke exposure that may be used to inform system enhancements, user training, NLP, and standards for tobacco use documentation. In addition, this work represents a preliminary step towards developing a systematic and semi-automated process for evaluating EHR structure and content. While the study was limited to a single institution with a particular EHR system and focused on a specific module in this system, the overall findings as described below are expected to be generalizable to other institutions that may have the same or different EHR system. It is also anticipated that the three-phased approach for generating, validating, and applying coding schemes to retrospective EHR data could be adapted and applied to accommodate for institutional variations including differences in EHR structure. More broadly, this approach is extensible for performing both quality assessment and content analysis of information in other EHR modules that include free-text and/or structured fields.

The more frequent reasons and contents of the free-text tobacco use comments suggest the need for flexibility in describing amount, frequency, and start or quit dates associated with different types of tobacco or smokeless tobacco (e.g., cigarettes, pipe, cigars, snuff, and chew) that could not be accommodated with existing structured fields (i.e., for packs/day, pack years, and start and quit dates that require that the month, day, and year be specified). Other temporal information within the comments included start age, quit age, and quit duration. Potential system enhancements include incorporating additional structured fields as well as values within existing fields to enable the capture of information such as "occasionally 1-2 cigarettes," "2 cans of chewing tobacco weekly," "quit 1980," "quit in her 20's," "quit x 21 years" in addition to "1 ppd for 5 years" and "quit 2/2/12". The five reasons related to misplaced information (i.e., free-text used instead of available structured fields) were less common; however, they are indicative of gaps in the documentation process that could potentially be addressed through improved user training (e.g., reminders of existing EHR functionality for structured data entry and guidance for when/how to use free-text comments).

While occurring less frequently, there were several entries that included multiple statements or values reflecting changes in status (e.g., patient quit, restarted, and then quit again with associated dates) or amount (e.g., from 0.5 packs to 1 cigarette per day). Such changes could potentially be reflected or re-created by accessing the audit trail for the social history module; however, there may be value in having a more readily-accessible, comprehensive, and flexible "tobacco use history" (or broader "nicotine use history") that could be guided by the findings of this study in addition to existing standards for the representation of tobacco use (Table 5). These standards include those from HL7[31] (e.g., "social history observation," "smoking status observation," and "tobacco use observation" in implementation guides associated with the HL7 Clinical Document Architecture[32-35]) and openEHR[36] (e.g., archetypes for "Tobacco Use" and "Tobacco Use Summary"[37]) that collectively specify the collection of elements and associated values for status, method of use, substance (or type), amount, frequency, start date or age, and quit date or age.

Table 5. Example Tobacco Use History.

Date	Status	Start Date or Age	Quit Date or Age	Duration	Type	Amount	Frequency
2009-07-20	Former	Age: teenager	Date: 30 years ago	Use: 20 years	cigarettes	few	daily
2012-04-16	Current	Date: 2010-08-15			cigarettes	1 pack	weekly
2012-04-16	Current	Date: 2011			cigar	2-3	monthly
2013-10-06	Former	Date: 2011-04		Quit: 6 months	cigar	1	occasionally
2013-10-06	Current	Date: 2010-08-15			cigarettes	0.5 pack	weekly

In addition to describing tobacco use, other uses and contents of the free-text comments were related to secondhand smoke exposure and tobacco cessation (including attempts and interventions). While the list of available values for the structured smoking status field includes one for passive smoking, this is limited to patients who have never smoked and therefore could not be applied to those who were former smokers or who are also current smokers. The occurrence of comments describing exposure or no exposure to secondhand smoke as well as details about experiencer (who smokes – e.g., "parents," "father," or "mother") and location (where smokes – e.g., "outside," "home," or "car") could be used to inform the development of a set of structured fields focused on secondhand smoke exposure.

For tobacco cessation, analysis was performed at a high-level in this study given the breadth of information where further examination is planned to better understand the contents, guide enhancements, and inform how the social history module could promote interventions (e.g., through decision support functionality such as alerts and reminders). As part of this effort, the coding scheme for elements could be extended based on cessation-related elements defined in standards such as the openEHR archetypes for Tobacco Use, Tobacco Use Summary, and Cessation Attempts[37]. Open questions also include determining where and how to document nicotine replacement therapies such as nicotine gum, patches, and inhalers or devices such as electronic cigarettes[38], which were initially coded as missing values for the structured type field when analyzing reasons for use.

Finally, a number of data consistency and other quality issues were noted that could present challenges in using tobacco use information from the social history module for decision support, research, public health, and other primary and secondary uses. In some cases, it was found that information in the free-text comment was inconsistent with the structured fields such as indicating a different status (e.g., never smoker vs. current smoker or former smoker) or number of packs/day where the former could lead to missing or incorrectly identifying patients for tobacco cessation interventions. In other cases, the free-text was found to be the only source of information such as indicating the type of tobacco use (e.g., cigarettes or cigars). For both cases, user training may be one approach for improving and ensuring consistency in documentation prospectively while NLP techniques could be developed to extract details about tobacco use, secondhand smoke exposure, cessation attempts, and interventions both retrospectively and prospectively. Next steps include examining the documentation of tobacco use in other parts of the EHR (e.g., problem list and clinical notes) to further characterize data consistency and quality issues as well as determine how to integrate this information for subsequent uses.

Conclusion

With the increased adoption of EHR systems, there is a need for efforts to explore their potential for improving tobacco use assessment and cessation. This study involved examining the current collection of tobacco use information in the social history module of an EHR with an emphasis on free-text documentation. Based on the preliminary findings, implications for improving the use of information related to tobacco use and secondhand smoke exposure in the EHR include system enhancements, user training, NLP, and standards.

Acknowledgments

Research reported in this manuscript was supported by the National Library of Medicine of the National Institutes of Health under award number R01LM011364. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Centers for Disease Control and Prevention. Smoking & Tobacco Use Fast Facts. [March 2014]; Available from: http://www.cdc.gov/tobacco/data_statistics/fact_sheets/fast_facts/.
2. U.S. Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. 2014 [March 2014]; Available from: http://www.cdc.gov/tobacco/data_statistics/sgr/50th-anniversary/index.htm.
3. World Health Organization. Tobacco Fact Sheet. [March 2014]; Available from: http://www.who.int/mediacentre/factsheets/fs339/en/.
4. World Health Organization. WHO Report on the Global Tobacco Epidemic, 2013. [March 2014]; Available from: http://www.who.int/tobacco/global_report/2013/en/.
5. Centers for Disease Control and Prevention. Smoking & Tobacco Use - Healthy People 2020. [March 2014]; Available from: http://www.cdc.gov/tobacco/basic_information/healthy_people/.
6. Healthy People 2020 Tobacco Use. [March 2014]; Available from: http://www.healthypeople.gov/2020/topicsobjectives2020/overview.aspx?topicid=41.
7. Centers for Disease Control and Prevention. National Tobacco Control Program. [March 2014]; Available from: http://www.cdc.gov/tobacco/tobacco_control_programs/ntcp/.
8. Moyer VA, U. S. Preventive Services Task Force. Primary care interventions to prevent tobacco use in children and adolescents: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2013 Oct 15;159(8):552-7.
9. U. S. Preventive Services Task Force. Counseling and interventions to prevent tobacco use and tobacco-caused disease in adults and pregnant women: U.S. Preventive Services Task Force reaffirmation recommendation statement. Ann Intern Med. 2009 Apr 21;150(8):551-5.

10. World Health Organization. Tobacco Free Initiative. [March 2014]; Available from: http://www.who.int/tobacco/en/.
11. Institute of Medicine. Capturing Social and Behavioral Domains in Electronic Health Records: Phase 1. Washington, DC: The National Academies Press; 2014.
12. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med. 2010 Aug 5;363(6):501-4.
13. Meaningful Use Stage 2 - Record Smoking Status. [March 2014]; Available from: http://www.healthit.gov/providers-professionals/achieve-meaningful-use/core-measures-2/record-smoking-status.
14. eCQM Library. [March 2014]; Available from: http://cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/eCQM_Library.html.
15. Boyle R, Solberg L, Fiore M. Use of electronic health records to support smoking cessation. Cochrane Database Syst Rev. 2011(12):CD008743.
16. Lindholm C, Adsit R, Bain P, Reber PM, Brein T, Redmond L, et al. A demonstration project for using the electronic health record to identify and treat tobacco users. WMJ. 2010 Dec;109(6):335-40.
17. Greenwood DA, Parise CA, MacAller TA, Hankins AI, Harms KR, Pratt LS, et al. Utilizing clinical support staff and electronic health records to increase tobacco use documentation and referrals to a state quitline. J Vasc Nurs. 2012 Dec;30(4):107-11.
18. Mathias JS, Didwania AK, Baker DW. Impact of an electronic alert and order set on smoking cessation medication prescription. Nicotine Tob Res. 2012 Jun;14(6):674-81.
19. Zheng K, Hanauer DA, Padman R, Johnson MP, Hussain AA, Ye W, et al. 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.
20. 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.
21. Chen ES, Garcia-Webb M. An analysis of free-text alcohol use documentation in the electronic health record: early findings and implications. Appl Clin Inform. 2014;5(2):402-15.
22. Zhou L, Mahoney LM, Shakurova A, Goss F, Chang FY, Bates DW, et al. How many medication orders are entered through free-text in EHRs?--a study on hypoglycemic agents. AMIA Annu Symp Proc. 2012;2012:1079-88.
23. Uzuner O, Goldstein I, Luo Y, Kohane I. Identifying patient smoking status from medical discharge records. J Am Med Inform Assoc. 2008 Jan-Feb;15(1):14-24.
24. Liu M, Shah A, Jiang M, Peterson NB, Dai Q, Aldrich MC, et al. A study of transportability of an existing smoking status detection module across institutions. AMIA Annu Symp Proc. 2012;2012:577-86.
25. Hazlehurst B, Frost HR, Sittig DF, Stevens VJ. MediClass: A system for detecting and classifying encounter-based clinical events in any electronic medical record. J Am Med Inform Assoc. 2005 Sep-Oct;12(5):517-29.
26. Hazlehurst B, Sittig DF, Stevens VJ, Smith KS, Hollis JF, Vogt TM, et al. Natural language processing in the electronic medical record: assessing clinician adherence to tobacco treatment guidelines. Am J Prev Med. 2005 Dec;29(5):434-9.
27. Wu CY, Chang CK, Robson D, Jackson R, Chen SJ, Hayes RD, et al. Evaluation of smoking status identification using electronic health records and open-text information in a large mental health case register. PLoS One. 2013;8(9):e74262.
28. Epic Systems Corporation. [March 2014]; Available from: http://www.epic.com/.
29. Chen ES, Manaktala S, Sarkar IN, Melton GB. A multi-site content analysis of social history information in clinical notes. AMIA Annu Symp Proc. 2011;2011:227-36.
30. Melton GB, Manaktala S, Sarkar IN, Chen ES. Social and behavioral history information in public health datasets. AMIA Annu Symp Proc. 2012;2012:625-34.
31. Health Level Seven International (HL7). [March 2014]; Available from: http://www.hl7.org/.
32. Dolin RH, Alschuler L, Boyer S, Beebe C, Behlen FM, Biron PV, et al. HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc. 2006 Jan-Feb;13(1):30-9.
33. HL7/ASTM Implementation Guide for CDA® R2 -Continuity of Care Document (CCD®) Release 1. [March 2014]; Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=6.
34. HL7 Implementation Guide for CDA® R2: History and Physical (H&P) Notes, Release 1. [March 2014]; Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=19.
35. HL7 Implementation Guide for CDA® Release 2: IHE Health Story Consolidation, Release 1.1 - US Realm. [March 2014]; Available from: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=258.
36. openEHR. [March 2014]; Available from: http://www.openehr.org/.
37. openEHR Clinical Knowledge Manager. [March 2014]; Available from: http://www.openehr.org/ckm/.
38. Fairchild AL, Bayer R, Colgrove J. The renormalization of smoking? E-cigarettes and the tobacco "endgame". N Engl J Med. 2014 Jan 23;370(4):293-5.