INTERNATIONAL JOURNAL OF MEDICAL INFORMATICS 76S (2007) S122–S128

journal homepage: www.intl.elsevierhealth.com/journals/ijmi

Copying and pasting of examinations within the electronic medical record

Stephen Thielke[a], *Kenric Hammond*[a,b,c,*], *Susan Helbig*[c]

[a] *Department of Psychiatry, University of Washington Seattle, Washington, United States*
[b] *Department of Medical Education and Biomedical and Health Informatics, University of Washington, Seattle, Washington, United States*
[c] *Department of Veteran's Affairs Puget Sound Health Care System, Seattle, Washington, United States*

ARTICLE INFO

Article history:
Received 9 March 2006
Accepted 20 June 2006

Keywords:
Computerized medical record
Physical examination
Mental status examination
Copying processes
Databases

ABSTRACT

Electronic patient records often include text that has been copied and pasted from other records. A type of copying that involves the highest risk for confusion, medical error, and medico-legal harm is the copying of the clinical examination. We studied this phenomenon using an automated text categorization algorithm to detect copied exams in a set of 167,076 VA records. Exam copying occurred frequently, in about 3% of all exams, or in 25% of patient charts. Thirteen percent of all authors had copied at least one exam, and 3% of authors had copied an exam from another author. There were significant differences between service types and levels of training of the authors. We speculate that copying and pasting of exams degrades the quality of the medical record, and that studying this behavior is integral to our understanding of phenomenology of the electronic medical record.

Published by Elsevier Ireland Ltd

1. Introduction

The electronic medical record makes copying and pasting of text an easy process. Some users have reported finding that another user had copied text they had written, and describe this as an unnerving experience akin to finding that one's work had been plagiarized. During my [Thielke's] first rotation at the VA Medical Center as a medical student in 2000, one of the residents told the students, "the VA is a great place to work because with the computers you can just copy and paste your notes, and you don't need to write a new note every day", and other students suggested copying other author's notes as a way to save time. Students frequently copied clinical exams from day to day on inpatient notes, and there was no discussion of the appropriateness of this practice, or its effect on documentation or patient care.

Copying and pasting of text within electronic medical records has been recently studied [1,2], but has not been investigated on a broad scale. In a manual review of 6322 records containing at least 40 words of copied text, Hammond et al. discovered 66 "highest risk" and 272 "high risk" copying events (see the scale of risk in [1]). About half (31) of the highest risk copying events were clinical examinations copied from more than 6 months in the past, or from another clinician. Because of the small sample size, it was not possible to subclassify the copied exams further.

The copying of a clinical examination was categorized as "highest risk" because the examination allegedly records the writer's direct observations in a clinical encounter. While most of the other content categories (i.e. present illness, past medical history, medications, family history, etc.) are relatively static and external to the visit, the examination is entirely dependent on what happened while the patient and clinician were present. If this text is copied from another note, the reader cannot be certain what form the second encounter took, or if it took place at all. Physicians, interviewed by Embi in

* *Corresponding author*. Tel.: +1 253 583 1708; fax: +1 253 589 4167.
E-mail address: khammond@u.washington.edu (K. Hammond).
1386-5056/$ – see front matter. Published by Elsevier Ireland Ltd
doi:10.1016/j.ijmedinf.2006.06.004

a survey of attitudes about computerized physician documentation, perceived that the use of the copy-and-paste function "resulted in the propagation of misinformation or even in frank errors" [3]. We thus hypothesized that from a medico-legal and quality assurance perspective, the copying of clinical examinations could be a problematic behavior, but we were unclear about who was doing it, in what settings, and how often.

It seemed important to characterize exam copying on a broad scale, but the slow pace of manually reviewing copying episodes precluded finding enough data to define more precisely how and when clinical exam copying happens [1]. Automated techniques permitted mining more data to characterize the phenomenon better and allow more meaningful speculation about the effects of exam copying.

2. Methods

The 31 cases of copied examinations disclosed in manual review (6322 copying episodes, 29,386 source notes) occurred at a low incidence of 0.1%, so a large sample size was needed to produce enough samples to describe this phenomenon. Starting with 167,076 notes for 1479 randomly selected patients from 1990 to 2002, 90,702 instances of at least 40-word duplications were selected using the methodology described by Hammond et al. [1].

We developed an algorithm for the automatic selection of clinical exams using lexical frequency. The concentrated, distinct vocabulary of clinical examinations facilitated the automated text classification task. Scripts written in PERL (ActiveState ActivePerl), and queries written in Microsoft Access carried out the functions described below.

First, we developed a training set of well-formed examinations. From 80,000 records, we used automated methods to extract 1850 instances of examinations, and confirmed these by manual review. The resultant sample examinations clustered lexically into three areas: complete and partial physical exams; podiatry exams limited to the foot; and mental status exams (using a psychiatric vocabulary). One thousand and one hundred sixty-eight physical exams, 224 podiatry exams, and 336 mental status exams were found. For each set, words frequency lists were produced. Table 1 lists the first 20 words in frequency for the mental status exam, physical exam, and podiatry exam.

"Negative" frequency lists were developed from the text of notes that remained after the examination portion was removed, for each of the three exam types. Thus, frequency lists were produced for (1) mental status exam, (2) not mental status exam, (3) physical exam, (4) not physical exam, (5) podiatry exam, and (6) not podiatry exam.

For each word found in a portion of copied text to be analyzed, a numerical score was assigned in order to relate the frequency of the word within the copied text and its frequency within the word lists, as shown in Eq. (1).

$$\log_2 \frac{\text{Freq(word|frequency list)}}{\text{Freq(word|not frequency list)}} \times \text{Freq(word|sample text)} \quad (1)$$

For instance, if a word occurred twice as often in the mental status frequency list as in the associated negative frequency list, and it occurred five times in a 50-word segment of text, the score for this word would be $\log_2(2/1) \times (5/50) = 0.1$. Words occurring more of in "negative" text than in exam text produced a negative number. These individual likelihood scores were summed for each word in the copied text segment using each of the three exam types, yielding three scores for the text segment, each representing the likelihood that the text was a physical exam, podiatry exam, or mental status exam.

In order to assess the sensitivity and specificity of this method for categorizing note text, and in order to define the appropriate cutoffs for inclusion in the three sets of exams,

Table 1 – Lexical frequency within different exam types; 20 most common words of each

Mental status		Physical		Podiatry exam	
Words	Frequency (%)	Words	Frequency (%)	Words	Frequency (%)
And	4.04	No	3.82	Bl	3.18
Is	2.68	And	2.28	And	2.59
He	2.17	To	1.57	With	2.42
To	1.77	In	1.57	No	2.09
Of	1.74	The	1.29	B	2.09
No	1.48	With	1.29	To	1.47
Or	1.30	Of	0.99	Of	1.38
The	1.26	Is	0.96	Foot	1.35
Was	1.26	Intact	0.90	Intact	1.09
With	1.12	Or	0.84	Pain	1.02
Mood	1.08	On	0.73	Derm	0.95
His	1.05	Clear	0.73	Neuro	0.95
Thought	1.05	Right	0.73	Pt	0.95
In	1.03	Edema	0.66	Met	0.88
Affect	0.98	Soft	0.64	Msk	0.85
Speech	0.90	Neck	0.60	1st	0.85
A	0.77	A	0.59	Vasc	0.85
But	0.72	Abd	0.59	R	0.74