Are Electronic Medical Records Trustworthy?
Observations on Copying, Pasting and Duplication

Kenric W. Hammond, MD[1,2], Susan T. Helbig[1], MA, RHIA,
Craig C. Benson[1,2], MS, Beverly M. Brathwaite-Sketoe[1], BS,
US Department of Veterans Affairs Puget Sound Health Care System[1] and
Department of Medical Education and Biomedical Informatics,
University of Washington[2],
Seattle and Tacoma, Washington

Abstract

As routine use of on-line progress notes in US Department of Veterans Affairs facilities grew rapidly in the past decade, health information managers and clinicians began to notice that authors sometimes copied text from old notes into new notes. Other sources of duplication were document templates that inserted boilerplate text or patient data into notes. Word-processing and templates aided the transition to electronic notes, but enabled author copying and sometimes led to lengthy, hard-to-read records stuffed with data already available on-line. Investigators at a VA center recognized for pioneering a fully electronic record system analyzed author copying and template-generated duplication with adapted plagiarism-detection software. Nine percent of progress notes studied contained copied or duplicated text. Most copying and duplication was benign, but some introduced misleading errors into the record and some seemed possibly unethical or potentially unsafe. High-risk author copying occurred once for every 720 notes, but one in ten electronic charts contained an instance of high-risk copying. Careless copying threatens the integrity of on-line records. Clear policies, practitioner consciousness-raising and development of effective monitoring procedures are recommended to protect the value of electronic patient records.

Introduction

Access to on-line text records affords rapid retrieval of more complete information and enhances care[1]. Achieving an on-line record environment requires clinician involvement, major capital investment, and visionary, effective leadership[2]. Transitioning an institution from handwritten records is a huge task, and requires adequate support for users to either type or dictate clinical notes. This support must include responsive transcription services, many workstations, and may employ specialized templates and automated insertion of patient data to aid direct entry. All of these methods were available when the Veterans Affairs (VA) Puget Sound Health Care System expanded its clinical information system, fully implementing order entry and patient document functions available in the nationally used Department of Veterans Affairs Computerized Patient Record System (CPRS). An unintended consequence of deploying on-line records (and the various methods to aid their creation) was growth in copying and re-use of text from elsewhere in the electronic chart. This paper describes the extent and implications of the phenomenon and discusses approaches to corrective action.

Background

VA Medical Records managers and clinicians across the country first began to notice unattributed copying of text from one clinical document into another using Windows® copy and paste functions in the late 1990s, paralleling expanded use of CPRS and replacement of data terminals with microcomputer workstations. The copying was easiest to note among records belonging to a single patient. Concern arose over potential for propagation of erroneous information, lapsed professional ethics, billing irregularities, and potential liability exposure. Disabling copy and paste functionality was considered impractical. Other worries were expressed about errors and poor readability caused by templates and inserted data. Starting in the year 2000, several draft policies enunciating guidelines and policies for electronic notes circulated in the VA community. Some of these recommended monitoring without specifying a methodology. Development of policy is ongoing.

In recognition of its quality improvement programs, Veterans Integrated Service Network 20, located in the Pacific Northwest, received a $500,000 Kenneth Kizer Quality Improvement Award (honoring the former VA Undersecretary for Health) to support

local quality improvement projects. The present project received a Kizer grant. Inspiration for the method adopted came from a related problem in academia – plagiarism – facilitated by word-processing software and availability of vast quantities of machine-readable text on the Internet.

Methodology

A survey revealed several commercial and public use software resources to detect copying. One of these, Copyfind, a well-documented GNU General Public Use License program developed by Bloomfield, showed promise[3]. Copyfind performs pair-wise comparisons of all documents in a set to identify identical word sequences. When a user-defined minimum sequence matches, the program creates two HTML files, underlining the matching portion in each document. Copyfind was modified (Copyfind-VA) to communicate with a relational database of patient documents using Open Data Base Connectivity (ODBC) conventions. Figure 1 shows an example of the output.

All patient documents and document identifiers present in the VA CPRS of VA Puget Sound between October 1990 and June 15, 2002 were downloaded. A second download captured document identifiers, case identifiers, creation dates, authors, and clinical departments associated with the documents. These data were imported into linked Access® databases containing 4.5 million documents and requiring 17 Gigabytes of disc storage. Notes for 98,753 individuals were captured. The project was an integral part of the facility's quality assurance program and patient consents were not obtained. To safeguard privacy all databases were encrypted and password-protected, and maintained on a secure machine in the medical file department.

The following procedures were performed to analyze records for copying. For each entry in a list of case identifiers, a read procedure retrieved all documents belonging to a case. Next, Copyfind-VA compared each retrieved document to every other document in the set. When at least forty consecutive words in compared documents matched exactly, a pair of marked-up text records and document identifiers was written to a database of "copyevents". The fully permuted comparison created two records for each episode of duplication, but document identifiers permitted identifying original and recipient documents. The procedure repeated until all cases were read and analyzed and associated copyevents recorded.

Our study examined a random sample of the approximately 15,000 VA Puget Sound patients who had at least one progress note recorded between May 15 and June 15, 2002. For each of these patients, all electronic notes were retrieved from the master databases. The resulting data set contained 167,076 progress notes for 1,479 patients. After analysis with Copyfind-VA, 181,404 copyevent records, representing 90,702 instances when pairs of documents contained identical forty-word sequences, were created.

Manual Review

A data viewing and coding screen was created using an Access® form. This form stepped through each "original" document for a patient, permitting viewing markups of "original" and "recipient" document pairs. Each instance was coded for risk severity and the duplicated information was classified. Risk was rated on a six-point scale and twenty-one categories of information were identified. (Tables 1 and 2).

Table 1. Severity scale

Code	Risk Description
1	Artifact, not misleading, no risk
2	Artifact, minimally misleading, minimal risk
3	Human, not misleading, no risk
4	Human, minimally misleading, minimal risk
5	Human, misleading, some risk
6	Human, clinically misleading, major risk

Table 2. Categories of duplicated information

Present Illness	Vital signs
Past History	Boilerplate
Medications	Plan
Problem List	Diagnosis
Family History	Chief Complaint
Social History	Diagnostic Data
Systems Review	System Error
Examination	Allergies
Laboratory Data	Other
Assessment	Multiple
Treatment	

Discussion of Risk Ratings

"Artifact" referred to system-created boilerplate and inserted patient data. Documents constructed with templates and data objects often yielded duplications identified by Copyfind-VA. "Human" referred to duplications caused by an author copying text

Outpatient Visit Note, 10/16/01	Student Note, 5/30/02
VITALS: BP:136/73 HR:80 Wt:246.4 lb PN 2/10 rt heel S. 57 year old RTC to p/u new FFO. Pt complains of heel pain rt only subsideing slowly with new orthoses; <u>PMH: PTSD, depressing, GERD 79 pack years, quit smoking three years ago. Currently sober & for THE PAST 3+ years O- Vasc: DP/PTpalable b/l, TTT intact b/l Neuro: Semes weinstein 5.07/10g monofilament wire sensation intact b/l epicratic sensation intact b/l Derm: toenails 1-5 b ft thickwened brittle incurvated painful with yelow subungual debris distal 1/3 only Musc: strenght intact, ROM intact FLEXIble PES cavus B/L Flexible hammertoes b/l Pinpoint pain with palpable medial heel r only A. 1. Plantar fasciitis r>l CHRONIC 2. B/l PES cavus 3. onychomycosis 1-5 b ft P. continue FOOTMAXX FFO</u> rtc May 02 renew naprosyn 2 tabs bid # 120	VITALS: 05/30/2002 08:55 BP:127/63 HR:72 Wt:255 lb [115.9 kg] S. 57 year old RTC to p/u new FFO. Pt complains of heel pain rt only subsideing with new orthoses; Pt cont to take the naproxen for pain relief. Pt states clotrimazole soln is working well for toenail fungus. <u>PMH: PTSD, depressing, GERD 79 pack years, quit smoking three years ago. Currently sober & for THE PAST 3+ years O- Vasc: DP/PT palable b/l, TTT intact b/l Neuro: Semes weinstein 5.07/10g mono-filament wire sensation intact b/l epicratic sensation intact b/l Derm: toenails 1-5 b ft thickwened brittle incurvated painful with yelow subungual debris distal 1/3 only Musc: strenght intact, ROM intact FLEXIble PES cavus B/L Flexible hammertoes b/l Pinpoint pain with palpable medial heel r only A. 1. Plantar fasciitis r>l CHRONIC 2. B/l PES cavus 3. onychomycosis1-5 b ft P. continue FOOTMAXX FFO</u> cont naproxen, cont clotrimazole soln rtc Aug 02 to be rescanned for new footmax ffos Pt and tx d/w Dr. XXXX.

Figure 1. Marked-up progress note showing copied text (and rated "Human, clinically misleading, major risk").

verbatim from an existing patient document. Duplicated information that appeared valid and incapable of leading a reader to an erroneous conclusion was classified as not misleading. Minimally misleading referred to possible but not definite doubt about the information, e.g., when part of a history or a problem list was copied from one time period to another. The judgment depended on both the type of information copied and the elapsed interval. A past medical history copied from three months before was judged minimally misleading because it might not have been verified. A history of present illness copied from three months before was classified as misleading because of the improbability that a new illness episode would present with *exactly* the same description. Copied information that was obviously inaccurate or inauthentic, e.g., a cloned, out of date physical exam, or a record for billed services where the fact of copying gave doubt that history taking or cognitive services were performed, was classified as clinically misleading. These attributes were combined into an estimate of risk, ranging from no risk to major *potential* risk of patient harm, fraud or tort claim exposure. A physician with twenty-five years of quality review and informatics experience and a medical records specialist trained in a university certificate program performed the ratings. Formal inter-rater reliability analysis was not performed, but raters discussed all records ranked as risk categories 5 or 6 until consensus was reached.

Restrictions

The initial observations showed that text duplication was extensive and complex. While we anticipated finding much "copying" due to boilerplate and inserted data, the observed ratio of copyevents to total documents, exceeding 50%, was striking. To conserve resources for expert review, the next phase of the study addressed filtering out both trivial, and non-trivial, but very frequent, instances of duplication as we proceeded. This key decision meant that the prevalence estimates cited below are *underestimates* of the true prevalence of problematic copying. This semi-quantitative approach was appropriate to a *quality improvement* goal of identifying as many opportunities as possible to improve care. Once identified, a problematic pattern of copying could be targeted for corrective action.

Accordingly, we identified document titles most often associated with copyevents. Initially we identified titles responsible for at least 1% of copy events and inspected them. When the note's template structure contained boilerplate or inserted data highly likely to register a copyevent, it was eliminated from consideration. High-yielding titles associated with human copying were retained, but some were eliminated when an obvious copying pattern was identified in the analysis. For example, practitioners in one clinic routinely copied and edited prior visit notes, often not very carefully, yielding many higher-risk copy events. No further analysis was required to identify

an opportunity for corrective action. Discharge summaries were eliminated because copying was justified. As the analysis proceeded, we found additional note titles invariably associated with copying due to artifact or systematic author practice, and eliminated them, after identifying remedial action. 63 note types (of 874 found) were eliminated; reducing the number of copyevents for manual review by 68%, but 29,009 copyevents for 1,479 patients remained.

Results

When the review ended, 6,322 copy events for 243 patients (1.6% of the one-month sampling cohort) were reviewed. Despite only scratching the surface, the yield of information was sufficient to show the pervasiveness of the copying phenomenon. We found the following results:

1. Cases examined: 243.
2. Time frame: 1993-2002.
3. Total (copied and uncopied) notes for cases between 1993 and 2002: 29,386.
4. Notes containing copied text: 2,645 (9.0% of total notes*).
5. Copying Risk 5 or 6: 338 (1.2 % of total notes).
6. Copying Risk 5: 294 (1.0 % of total notes).
7. Copying Risk 6: 44 (0.15 % of total notes).
8. *Case* prevalence of either type 5 or 6: 89 (36.6%).
9. *Case* prevalence of type 5: 75/243 or 30.8%.
10. *Case* prevalence of type 6: 25/243 or 10.2%.
11. Prevalence of human copying: 63% of copied notes; 5.2% of total notes.

* The true prevalence of copying or duplication. A single note with content duplicated in several prior notes signals one actual instance of copying plus numerous possible instances of copying.

Table 3 traces the copying phenomenon that began in 1995 and grew precisely when workstations were introduced. Table 4 categorizes patient information copied among the 44 notes in the highest risk group.

Billed Encounters

In May 2002, 164 visits for our sample resulted in third party billings. We analyzed notes documenting these encounters and found the risk distribution shown in Table 5. The highest-risk note consisted of 80% copied text.

Table 3: Annual distribution of copying for 243 cases

Copyevents		Notes with copied text	Total number of notes
1995	6	2	20
1996	8	6	32
1997	37	13	171
1998	672	164	1,047
1999	805	265	3,157
2000	1,667	631	6,379
2001	2,200	867	10,989
2002*	925	422	7565

Total notes for 1995-2002: 29,360
* 2002 data based on 5.5 months experience

Table 4. Categories of highest risk copyevents.

Type of information copied	Occurrence
Examination (physical, mental)	31
History of Present Illness	9
Past Medical History	8
Assessment	5
System Error	4
Problem List	3
Review of Systems	2
Multiple Type	2
Chief Complaint	1
Other	1

Table 5. Risk distribution for billed visits.

Risk	N	%
Major Risk (human)	1	0.6 %
Some Risk (human)	2	1.2 %
Minimal Risk (human)	8	4.9 %
No Risk (human)	26	16.0 %
Minimal Risk (artifact)	2	1.2 %
No Risk (artifact)	9	5.5 %

Discussion

These results reveal behavior that cannot comfortably be ignored. Authors were responsible for 63% of the copying found in notes containing copied material (after filtering). Unlike machine copy-artifact, human copying is hard to detect without technical aid. Re-copied problem lists usually were benign and useful, but sometimes appeared to propagate errors. The variety, creativity and subtlety of human copying ef-

forts were broad. Without Copyfind-VA it would have been very difficult to distinguish valid from invalid records. With it, many innocent-appearing records raised doubts.

Motivations for author-based copying surely varied, but generally, we infer that the primary motivation was to produce effective documentation efficiently. Most copying appeared to be the work of busy clinicians striving to be productive. Sometimes we pictured industrious students synthesizing complex histories to benefit the teaching team. At other times it appeared that the copier was naïve, but we also saw authors being careless and possibly lazy. A few times we believed that unethical and potentially dangerous behavior had occurred.

Early on we envisioned using Copyfind-VA to monitor and encourage reasonable copying practices, but found that the volume of copying overwhelmed our technology. Instead of giving useful feedback to individuals, we earned the awkward task of delivering good news and bad news to executive leaders. For example, although we found a very low rate of highest-risk copying (one type 6 event per 720 notes), *one in ten* cases reviewed registered at least one high-risk event. We examined a twelve-year time frame for copying, but very little copying occurred before 1998 (Table 3), showing that the problem is very current. Restricting the analysis to notes written in the most recent twelve *months* would only have reduced the number of filtered copyevents for the 1,479-patient sample by 43%.

After our study was completed we learned that Weir et al.[4] found a 20% prevalence of notes containing copied text in a manually reviewed set of 60 inpatient CPRS charts at the Salt Lake City VA Health Care System. Their detailed analysis found an average of one factual error introduced into the electronic record per human- or machine-effected copying episode. The higher prevalence of copying they found likely came from examining inpatient notes rather than all notes.

Copyfind-VA succeeded in diagnosing and more precisely characterizing the extent of a problem we knew existed but had only vaguely perceived. In its present form it lacks the specificity needed to make it a practical and efficient tool for general record review and monitoring, but it may be of value in selective applications such as auditing records submitted for billing. It is clear that an effective remedy must be pursued. Disabling copy/paste functionality is not recommended because its crippling effect on the valuable labor of composing and editing electronic

notes. We believe that a more effective approach will involve user education, enunciation of strong guidelines for note-writing practices, and introduction of effective monitoring systems with user and supervisory feedback. To this end we propose that institutions using or contemplating adoption of electronic medical records consider the following:

1. Re-engineer templates to avoid unnecessary duplication artifact.
2. Minimize inserting patient data available elsewhere into the narrative record.
3. Develop medical history and examination data objects that can be reviewed, amended and re-used.
4. Enhance the problem list function as a better alternative to copying text lists.
5. Enhance automated methods to more efficiently monitor for dangerous and misleading copying.
6. Caution clinical departments against excessive use of copying to boost productivity.
7. Teach practitioners and students that careless copying creates untrustworthy records.
8. Empower teachers to monitor the writings of trainees with automated methods.
9. Adopt policy stating that unethical copying is unacceptable.
10. Promote ethical electronic documentation early in training.
11. Require source attribution when copied text is re-used in patient records.

References

1. Tang PT, LaRosa PA, Gorden SM. Use of computer-based records, completeness of documentation and appropriateness of documented clinical decisions. J Med Inform Assoc 1999; 6 (3): 245-251.

2. Doolan DF, Bates DW, James BC. The use of computers for clinical care: a case series of advanced U.S. sites. J Am Med Inform Assoc 2003; 10 (1):94-107.

3. Bloomfield, L: Plagiarism Resource Center, http://plagiarism.phys.virginia.edu/

4. Weir CR, Hurdle JF, Felgar MA, et al. Direct text entry in electronic progress notes: an evaluation of input errors. Methods Inf Med 2003; 42: 61-67.