

Understanding the EMR Error Control Practices among Gynecologic Physicians

Ritu Khare

Drexel University College of
Medicine

ritu.khare@nih.gov

Yuan An

Drexel University College of
Information Science & Technology

yuan.an@ischool.drexel.edu

Sandra Wolf

Drexel University College of
Medicine

sandra.wolf@DrexelMed.edu

Paul Nyirjesy

Drexel University College of
Medicine

paul.nyirjesy@DrexelMed.edu

Longjian Liu

Drexel University School of Public
Health

l185@drexel.edu

Edgar Chou

Drexel University College of
Medicine

edgar.chou@DrexelMed.edu

Abstract

An “EMR error” refers to any incorrect, incomplete, or inconsistent patient information entered into electronic medical records (EMRs). Currently, the administering clinicians “manually” resolve such errors. Designing automated error control algorithms is a significant, and yet under-explored, informatics problem. In this study, we assess the EMR error detection abilities of physicians, reveal their strategies, and draw implications for computational algorithm design. Focusing on gynecologic practice, we conducted an error simulation study by fabricating several “erroneous” patient visit notes. We presented these notes to 20 experienced gynecologists, and asked them to detect any errors. Despite devoting substantial time, the participants could detect <50% of the introduced errors. Nevertheless, the successful cases helped reveal the 5 kinds of automatable “triggers” that helped participants sense an error candidate. The participants were able to recognize these triggers because of their comprehensive gynecologic knowledge accumulated through experience and medical school training.

Keywords: algorithms, data errors, EMRs, physicians, user study

Motivation

Electronic medical records (EMRs) have revolutionized the accessibility, legibility, and decisive ability of patient health information. However, the unusable EMR interfaces, situated within a demanding clinical environment, make the process of data-entry very error-prone. Quite often, clinicians inadvertently make mistakes while documenting patient visits and diagnosis information, and thereby commit “EMR errors,” which include incomplete, inaccurate, or inconsistent information (Brown & Patterson, 2001; Phillips & Gong, 2009). EMR errors are expensive; not only do they lead to poor data quality, but also they have the potential to cause unsafe quality of care, and to hold the physicians liable for medical malpractice (Classen, Pestotnik, Evans, Lloyd, & Burke, 1997; Fichman, Kohli, & Krishnan, 2011). Given the gravity of the problem, informatics research should actively engage in developing computational error control algorithms to alert the physicians in real time, and minimize further medical errors (Redwood, Rajakumar, Hodson, & Coleman, 2011). Since the EMR errors are largely underexplored, it is important

Acknowledgements: We sincerely thank the two anonymous reviewers for providing valuable suggestions for revising this paper.

This research is supported in part by the Drexel’s Jumpstart grant for health informatics. The user study was approved from the Institutional Review Board at Drexel University. We thank all the user study participants for sharing their wisdom and time.

Khare, R., An, Y., Wolf, S., Nyirjesy, P., Liu, L., & Chou, E. (2013). Understanding the EMR error control practices among gynecologic physicians. *iConference 2013 Proceedings* (pp. 289-301). doi:10.9776/13197

Copyright is held by the authors.

to follow an “inside-out” approach to develop algorithms, i.e., to (i) first understand the existing error control mechanisms, and (ii) then design the algorithms according to the observed limitations, and opportunities. Given the limited error control functionality provided by the existing vendor-designed EMRs, clinicians resort to certain “manual” techniques to review, detect, and resolve the errors (Phillips & Gong, 2009). In this study, we take the first step toward algorithm development, and systematically investigate the manual error control practices. We assess the abilities of physicians to detect a variety of EMR errors, elucidate their strategies, and accordingly, derive implications for algorithm design.

To understand the existing error control practices, we conducted error simulation and user study on the physicians who are responsible for electronically documenting a wide range of patient problems and visit information. We selected the extremely vital medical field of gynecology because one of the key investigators of the study worked as a data scientist in a women’s health research team affiliated with the College of Medicine at Drexel University. As a result, the team had a close interactions with the gynecologic physicians who document a variety of information pertaining to yeast infections, bacterial vaginitis, menstrual cycle issues, pre-natal and post-natal complaints, regular gynecologic examination, etc. In the future, we plan to conduct similar studies in other medical areas to validate the findings of this gynecology-specific study.

We conducted the study in context of outpatient clinics wherein clinicians document the patient visits into the EMRs in an on-the-spot “narrative” manner. The documentation occurs under extreme time constraints, and hence such an unstructured documentation was conducive to a variety of data errors (George & Bernstein, 2009). To simulate the clinic environment, we fabricated 7 gynecologic visit scenarios, and developed the corresponding EMR patient visit notes; one such note is shown in the Figure 1. To simulate the error-prone nature of data-entry, we purposefully introduced 97 errors of 5 different kinds into the notes. We conducted a user study individually with 20 gynecologic physicians having extensive experience with EMR visit documentation. The participants were presented with the flawed notes, and were asked to identify any data errors. The error detection step was followed up with a debriefing discussion to reveal the error detection and resolution strategies adopted by the participants.

We find that the participants could detect only 49% of the inaccuracy and inconsistency errors from the notes, and only 36% of the omission errors from the notes. This clearly indicates the need for developing automated algorithms that not only save time, but also provide a more effective error control solution. While the task performance of participants was very limited, we find that the strategies adopted by the participants are very important in developing guidelines for designing computational algorithms. The debriefing discussion suggests that there are certain data triggers that naturally prompt participants to sense a potential error. In our user study with gynecologic narrative notes, the participants relied on five kinds of error detection triggers: *detection of abnormal examination results; recall of generic clinical guidelines; detection of abnormal history events; observation of discrepant information; and identification of broken information links*. While decoding their strategies, we find that participants not only have the natural language processing (NLP) abilities, but also an immense amount of intuitive domain knowledge accumulated through experience and medical school training. To simulate such behavior, in addition to sophisticated NLP techniques, the algorithms should incorporate a wide range of federally¹ established free resources for clinical guidelines, controlled vocabularies, drugs, diseases, drug indications, gynecologic best practices, drug interactions, etc. We briefly provide the linkages between the triggers, and the relevant trustworthy knowledge sources. The key contribution of this study is that, as a pre-step to design EMR error control algorithms, we explore an untapped knowledge source, i.e., the physicians, and learn algorithm design lessons from their abilities and behaviors. We plan to use the results of this study in implementing customized algorithms for the narrative EMR data specific to gynecologic patients.

The remainder of the paper is organized in the following manner. We first provide a background on patient visit notes and the typical errors associated with them. We then describe the results of the user study conducted with the gynecologic physicians, discuss the related literature, and conclude the paper.

¹ <http://www.hhs.gov>

ABC, ABC, 32 yr old F, DOB: DD-MM-YY	
Type: Amb Ret Pt Visit Owner: XYZ, XYZ	Status: Final
Reason for Visit ANNUAL VISIT	Inconsistent Information
Allergies No Known Drug Allergy	
Current Meds None	Miscellaneous Error
Active Problems Vaginal Polyp(623.7); c/w hymenalring Visit For: Postpartum Exam(V24.2)	
PMH Patient reported no PMH	Missing Sections
PSH Supervision of normal first pregnancy (V22.0)	
Family Hx Hypertension Diabetes mellitus No cancer	Incomplete Information
Gyn History PAP HISTORY Date of last pap: a year ago History of Abnormal Pap: Yes	
OB History G1P1001 svd x 1	Incorrect Information
ROS ROS: reviewed intake sheet	
Vital Signs Vital Signs Recorded by LMN on MM-DD-YY BP:122/84 mm Hg, Height: 70in, Weight: 224 lb, BMI: 32.07, BSA: 2.19	
Physical Exam All normal	
Assessment • Routine pelvic exam with cervical pap smear(V72.31)	
Plan - rx for pnv - sbe/diet/exercise	
Orders PNV- DHA 27-0.6-0.4-300MG Capsule; take 2 capsules daily; Qty 30; R11; Rx	
Signature Electronically Signed by XYZ, MM-DD-YY	

Figure 1. A (fabricated) visit note highlighting different kinds of errors

EMR Visit Notes and Data-entry Errors

In a typical outpatient clinic, the affiliated physicians administer several patient visits on a daily basis. For each visit, the associated provider is solely responsible for documenting complete visit information into the EMRs. Despite the provision to enter structured information into EMRs, providers find it far more efficient to enter impromptu narrative notes during patient visits (Doğan et al., 2010). Such a “patient visit note” should ideally record complete information associated with the visit; e.g., a typical gynecologic note is organized into the following 19 “sections” in that order: (1) *Reason for Visit*, (2) *Chief Complaint*, (3) *History of Present Illness (HPI)*, (4) *Allergies*, (5) *Current Medicines*, (6) *Active Problems*, (7) *Past Medical History (PMH)*, (8) *Past Surgical History (PSH)*, (9) *Family History*, (10) *Personal or Social History*, (11) *Gynecologic History*, (12) *Obstetric History*, (13) *Review of Systems*, (14) *Vital Signs*, (15) *Physical Examination (PE)*, (16) *Assessment*, (17) *Tests*, (18) *Plan*, and (19) *Orders*. Figure 1 shows a gynecologic note for a fabricated annual patient visit.

Although such notes capture rich information about a visit, their free-text nature, combined with the demanding clinical settings, often causes the administering clinician to commit errors. For instance, some of the errors in the above note are: (i) the physical examination section is not detailed enough, (ii) it is not specified who in the patient’s family had been diagnosed with hypertension and diabetes, and (iii) 5 out of the 19 sections are missing from the note. While appearing naïve and harmless, such entry-level errors often advance into more serious forms such as medication and prescription errors (J. C. Pham et al., 2012; Wetterneck et al., 2011). From the context of these narrative sectioned visit notes, we classify the entry-level errors into 5 broad categories: (a) *inconsistent information*, wherein the information presented between any two or more sections is contradictory to each other; (b) *incorrect information*, wherein the information is incorrect with respect to the scenario presented in the note, and to the clinical guidelines; (c) *incomplete information*, wherein certain essential information is omitted from the note; (d) *missing section*, wherein an entire required section is missing from the note; (e) *miscellaneous errors*,

such as placement of information into an inappropriate section, or usage of an un-established acronym² in the note. An example of each type of error is highlighted in the Figure 1 and is elaborated in the Table 1.

Table 1
The 5 Kinds of EMR Errors

Kinds of Errors	Error Examples
a. Inconsistent Information	The <i>reason for visit</i> conflicts with the <i>active problem list</i>
b. Incorrect Information	The prescribed dosage for prenatal vitamins is 1 capsule (instead of 2 capsules) daily.
c. Incomplete Information	BMI suggests that the patient is obese, and yet the corresponding diagnosis and plan are omitted from the note
d. Missing Section	Certain sections, like <i>chief complaints</i> , <i>personal history</i> , <i>tests</i> , etc. are missing from the note
e. Miscellaneous	“supervision of normal first pregnancy” is placed in an inappropriate section

It should be noted that all the errors described above can be perceived by carefully reviewing the note, some other kinds of errors such as, the patient forgetting to mention about an active medication, are beyond the scope of this work. In this study, we investigate whether physicians can detect the aforementioned 5 kinds of errors from patient visit notes. In addition, we reveal their strategies, and derive implications for algorithm design.

The Study with Gynecologic Physicians

The objective of this research is to study the manual error control practices among physicians. Through this study, we anticipate to learn from the physicians, and accordingly design computational error detection algorithms. To accomplish this, we conducted a user study with 20 gynecologic physicians, and presented them with several erroneous notes belonging to different hypothetical patients. The participants were asked to carefully audit the notes, and detect the errors. Each session with a participant was followed up with an open-ended debriefing interview. As a result of this study, we accomplished the following goals: (i) assess the participants' ability to detect and resolve errors, (ii) explicate their intuitive strategies, and (iii) infer guidelines for algorithm design.

User Study Design

We recruited 20 gynecologic physicians (11 females, 9 males) working with the Department of Obstetrics and Gynecology in the Drexel University College of Medicine. Each participant had extensive documentation experience with the Allscripts³ EMR deployed into various affiliated clinics. The experience distribution of participants is shown in the Figure 2. Since EMRs were introduced to the facilities in 2008, the participants had at most 4 years of experience until the commencement of this study.

² <http://www.tabers.com/tabersonline/ub>

³ <http://www.allscripts.com>

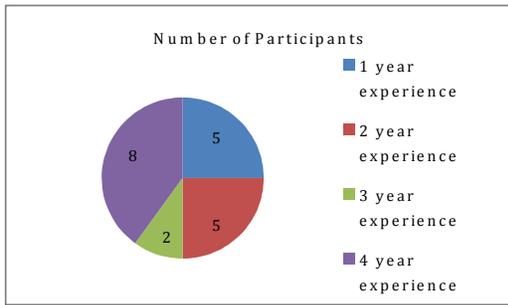


Figure 2. EMR Documentation Experience

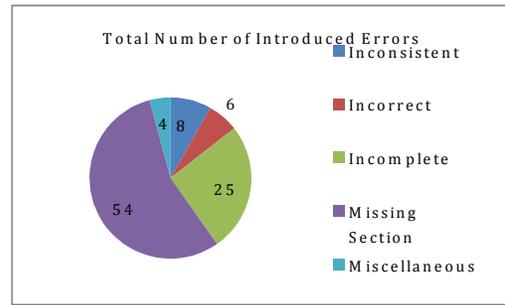


Figure 3. Description of Introduced Errors

For the conduction of this study, we fabricated 7 visit notes belonging to different hypothetical gynecologic patients, and introduced several errors in each note. These erroneous notes were designed after discussion with the clinical investigators of this study, which had more than 20 years of experience with patient visit documentation. Though each introduced error is absolutely fabricated, it is largely inspired by real-world clinical documentation malpractices. The introduced errors comprise our “gold standard” list to be used for evaluation. We introduced a large number (total: 97) of errors in order to increase the odds of errors being detected by the participants, and thereby to increase the odds of learning about the manual strategies. Figure 3 shows the error distribution; as apparent, we introduced a large number of missing section (54) and incomplete information (25) errors because missing information is much more prevalent than other kinds of errors in practice (Botsis, Hartvigsen, Chen, & Weng, 2010; George & Bernstein, 2009; Smith, Banner, Lozano, Olney, & Friedman, 2009).

The study was conducted individually with each participant. During a typical session, the investigator had an in-person meeting with the participant; the participant was briefly introduced to the study, and was demonstrated some examples from the various kinds of errors. Each session was divided into two stages:

- I. *Analysis Stage*: During this stage, the participant was presented with the paper prototypes of the patient notes. For each note, the participant was asked to carefully study the note, detect any data error(s), and document/annotate them on the same sheet of paper. In order not to overwhelm the participants, we did not ask them to categorize the errors.
- II. *De-briefing Stage*: During this stage, the participant was asked of certain follow-up questions regarding the detected errors, e.g., what makes you conclude that certain data are erroneous? what in your medical training allowed you to detect this error? what measures would you take to resolve a certain error? why do you think these errors occur? The responses were transcribed for further analysis, and were synthesized to reveal their personal strategies and thought processes.

Throughout the sessions, we did not impose any time restrictions, and facilitated the participants to perform in an uninhibited manner.

Results and Findings

The Note Analysis Stage. During the analysis stage, each participant spent, on an average, 32.6 minutes to review the 7 visit notes, and detect the embedded errors. Interestingly, each detected error could be mapped to an item from the gold standard list of errors, leading to a perfect *error precision* by each participant. This finding stands in contrast to the existing EMR alert systems that often report false positives. To assess the participants’ detective abilities, we compute the *error recall*, which is the proportion of the errors detected by the participants with respect to the gold standard list of errors. To organize the results, we categorize the 5 kinds of errors into two bins based on the extent of physician liability: (i) *hi-liability errors*: This bin includes the errors that potentially affect the quality of patient care, and hence lead to higher liability to the concerned physician. This includes the *incorrect* and *inconsistent information* errors; (ii) *mod-liability errors*: This bin includes the errors that potentially affect the quality of data, and hence are relatively lower liability catalysts. This includes *missing sections*, *missing information*, and *miscellaneous errors*. It should be noted that the classification of an error kind into hi-liability and mod-liability is not absolute, and would ideally depend on the entire medical history of the patient. The graph in the Figure 4 shows the error recall performance (y-axis) of the participants (x-axis). On an average, the hi-liability and the mod-liability performances are 0.49 and 0.36, respectively. Based on the

within subject t-test, there is a statistically significant ($p \leq 0.05$) difference between the performances of the participants for the two categories of errors.

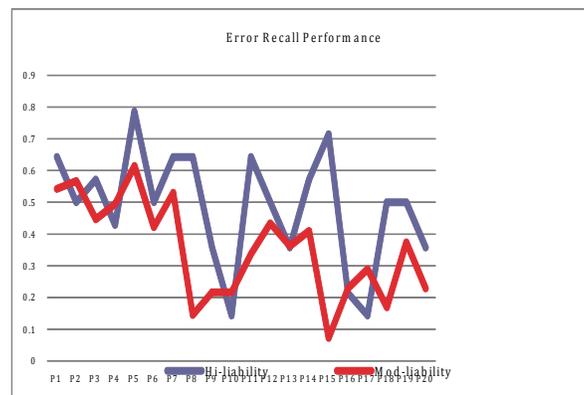


Figure 4. Error Recall Performance

We also measure the association of the task performance with the task duration, and the number of years of experience. The Pearson correlation coefficients are shown in the Table 2; these correlations are not statistically significant at the 0.05 level. Nonetheless, the best error recall performance (70%) was delivered by participant P5, who had 4 years of experience, and who spent 53 minutes auditing the notes. The lowest recall (17%) was achieved by participant P10, who had 1 year of experience, and who only spent 19 minutes analyzing the notes.

Table 2
Correlation: Performance and Time/Experience

	Time spent	Experience
Hi-liability	0.29	0.43
Mod-liability	0.1	0.19

The Debriefing Stage. The debriefing session lasted for an average 13.3 minutes per participant. During this stage, we encouraged the participants to think aloud about the errors and the adopted detection strategies. The results for this stage are also based on our observations during the analysis stage. All the participants were very confident of their performances during the note analysis stage, and were very vocal about their experiences. We discussed and attempted to uncover their abilities and strategies. In particular, we were interested in the provenance of detective abilities, the triggers that caused them to “sense” (Aron, Dutta, Janakiraman, & Pathak, 2011) the errors (See Table 3), and their opinions on the causes of EMR errors.

How do you gain the ability to detect errors? Upon being asked about the provenance of their abilities, 6 participants attributed to the field experience of writing EMR notes in clinical settings; 4 attributed to their training in the medical school; 5 clearly mentioned that while the *mod-liability* errors can be detected by someone fresh out of medical school, the *hi-liability* errors could only be detected by someone who has extensive experience with on-the-spot clinical documentation. The remaining 5 participants attributed the origin of their abilities to both academic training and field experience.

What are the triggers for detecting mod-liability errors? In context of the *missing section errors*, the participants unanimously mentioned that the format of a note is wired into their brains, and hence it is very easy for them to spot any missing sections. For example, the first note element a physician looks at is the *chief complaint* section, and it is easy to identify if this section is missing.

Identifying *incomplete* or *omitted information*, however, requires more attentive analysis of the note. We observed two different kinds of triggers that prompted the participants to detect such errors.

- *Detection of abnormal history events*: Participants could detect the *incomplete information* errors if a partially written abnormal history event caught their attention. For instance, the note in the Figure 1 shows a history of abnormal pap smear in the *gynecologic history* section; however, to completely qualify this event, more information, e.g., the diagnosis date, should be documented. As another example, the same note demonstrates a history of hypertension in the *family history* section, but does not state which family member suffered from hypertension.
- *General Information Recall*: The participants were able to detect omitted information when they could recall some general and mandatory checks associated with annual visits, e.g., the Centers for Disease Control and Prevention⁴ stipulate that HIV screening should be offered on an yearly basis, and if the participant recalled this information, he/she could verify whether HIV screening was present in at least one of the sections.

What are the triggers for detecting hi-liability errors? With regard to the *high-liability errors*, we observed the following triggers that led the participants to detect the *incorrect* and *inconsistent information*.

- a. *Observation of discrepant information between two sections*: Certain errors could be detected as soon as the participants noticed a clear mismatch between the same kinds of information from different sections. For instance, in one of the notes, the strength of the drug “Diflucan” is mentioned as 200mg in the *Plan* section, and the strength of the drug “Fluconazole” is mentioned as 20mg in the *Orders* section. Both the sections refer to the same drug because Diflucan is a market brand name for the ingredient Fluconazole. The strength information in the *Orders* section is incorrect because the tablet form of this drug only comes in 150 or 200mg strength. In our user study, only 5 participants detected and resolved this error, 1 participant identified the error but couldn’t resolve, and the rest either couldn’t figure out the link between the two drugs, or simply missed out due to lack of attention.
- b. *Detection of abnormal results*: Certain errors could be detected as soon as the participants could detect some abnormal, and yet unaccounted for, numerical results. For instance, in one of the notes, the blood pressure of the patient was “150/80” that suggests systolic stage 1 hypertension and diastolic borderline prehypertension, and yet this abnormal result was not alerted in the note. In our user study, 4 participants highlighted this error and believed that the corresponding plan and assessment should be created; and 3 participants detected the error but weren’t sure if it was a typing error, or an omission error by the physician.
- c. *Identification of broken information links*: In a high-quality note, there is a story-like logical flow from across sections, e.g., the *reason for visit* is investigated in the *physical examination* section, and the results of the examination are diagnosed in the *assessment* section, and the appropriate recommendation is documented in the *plan* section. Certain errors were detected once any “broken information link” across multiple sections was observed. For example, in the note in the Figure 5, the results of the wet mount, described in the *physical examination* section, suggest that the patient has yeast infections and bacterial vaginitis. Although this information is acknowledged in the *plan* section, it is omitted from the *assessment* section. The assessment is rather mentioned as trichomoniasis even though there is no corresponding indication in the results. In our user study, 8 participants were confident that trichomoniasis has been accidentally written in lieu of bacterial vaginitis, 5 participants were confused whether the information in *physical exam* is incorrect, or the *assessment* has been entered incorrectly.

⁴ <http://www.cdc.gov>

Physical Exam

Abd: soft obese, point tenderness on abd 3cm superior to R side of incision, 4 cm lateral to midline.
No masses palpated. No CVA tend b/l.

SSE: scant thick white d/c, no internal tenderness on pelvic exam, bladder nontender.

Wet mount : +clue cells, scant hyphae , + whiff

Urine dip: 2+ protein, neg nitrates, tr leuks, moderate bld.

Current Meds

Lithium

Assessment

- RLQ abdominal pain
- Candida albicans vaginitis
- Trichomoniasis

Plan

RLQ Abd Pain? Muscle / nerve entrapment from surgery, RTO for trigger point injection

BV: Tx with flagyl 500mg bid x 7 days

Yeast: Diflucan 200mg x 1 , repeat 1 wk later

Pt given referral for Urology consult 2' frequent UTIs, hematuria.

Orders

Metronidazole 500mg tablet, take 1 tablet 2 times daily after meals; Qty 14, R0, Rx

Fluconazole 20 mg tablet , take 1 tablet 1 time only Qty 1 R1 Rx.

Figure 5. A (portion of) visit note demonstrating *broken information links*

Why do the errors occur? The participants also presented their views on the causes of errors. A majority (17 participants) commented that physicians are primarily responsible for the errors, and that they should improve their documentation behavior. The participants offered the following reasons and suggestions: (i) since most providers write for themselves, they make a lot of assumptions, leading to a poor quality note. On the other hand, the physicians who share patients, such as residents in training, write higher quality notes, (ii) clinicians should ask more questions of the patient to ensure complete information, and (iii) whenever possible, clinicians should write in a list format instead of a narrative format, since lists are more likely to get correctly audited. Only 5 participants attributed the errors to system design: (i) the Allscripts EMR propagates all the active problems through previous visits, creating a lot of inconsistent and obsolete information, (ii) the clinicians tend to write free-text notes because the structured interface is not user-friendly.

Implications for Error Control Algorithm Design

The error recall performance suggests that despite their expertise and experience, participants could not detect more than 50% of the errors. Only 3 participants identified more than 55% of the errors, and the performances for at least 5 participants was below 30%. These results clearly indicate that the existing manual strategies for error control are not foolproof, and it is imperative to replace them with effective computational algorithms. The participants delivered statistically better performance for the hi-liability errors than the mod-liability errors. This further underlines the significance of learning from their expert abilities to minimize potential physician liability. None of the computed correlations between performance and experience/time were significant. So, we cannot draw any clear conclusion regarding learning-based error control algorithms.

Although the participants' quantitative performance is not impressive, the results of the debriefing stage teach us important lessons on algorithm design. The results on the provenance of abilities suggest that the future algorithms should incorporate domain knowledge from a wide range of sources, and also be able to learn and infer from the contextual information in the EMR data. More importantly, the analysis of the triggers for error detection and the associated strategies provided us with some concrete implications on gynecologic error detection algorithms.

Table 3
Error Detection Trigger Categories and the Algorithm Design Implications

Trigger Category	Associated Note Sections	Examples	Algorithm Design Guidelines
<i>General Information Recall</i>	<i>Physical Examination, Plan</i>	If it is an annual visit, then all 19 sections should be present in the note. If it is an annual visit, then HIV screening should be performed and documented in the <i>Physical Examination</i> section. If the patient is over 60 years of age, then a health monitoring plan should be created and specified in the <i>Plan</i> section	Computational Technique: Basic if-then rules, extraction of key information such as age, visit type, etc. Knowledge Sources: Clinical guidelines <ul style="list-style-type: none"> • Centers for Disease Control and Prevention • American Congress of Obstetricians and Gynecologists
<i>Detection of Abnormal Results</i>	<i>Physical Examination, Vital Signs, Review of Systems, Plan</i>	Any any abnormal body mass index should be alerted in the <i>Plan</i> and <i>Assessment</i> sections. Any abnormal blood pressure should be alerted in the <i>Plan</i> and <i>Assessment</i> sections.	Computational Technique: Basic if-then rules, automatic extraction of examination results. Knowledge Sources: <ul style="list-style-type: none"> • Davis's Laboratory and Diagnostic Tests • Agency for Healthcare Policy and Research • Archimedes 360 Medical Calculator
<i>Detection of abnormal history event</i>	<i>History of Present Illness, Current Medicines, Past Medical History, Past Surgical History, Family History, Personal or Social History, Gynecologic History, Obstetric History</i>	If an abnormal pap smear was observed, then the diagnosis date should be noted. For observation of white discharge, the duration and the odor should be specified. If herpes is noted as part of the gynecologic history, then the diagnosis date and the frequency of outbreaks should be specified. For the <i>family history</i> of breast cancer, the relationship of the family member should be specified. For the <i>social history</i> of smoking, the duration and frequency should be specified, and an appropriate counseling should be specified in the <i>Plan</i> section. For any <i>current medications</i> , the dosage, frequency, and administration route should be specified.	Computational Technique: Advanced if-then rules, extraction of abnormal events and their attributes from relevant sections, extraction of medications and their attributes. Knowledge Sources: the 5Ws of information gathering basics, Conceptual model for drugs, disease conditions, and habits. <ul style="list-style-type: none"> • UMLS RxNorm • DailyMed: Current Medication Information • MedlinePlus: Trusted Health Information for You
<i>Observation of discrepant information between two sections</i>	All drug-related sections, <i>Reason for Visit, Active Problem List</i>	The <i>reason for visit</i> should be consistent with the <i>active problem list</i> . The information on the same drugs should match across different sections, e.g., <i>plan</i> and <i>orders</i> Different drugs should be	Computational Technique: Comparison of problems, and medications across sections, Drug and Disease recognition Knowledge Sources: Controlled vocabulary for describing problems and drugs, linkages between drug ingredients, and brand names (Li, Khare, & Lu, 2012), drug-drug interactions.

		<p>compared to verify any possible adverse interaction, e.g., in the Figure 5, Lithium (in <i>Current Medications</i>) and Metronizadole (in <i>Plan</i>), are pharmacologically incompatible. In our user study, only 5 participants detected this error.</p>	<ul style="list-style-type: none"> • DrugBank: Open Data Drug and Drug Target Database • Unified Medical Language System • FDA National Drug Code Directory • Classification of Diseases, Functioning, and Disability • RxDrugs • Davis's Drug Guide
<p><i>Identification of broken links across multiple sections</i></p>	<p><i>Physical Examination, Assessment, Plan, Reason for Visit, Orders</i></p>	<p>Each abnormal result from <i>Physical Examination</i>, should be linked to a corresponding diagnosis in the <i>Assessment</i> section</p> <p>Each diagnosis item should have a corresponding item in the <i>Plan</i> section</p> <p>Each <i>Plan</i> item should optionally have at least corresponding order in the <i>Order</i> section.</p>	<p>Computational Techniques: Extraction of results, diagnosis, plan, order information, linking items from different sections, and discovering the missing links.</p> <p>Knowledge Sources: drug indication (Névéal & Lu, 2010), prescriptions, physical examination resources.</p> <ul style="list-style-type: none"> • SIDER 2: Side Effect Resource • Health Assessment Through the Life Span • Outlines in Clinical Medicine • DailyMed: Current Medication Information • National Drug File – Reference Terminology Source Information

The implications for algorithm design are summarized in the last column of the Table 3. The key step in algorithm design is to be able to programmatically fire the triggers that allowed participants to detect the errors. This requires a combination of computational NLP techniques, and a wide range of medical knowledge resources. For each trigger, some suggested computational techniques, and certain specific authenticated knowledge sources are provided in the table. While a plethora of NLP techniques have been proposed earlier, the use of existing knowledge sources to simulate physicians' knowledge is still limited and under-explored. In the future, we intend to systematically use these findings, and build algorithms for each trigger, while integrating the relevant knowledge sources.

Study Limitations

The biggest limitation of this study is the demanding schedule of our participants. Besides seeing patients, clinicians engage in research, conduct clinical trials, and serve on multiple committees. Thus, the participants could only devote very limited time to this study. Therefore, the set of derived implications are by no means complete. Also, the study has an inherent bias; while the participants knew in advance that the notes contain errors, in reality, such assumptions are not made while reviewing the EMRs. In addition, the frequency of errors (average 13) introduced in each note is not based on any empirical evidence due to the lack of related work. There is a possibility that some participants assumed the notes to contain fewer errors, and terminated their analysis upon finding certain number of errors.

Related Literature

There has been a considerable interest in medical errors in the past; medical errors occur due to a variety of reasons (Wagner & Hogan, 1995) including, un-captured handwritten changes made in the paper chart, electronic data-entry errors, patient-initiated changes in medications, changes made by external clinicians, etc. There has been a substantial literature discussion on the kinds of medical errors. There are two ways of perceiving medical errors: at the inception, and at the conclusion. The latter category includes duplicate order entry, pharmaceutical, adverse events, and medication errors (J. C. Pham et al., 2012; Wetterneck et al., 2011). In this work, we are interested in the errors at their inception, i.e., at the patient record entry level. At this level, several classification schemes are possible. Wagner et al. (Wagner & Hogan, 1995) describe two kinds of clinical errors, *incompleteness*, i.e., missing observations, and *incorrectness*, i.e., inaccuracy in recording information. Aron et al. (Aron et al., 2011)

classify the errors as *procedural errors*, which are not justifiable under any circumstances, and *interpretive errors*, which are qualified based on circumstances and other contextual information. Redwood et al. (Redwood et al., 2011) classify errors based on user intentions as *unintended errors*, e.g., accidentally typing 100 instead of 10, and *unanticipated errors* that occur when a user deliberately deviates from standard clinical guidelines. Botsis et al. (Botsis et al., 2010) classify errors as incomplete, inconsistent, and inaccurate, in the context of clinical narrative text. In this work, we build upon the existing taxonomy, and focus on a specific EMR artifact, the patient visit note. We do not take into account the user intention, and only focus on error as it appears on the document. We classify errors into 5 categories: *inconsistent information*, *incomplete information*, *omitted information*, *missing sections*, and *miscellaneous errors*.

While data entry errors are very prevalent, the existing EMR systems are very limited in catching the errors and alerting the users. The EMRs usually provide error control for structured data. For instance, the Allscripts EMR provides warnings on drug-drug interaction, allergies, and duplicate orders. However, the number of warnings is so high, and their relevance is so low that this often causes confusion and possibilities of more errors (Goldberg, Shubina, Niemierko, & Turchin, 2010). Also, the list of diseases to choose from is too lengthy, e.g., a filtered search on “diabetes” shows a list of 150 options.

There are few existing works that focus on designing algorithms for minimizing and controlling the clinical errors. Wilderman et al. (Wilderman et al., 2011) develop algorithms to minimize errors in clinical trial databases. The algorithm relies on validation rules, warning messages, range checks, and mandatory field checks to minimize errors on the data entry forms. Mitchel et al. (Mitchel et al., 2011) develop error control mechanism for electronic data capture system wherein only critical error-prone fields are targeted, and the validation rules are designed accordingly. As opposed to targeting the research databases as the previous two works, Goldberg et al. (Goldberg et al., 2010) target the EMRs, and design algorithms for detecting errors in the quantitative patient weight information. They develop two versions of algorithm: one that detects errors in real-time, and the other that works in a retrospective manner. To detect the possible outliers, and hence the erroneous entries, they use a combination of statistical techniques such as threshold analysis, change threshold analysis, difference from mean, etc. They evaluate the algorithm on 186 weight entries from real EMRs, and find that the real-time version is 81% accurate, and the retrospective version is 96% accurate as compared to expert judgment.

There are several limitations of the existing algorithms. First, they are designed for structured clinical data, and are hence inapplicable to a large amount of EMR data, which are narrative in nature. Second, they are largely based on the *detection of abnormal results* trigger adopted by physicians. However, the remaining 4 triggers (See Table 3) that we derived from our study are not yet incorporated. This makes the existing algorithms largely incomplete and incompetent for catching different kinds of data errors. Finally, these algorithms do not consider using any established medical knowledge resources, and are hence are less likely to simulate the abilities of the knowledgeable physicians.

In this study, we do not necessarily propose a complete error control algorithm, but we investigate the manual strategies adopted by expert physicians to detect and resolve errors, and learn several lessons on effective algorithm design. We focus on the gynecologic field of medicine, and understand the nature of errors associated with a specific EMR document, the patient visit note. We conduct a user study with the gynecologic physicians who have substantial experience with note documentation. We identify the specific triggers for error detection, the associated computational strategies, and the trustworthy knowledge sources to be incorporated in the future error control algorithms.

Conclusions

In the United States, medical errors kill more people than highway accidents every year (Kohn, Corrigan, & Donaldson, 1999). Although federally funded EMRs have been installed in several major hospital and clinic facilities, there is no evidence of decline in medical errors. In fact, EMRs themselves lead to a new family of errors, the “EMR errors” (Phillips & Gong, 2009; Thyvalikakath et al., 2012). Our ultimate goal is to design sophisticated computational algorithms to alert the physicians about the EMR errors in a real-time fashion.

In this paper, we have taken the first step to algorithm design, and have explored an untapped knowledge resource, i.e., the physicians. We have explored their abilities to detect EMR data errors, and have derived algorithm design implications from their intuitive knowledge and personal strategies. To accomplish this, we have conducted a user study with 20 gynecologic physicians, wherein we presented

them with prototypes of several erroneous EMR patient visit notes, and asked them to detect *incomplete*, *inconsistent*, or *incorrect information* errors. The error recall performance (<50%) of the participants indicated that the existing manual abilities are neither efficient nor guaranteed. However, an in-depth investigation of manual strategies helped us learn several lessons on error control algorithm design. The study helped reveal 5 kinds of triggers that help the physicians sense an error candidate. To identify the triggers, and to detect and resolve the specific errors, physicians rely on an implicit knowledge base of clinical guidelines, and gynecologic best practices. Such a knowledge base has been learned and accumulated through several years of field experience and medical school conditioning.

In comparison to the manual expert strategies, the existing automated algorithms only scratch the surface of error control. In the future, we plan to leverage the findings from this study and design customized algorithms for gynecologic notes. In particular, we would build on the identified triggers, and design algorithms accordingly. To simulate the narrative information extraction, several NLP algorithms exist to extract drug, disease and specific clinical information from texts (Doğan et al., 2010; Li et al., 2012; Névéol & Lu, 2010). To simulate the physicians' knowledge in the head, we intend to utilize, integrate, and organize several available trustworthy knowledge sources hosted by the US Government.

References

- Aron, R., Dutta, S., Janakiraman, R., & Pathak, P. A. (2011). The Impact of Automation of Systems on Medical Errors: Evidence from Field Research. *Information Systems Research*, 22(3), 429-446.
- Botsis, T., Hartvigsen, G., Chen, F., & Weng, C. (2010). Secondary Use of EHR: Data Quality Issues and Informatics Opportunities. *AMIA Summits Transl Sci Proc.*, 1-5.
- Brown, A., & Patterson, D. A. (2001). *To Err is Human*. Paper presented at the First Workshop on Evaluating and Architecting System dependability (EASY '01), Göteborg, Sweden.
- Classen, D. C., Pestotnik, S. L., Evans, R. S., Lloyd, J. F., & Burke, J. P. (1997). Adverse drug events in hospitalized patients: excess length of stay, extra costs, and attributable mortality. *The Journal of American Medical Association*, 277-301.
- Doğan, R. I., Mork, J. G., Bodenreider, O., Demner-Fushman, D., Lang, F.-M., Lu, Z., . . . Aronson, A. R. (2010). Extracting Rx Information from Clinical Narrative. *J Am Med Inform Assoc*, 17(5), 536-539.
- Fichman, R. G., Kohli, R., & Krishnan, R. (2011). The Role of Information Systems in Healthcare: Current Research and Future Trends. *Information Systems Research*, 22(3), 419-428.
- George, J., & Bernstein, P. S. (2009). Using electronic medical records to reduce errors and risks in a prenatal network. *Current opinion in obstetrics & gynecology*, 21(6), 527-531.
- Goldberg, S. I., Shubina, M., Niemierko, A., & Turchin, A. (2010). *A Weighty Problem: Identification, Characteristics and Risk Factors for Errors in EMR Data*. Paper presented at the AMIA 2010 Symposium Proceedings.
- J. C. Pham, J. C., Aswani, M. S., Rosen, M., Lee, H., Huddle, M., Weeks, K., & Pronovost, P. J. (2012). Reducing medical errors and adverse events. *Annu. Rev. Med.*, 63, 447-463.
- Kohn, L., Corrigan, J., & Donaldson, M. (1999). *To Err Is Human: Building a Safer Health System*.
- Li, J., Khare, R., & Lu, Z. (2012). *Improving Online Access to Drug-related Information*. Paper presented at the The 2nd IEEE Conference on Health Informatics, Imaging, and Systems Biology, La Jolla, California.
- Mitchel, J. T., Kim, Y. J., Choi, J., Park, G., Cappi, S., Horn, D., . . . Jr., R. B. D. a. (2011). Evaluation of Data Entry Errors and Data Changes to an Electronic Data Capture Clinical Trial Database. *Drug Information Journal*, 45(4), 421-430.
- Névéol, A., & Lu, Z. (2010). *Automatic integration of drug indications from multiple health resources*. Paper presented at the Proceedings of the 1st ACM International Health Informatics Symposium.
- Phillips, W., & Gong, Y. (2009). *Developing a Nomenclature for EMR Errors*. Paper presented at the Proceedings of the 13th International Conference on Human-Computer Interaction. Part IV: Interacting in Various Application Domains, San Diego, CA.
- Redwood, S., Rajakumar, A., Hodson, J., & Coleman, J. J. (2011). Does the implementation of an electronic prescribing system create unintended medication errors? A study of the sociotechnical context through the analysis of reported medication incidents. *BMC Medical Informatics and Decision Mak*, 11(29).

-
- Smith, L. B., Banner, L., Lozano, D., Olney, C. M., & Friedman, B. (2009). Connected Care: Reducing Errors Through Automated Vital Signs Data Upload. *CIN: Computers, Informatics, Nursing*, 27(5), 318–323.
- Thyvalikakath, T., Dziabiak, M., Johnson, R., Urquidy, M. T.-., Acharya, A., & Jonathan Yabes, T. S. (2012). *Designing Clinical Data Presentation in Electronic Dental Records Using Cognitive Task Analysis Methods*. Paper presented at the The 2nd IEEE Conference on Health Informatics, Imaging, and Systems Biology, La Jolla, California. .
- Wagner, M. M., & Hogan, W. R. (1995). The Accuracy of Medication Data in an Outpatient Electronic Medical Record. *Journal of the American Medical Informatics Association*, 3(3).
- Wetterneck, T. B., Walker, J. M., Blosky, M. A., Cartmill, R. S., Hoonakker, P., Johnson, M. A., . . . Carayon, P. (2011). Factors contributing to an increase in duplicate medication order errors after CPOE implementation. *J Am Med Inform Assoc*, 18, 774-782.
- Wildeman, M. A., Zandbergen, J., Vincent, A., Herdini, C., Middeldorp, J. M., Fles, R., . . . Tan, I. B. (2011). Can an online clinical data management service help in improving data collection and data quality in a developing country setting? *Trials*, 12, 190-196.