Systems for Improving Electronic Health Record Note Comprehension

Balaji Polepalli Ramesh
University of Wisconsin Milwaukee
3200 N Cramer St, Milwaukee, WI
+1-414-229-4677
brpj@uwm.edu

Hong Yu
University of Massachusetts Medical School, VA Central Western Massachusetts
368 Plantation St, Worcester, MA
+1-508-856-3474
hong.yu@umassmed.edu

ABSTRACT
Allowing patients access to their physicians’ notes has the potential to enhance their understanding of disease and improve medication adherence and healthcare outcomes. However, a recent study involving over ten thousand patients showed that allowing patients to read their electronic health record (EHR) notes caused confusion, especially for the vulnerable (e.g., lower literacy, lower income) groups. This finding is not surprising as EHR notes contain medical jargon that may be difficult for patients to comprehend. To improve patients’ EHR note comprehension, we are developing a biomedical natural language processing system called NoteAid (http://clinicalnotesaid.org), which translates medical jargon into consumer-oriented lay language. The current NoteAid implementations link EHR medical terms to their definitions and other related educational material. Our evaluation has shown that all NoteAid implementations improve self-rated EHR note comprehension by 23% to 40% of lay people.

Categories and Subject Descriptors
H.1.2 [User/Machine Systems]: Human information processing

Keywords
Electronic Health Records, Consumer Health, Information Retrieval, Natural Language Processing

1. INTRODUCTION
Allowing patients direct access to their electronic health record (EHR) notes has been shown to enhance medical understanding and improve medication adherence and healthcare outcomes [1]. However, a recent study involving over ten thousand patients showed that allowing patients to read their EHR notes caused confusion, especially for vulnerable (e.g., lower literacy, lower income) groups [1].

The level of a patient’s EHR note comprehension depends on his/her level of health literacy. The Institute of Medicine defined health literacy as “the degree to which individuals have the capacity to obtain, process, and understand basic information and services needed to make appropriate decisions regarding their health”[2]. However, the average American reads at or below an 8th grade level, and over 90 million Americans have limited health literacy [3]. Jones et al. (1992) [4] showed that 50% patients do not understand at least one term in their medical problem list. Lober et al. (2006) [5] found that medical terminology presented a barrier to almost one third of population. EHR notes contain complex medical conditions; abbreviations and other medical jargon that make it hard for patients to comprehend. Example 1, below, shows an excerpt from a de-identified EHR progress note. A patient might find it hard to understand the abbreviations “hx”, “ED”, “SOB” and the medical terms “psoriasis” and “bronchitis”.

Example 1: A patient with hx of active tobacco abuse, bronchitis, and psoriasis presented to ED earlier today with c/o SOB, mild wheezing, chest congestion and chills.

We are therefore developing a biomedical natural language processing (NLP) system called NoteAid that translates medical jargon into lay language. Studies have shown that patient education is effective in improving health literacy, decreasing disease severity, improving self-management behaviors, and reducing hospitalizations [6-8]. We hypothesize that NoteAid will improve patients’ comprehension of their clinical notes and ultimately their healthcare outcome. In our previous evaluation study [9] we reported that NoteAid system improved self-rated EHR note comprehension. In that evaluation, a subject was provided with a clinical note without and with the NoteAid system (in that order) and was asked to report comprehension score of the note. The evaluation design introduced an ordering bias. In this study, we eliminate the ordering bias by randomly assigning EHR note, either with or without the NoteAid system and examine patient EHR comprehension with the systems.

2. RELATED WORK
There is a rich literature related to health literacy and comprehension. A substantial amount of work has been done to compile consumer health vocabulary (CHV) [10], [11]. Elhadad (2006) [12] provided definitions of unfamiliar terms and found that such an approach significantly improved reader’s comprehension of online news stories. Zeng-Treitler et al. (2007) [13] designed and implemented a text translator that identifies difficult terms and replaces them with easier-to-read synonyms. Approaches have been also been developed to predict term familiarity with linguistic/stylistic features [14], term frequency [12], machine learning [15] as well as machine translation [16]. Tools have also been developed to simplify EHR note content using both syntactic and semantic approaches (e.g.,[17], [18]). Smith et al. (2011) [19] improved coherence by manually rewriting a clinical note and found increased comprehension by lay people.

InfoButtons [20] and the Patient Clinical Information System (PatCIS) [21] provided patients with online information resources and educational material [22]. However, the education material was manually compiled by the researchers after reading the EHR notes. In contrast, the NoteAid system [9] automatically extracts
complex medical jargon from EHR notes and links them to patient education material.

3. MATERIALS AND METHODS

3.1 The NoteAid System
As shown in Figure 1, the NoteAid system has three modules: Concept Identifier (CI), Definition Locator (DL) and Definition Filter (DF). CI processes the input text and maps terms to the corresponding UMLS concepts. DL fetches definitions from Medline Plus, Unified Medical Language System (UMLS) or Wikipedia (Wiki) if the term definition is found. We improved the quality of definitions fetched by Wiki by adding a DF, which fetches a definition if article is health-related. Wiki assigns each article a set of categories, which are organized into a direct acyclic graph. We recognize an article as health-related if any of the assigned category or the corresponding hierarchical categories belong to the following two terms: clinical and health.

![Figure 1 – Schematic representation of the NoteAid system](image)

3.2 Evaluation Procedure and Metrics
In this study, we evaluated four NoteAid implementations: MedlinePlus (linking EHR concepts to definitions in Medline Plus), UMLS (linking EHR concepts to their synonyms and definitions in the Unified Medical Language System), Wikipedia (Wiki, linking EHR concepts to health related articles in Wikipedia) and the hybrid system that integrates the three aforementioned implementations.

3.2.1 Subjects
With the IRB approval, we recruited subjects from the Amazon Mechanic Turk (AMT). We used AMT because the subjects have various background and qualifications, and therefore are representative in terms of health literacy. Many research studies use AMT for data collection and survey and have proven to be a reliable resource [23].

3.2.2 Evaluation Data
We randomly selected 20 de-identified progress notes (PGN) from the Pittsburgh NLP repository [24]. We measure the readability of notes with Flesch-Kincaid grade level [25] and self-rated comprehension scores. We also evaluate whether education is correlated with self-rated comprehension scores and report the Pearson coefficient value between the education level (four scales in decreasing order: Master, Bachelor, Associates, and High School).

3.2.4 Evaluation Criteria
We report the average self-rated comprehension scores and used the Mann-Whitney-Wilcoxon test to compare the comprehension score with and without a NoteAid implementation and between different NoteAid implementations.

3.2.5 Demographic Information of Subjects
Twenty-one subjects (9 female and 12 male) completed the evaluation. The number of White American (White), Asian, and Black American (Black) were 15, 4, and 2 respectively. The subjects in the study had a wide range of educational (Edu) backgrounds. Six (28.57%) subjects had a Masters (MA, MS) degree and 6 (28.57%) had Bachelors (BA, BS) degree each, 2 (9.52%) of them had an Associate (Asso) degree and the remaining 7 (33.34%) subjects had a high school diploma (High Sch). Table 1 details the demographic information of subjects as well as their education status.

<table>
<thead>
<tr>
<th>Race</th>
<th>Edu</th>
<th>Notes Alone</th>
<th>Medline Plus</th>
<th>UMLS</th>
<th>Wiki</th>
<th>Hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Asian</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Black</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

4. RESULTS
The 20 PGNs comprise of 473 sentences, 4862 words and have an average Flesch Kincaid Grade Level of 9.8. Table 2 below shows the average comprehension scores of PGNs without any NoteAid implementation and with each of the four NoteAid implementations. The average comprehension score of subjects and Flesch-Kincaid grade level had a spearman ranked correlation coefficient of rho = -0.77 (p<0.01).

As shown in Table 2, all NoteAid implementations improved self-rated PGN note comprehension and the improvements were statistically significant (p<0.01, the Mann-Whitney-Wilcoxon test). The difference in comprehension scores between different NoteAid implementation was not statistically significant except for the difference between the MedlinePlus and the UMLS implementations (p<0.01, the Mann-Whitney-Wilcoxon test). Table 2 also shows the number of concepts identified by each of the NoteAid implementations.
Table 2 - The average self-rated comprehension values (average ± std dev) and number of concepts identified by NoteAid implementations. (*p<0.01)

<table>
<thead>
<tr>
<th>System</th>
<th>Notes Alone</th>
<th>Medline Plus</th>
<th>UMLS</th>
<th>Wiki</th>
<th>Hybrid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>2.95 ±0.67</td>
<td>4.12* ±0.33</td>
<td>3.63* ±0.57</td>
<td>3.85* ±0.47</td>
<td>3.92* ±0.40</td>
</tr>
<tr>
<td># conc</td>
<td>NA</td>
<td>52</td>
<td>352</td>
<td>436</td>
<td>476</td>
</tr>
</tbody>
</table>

Figure 2 shows the average self-rated comprehension scores of all NoteAid implementations for every PGN note, and Figure 3 shows a scatter plot of the average self-rated comprehension scores with notes alone and with the MedlinePlus implementation. The results as shown in both figures demonstrate a strong and consistent improvement of self-rated comprehension scores with NoteAid implementations for every PGN note.

The Pearson coefficient values between the subject education level and comprehension scores are: Note Alone: 0.98, MedlinePlus: 0.31, UMLS: 0.71, Wiki: -0.47 and Hybrid: 0.04.

5. DISCUSSION AND FUTURE WORK

Our results show that, when the clinical notes are presented alone, the self-rated comprehension scores are highly correlated (0.98 Pearson coefficient) with the education levels of the subjects. The results support the validity of self-rated comprehension scores. In contrast, the correlation results are mixed with different NoteAid implementations. While the UMLS has a correlation value of 0.71, the MedlinePlus and Hybrid implementations decrease to 0.31 and 0.04. The Wiki implementation has a negative correlation: -0.47. Several factors may contribute to the results. First, the definition quality of the UMLS, Wiki, MedlinePlus and Hybrid resources are not yet evaluated and it is unclear whether the definitions correctly represent the semantic meanings of the notes. Secondly, although providing definitions may help comprehension, providing too much or unnecessary information (such as Wiki) may hurt those who have a better education level.

In future work, we need to conduct a comprehensive “think aloud” evaluation study to understand the behavior of users. We will also need to evaluate the quality of definitions of different NoteAid implementations and patient comprehension by replacing complex medical jargon with its equivalent lexical lay term variants [26,27] in EHNs.

The significant improvement of MedlinePlus over the UMLS implementation may be attributed to the lower readability of content in UMLS. For example, the definition of “malnutrition” is complex in UMLS and has a Flesch-Kincaid grade level of 19. Whereas the definition of “malnutrition” in Medline Plus has a Flesch-Kincaid grade level of 11. Similarly, Wiki has a grade level of 13 for “malnutrition.” Although the improvement in comprehension of Wiki over MedlinePlus implementation was not statistically significant, Wiki content may not be accurate as discussed earlier.

Our results show that all four NoteAid implementations improved EHR note self-rated comprehension significantly over Notes alone. The results are largely consistent with our previous evaluation [9] in which NoteAid implementations were evaluated in a before-and-after fashion but there are differences between the two evaluation results. In our previous study, we found that the Wikipedia implementation had the largest improvement and that the MedlinePlus implementation decreased the self-rated comprehension scores. Such discrepancy can be explained by the limitations of our study.

First, we report subjects’ self-rated note comprehension but did not evaluate to what extent they accurately comprehended the note content. The before-and-after evaluation design [9] may be a better model as we force a subject to read the EHR note prior to her/his exposure to the improved note (note+NoteAid). Our before-and-after evaluation results are also consistent with the number of concepts recognized by each NoteAid implementation. The MedlinePlus implementation has the least number of concept recognition, and therefore its comprehension improvement may be small. A randomized design, as we have done in this study, may provide an evaluation subject little incentive for comprehending the note content. In the future work, we will test subjects’ comprehension based on content analyses of every clinical note. Furthermore, we will evaluate subjects’ health literacy [28].

Secondly, the number of subjects is small in this study. As a result, we can’t evaluate the impact of moderators. For example, the data size is not well rounded to conclude that the subjects’ education levels impact self-rated comprehension scores. Other limitations of the study include that lay people performed our evaluation but not patients who comprehend their own EHR notes.

6. CONCLUSION

Our evaluation results show that NoteAid improves EHR note self-rated comprehension of lay people in a randomized evaluation study and the MedlinePlus implementation demonstrated the highest improvement.

7. ACKNOWLEDGEMENTS

We thank Tamsin Maxwell for providing helpful comments. Research reported in this publication was supported in part by 1R01GM095476 to Hong Yu, by a start-up fund from University of Massachusetts Medical School to Hong Yu, and by the National Center for Advancing Translational Sciences of the National Institutes of Health under award number UL1TR000161. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.
8. REFERENCES


