MEDINFO 2013
C.U. Lehmann et al. (Eds.)

doi:10.3233/978-1-61499-289-9-714

Improving Patients' Electronic Health Record Comprehension with NoteAid

Balaji Polepalli Ramesh[a], Thomas Houston[b,c], Cynthia Brandt[d,e], Hua Fang[b], Hong Yu[b,f]

[a]*Biomedical and Health Informatics, University of Wisconsin Milwaukee, Milwaukee, WI, USA*
[b]*Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, USA*
[c]*Bedford VA Medical Center, Bedford, MA, USA*
[d]*Center for Medical Informatics, Yale University, New Haven, CT, USA*
[e]*VA Connecticut Health Care System, West Haven, CT, USA*
[f]*VA Central Western Massachusetts, Northampton, MA, USA*

Abstract

Allowing patients direct access to their electronic health record (EHR) notes has been shown to enhance medical understanding and may improve healthcare management and outcome. However, EHR notes contain medical terms, shortened forms, complex disease and medication names, and other domain specific jargon that make them difficult for patients to fathom. In this paper, we present a BioNLP system, NoteAid, that automatically recognizes medical concepts and links these concepts with consumer oriented, simplified definitions from external resources. We conducted a pilot evaluation for linking EHR notes—through NoteAid—to three external knowledge resources: MedlinePlus, the Unified Medical Language System (UMLS), and Wikipedia. Our results show that Wikipedia significantly improves EHR note readability. Preliminary analyses show that MedlinePlus and the UMLS need to improve both content readability and content coverage for consumer health information. A demonstration version of fully functional NoteAid is available at http://clinicalnotesaid.org

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

Introduction

Allowing patients direct access to their electronic health record (EHR) notes has been shown to enhance medical understanding and may improve healthcare management and outcomes [1]. However, over 90 million Americans have limited health literacy [2]. Patients who have limited health literacy may have difficulty understanding written information in their medical notes and reports, communicating health related problems with their healthcare providers, and navigating complex Electronic Health Record (EHR) systems.

The level of a patient's EHR note comprehension is related to his/her level of health literacy, which is defined by the Institute of Medicine 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" [3]. EHR notes contain medical terms, abbreviations, concepts, and domain-specific jargon that are difficult to comprehend.

We are therefore developing NoteAid, a biomedical natural language processing (BioNLP) system to improve patient comprehension of EHR notes by providing comprehensible terms and concepts tailored to the patient NoteAid also links EHR notes to external patient education materials. Studies have shown that patient education can improve health knowledge, and education-based behavioral intervention can help improve self-management behaviors and reduce hospitalizations [4–6]. We therefore speculate that NoteAid will increase patient comprehension of their EHR notes and therefore improve the quality of patient care.

Related Work

Research related to health literacy and comprehension is rich. Studies have shown that consumers use a different vocabulary than clinicians when searching for health information [7]. Mapping between these vocabularies is a necessary step in building effective communication between clinicians and patients. A substantial amount of work has been done to compile a consumer health vocabulary [8,9] by analyzing user queries to Web sites at the National Library of Medicine [7,10]; consumer texts [11,12]; social media, including email content [13], and online support groups (e.g., PatientsLikeMe [14]). Approaches have been developed to predict term familiarity with linguistic/stylistic features [15], term frequency [16], as well as machine learning approaches [17]. Tools have also been developed to simplify EHR note content using both syntactic and semantic approaches (e.g., [18,19]). Approaches developed to predict unfamiliar terms found that providing definitions of unfamiliar terms significantly improved the comprehension of a collection of online news stories [16].

The Patient Clinical Information System (PatCIS) [20] was created to serve as a test bed for exploring issues related to patient access of EHR records. It provides patients with online information resources and educational material, and evaluations by patients have been positive [21]. However, the education material in the PatCIS system was mainly compiled by the researchers manually after reading the EHRs. In contrast, we are developing NoteAid, a fully implemented system that automatically links EHR notes to patient education materials to assist their EHR note comprehension.

Materials and Methods

NoteAid has two main components: A knowledge resource comprised of patient education materials and BioNLP ap-

proaches that link EHR notes to the knowledge resource. In the following, we first describe three knowledge resources. We then describe BioNLP approaches and conclude with an evaluation design.

External Knowledge Resources

The Unified Medical Language System (UMLS) [22] is a rich biomedical knowledge resource; Metathesaurus (MT) is a large, multi-purpose, and multi-lingual thesaurus that contains millions of biomedical and health related concepts, their synonym names, and their relations, from over 150 vocabularies. UMLS makes available the lexical tool MetaMap [23], which maps text to UMLS concepts and semantic types. We use UMLS MT version 2011AB in our system.

MedlinePlus [24] provides current and reliable information about over 900 diseases, conditions and treatment to users in simple language. The links to various health topics are added daily and the content is reviewed once every six months.

Wikipedia (Wiki) is a collaborative, community developed web-based encyclopedia that has evolved to be an important medical resource for a wide spectrum of audiences including healthcare professionals [25]. Among online health information resources, Wiki has shown to be a prominent source, ranking among the first ten results in 71-85% of search engines and keywords tested [26].

The NoteAid System

Our goal was to assist patients to understand the content of their EHR notes. For this purpose, we decided to link the complex medical concepts that appear in the text to simple consumer oriented definitions and explanations from external sources of information as described earlier. These definitions describe the complex medical concepts and jargon that appear in these EHR notes.

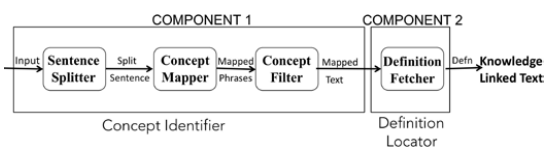

Figure 1- Schematic representation of the NoteAid system

Figure 1, above, contains the schematic representation of the NoteAid system. The system is comprised of two components. The first component is Concept Identifier (CI). CI processes input text and maps terms to the corresponding UMLS concepts. The second component is Definition Locator (DL). DL fetches definitions from UMLS, MedlinePlus and Wikipedia.

CI consists of three modules: Sentence Splitter, Concept Mapper, and Concept Filter. Sentence Splitter splits input text into individual sentences. Concept Mapper is built upon the Metamap tool [23] which identifies concepts and their UMLS semantic types. Concept Filter identifies clinical concepts by selecting ones with the following UMLS semantic types that appear in the sentence: Acquired Abnormality, Antibiotic, Cell or molecular Dysfunction, Clinical Attribute, Diagnostic Procedure, Disease or syndrome, Experimental model of disease, Finding, Laboratory procedure, Laboratory or Test result, Organ or Tissue function, Pathologic function, Physiologic function, Pharmacologic substance, Sign or symptom and Therapeutic or preventive procedure.

After concepts are identified, DL retrieves definitions from UMLS, MedlinePlus and Wikipedia using Definition fetcher module. The UMLS MRDEF file contains definitions of 107,604 unique concepts. We parsed the MedlinePlus content and extracted over 900 health related topics and their summaries. We automatically extracted definitions from the summaries by using handcrafted rules. For Wiki, we made use of the web service WikiAPI to return a Wiki page given a query topic (concept). When a page is returned, DL extracts the first three lines of the Wikipedia content. We found such a simple method works very well for extracting definitions from Wikipedia.

User Interface Design

We designed and implemented a user-interface with Java servlets and the tooltip with Javascript. Figure 2 below shows the screenshot of the interface.

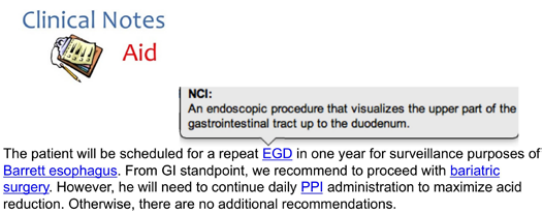

Figure 2- A screen shot of the NoteAid system. NoteAid takes a clinical narrative and outputs the narrative in which the clinical terms are underlined. The tooltip displays the definition(s) of a concept from the knowledge resources when a user hovers the cursor over a highlighted concept. In the current example, the definition of "EGD" is from the NCI (National Cancer Institute), a source vocabulary from UMLS.

Evaluation

To evaluate whether NoteAid improves EHR note comprehension, we evaluated four NoteAid implementations, namely: MedlinePlus, UMLS, Wikipedia, and a Hybrid (combined knowledge resource), using de-identified EHR notes.

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 as it has proven to be a reliable resource [27].

Evaluation Data and Readability Score

We randomly selected 20 de-identified progress note reports (PGN) and 20 de-identified discharge summary reports (DS) from the Pittsburgh NLP repository [28], which contains a variety of de-identified clinical reports including discharge summaries and progress notes. We used both Flesch-Kincaid ease score and Flesch-Kincaid grade level [29] to score readability; the higher the Flesch Readability ease scores, the higher the readability. In contrast, a lower Flesch-Kincaid grade level indicates higher readability.

Evaluation Process

For each NoteAid implementation, we asked each subject to read each assigned EHR note before and after the NoteAid system and score his/her level of comprehension (on a scale of 1—5, with 1 the poorest and 5 the best comprehension). Each subject was asked to complete the evaluation of either 20 PGNs or 20 DSs. Each subject was given a link to a welcome page describing the study, followed by demographic infor-

mation page, qualifying question page, pages containing EHR notes to evaluate, and finally the thank you page along with the validation code. For quality control, we gave each subject a question related to his/her evaluation data. The evaluation was hosted and stored on a local server. At the end of the evaluation, subjects received a code to confirm their participation in the study and receive payment for the task. Each subject spent 30-40 minutes to complete the entire evaluation and s/he was paid $4. We recruited 64 subjects: 8 subjects for each of the 8 evaluation tasks (4 systems, 2 types of EHR notes). A total of 3 subjects did not complete the evaluations and 2 subjects withdrew from the study. Our results were based on the analyses of the evaluation of the remaining 59 subjects who completed their tasks.

Demographic Information of Subjects

Of the 59 subjects (23 female and 36 male) completed the evaluation. The number of Asian, White, African American and Alaskan Native was 34, 23, 1, and 1, respectively. Nearly 24% of all subjects reported having Hispanic or Latin ethnicity. The subjects of the study had a wide range of educational backgrounds. Twenty-three (39%) of them had Bachelors degree, 15 (25.4%) of them had a Masters degree, 12 (20.3%) of them had an Associate degree and the remaining 9 (15.3%) had a high school diploma.

Evaluation Criteria

We report the average comprehension scores before and after each of the NoteAid implementations: Medline Plus, UMLS, Wiki, and Hybrid. The non-parametric Wilcoxon signed-rank test was used to compare subjects' scores on PGNs or DSs before and after each implementation. Unlike t-tests, the Wilcoxon signed-rank test does not assume data are normally distributed.

In order to evaluate whether the comprehension scores represent readability, we report both Flesch-Kincaid ease score and Flesch-Kincaid grade level and calculate the non-parametric Spearman correlation coefficient. We also show the scatterplot of the comprehension scores before and after the NoteAid systems, between the two readability scores, and between the comprehension and the readability scores.

Results

Table 1 shows the characteristics of the EHR note data used in the evaluation. The DS and PGN have an average Flesch Readability ease score of 38.5 and 43.9 and an average Flesch-Kincaid Grade Level of 8.8 and 9.76, respectively.

Table 1 – Statistics of the Evaluation Data

Type	Discharge Summaries	Progress Notes
No. of Reports	20	20
Total (Avg) # of sentences	355 (17.8)	473 (23.7)
Total (Avg) # of Words	2362 (118)	4862 (243)
Avg Flesch ease score	38.5	43.9
Avg Flesch-Kincaid Grade Level	8.8	9.8

Table 2 shows the average comprehension scores of the four NoteAid implementations (before and after each implementation). As shown in the table, all three NoteAid implementations except for MedlinePlus improve the comprehension in both DSs and PGNs. None of the improvement is statistically significant except for the Wiki implementation on PGNs. The Hybrid implementation has a p value of 0.06 for improvement on PGNs.

Table 2 – Average ± standard deviation of comprehension values of four NoteAid implementations

System	Discharge Summaries		Progress Notes	
	Before	After	Before	After
MedlinePlus	3.52±0.73	3.49±0.87	3.18±0.38	2.86±0.55
UMLS	3.80±0.16	3.81±0.48	3.75±0.55	4.01±0.86
Wiki	3.57±0.68	4.14±0.49	3.45±0.55	4.53±0.71*
Hybrid	3.86±0.69	4.02±0.73	3.40±0.55	4.54±0.53

*: $p < 0.05$

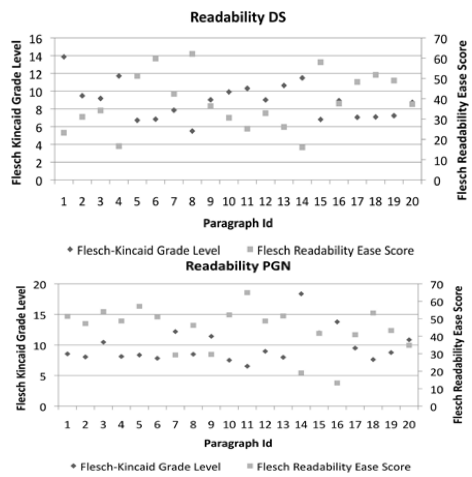

Figure 3 - Readability of Evaluation data

Figure 3 shows the scatter plot of the Flesch-Kincaid Grade Level and the Flesch Readability ease score calculated from the 20 DSs and 20 PGNs we used for the evaluation. The Spearman rank correlation on Flesch-Kincaid Grade Level and the Flesch Readability ease score demonstrated the consistency of data. (for PGN: rho = -0.807 $p < 0.0001$, for DS: rho =-0.970, $p < .0001$).

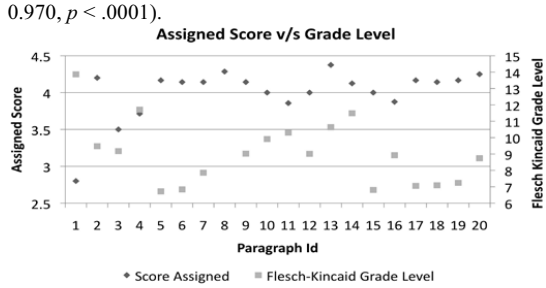

Figure 4 – Scatter Plot of the assigned score and Flesch-Kincaid Grade Level in the evaluation EHR notes

Figure 4 shows the scatter plot of the Flesch-Kincaid Grade Level and text comprehension score after NoteAid system on DS reports. Table 3 shows the total number of concepts that were recognized by three different NoteAid implementations on the 20 DSs and 20 PGNs

Table 3 – Number of concepts that were linked to different knowledge resources by the NoteAid system

System	Discharge Summaries	Progress Notes
MedlinePlus	37	53
UMLS	171	362
Wiki	190	427

Discussion

According to the average Flesch-Kincaid Grade Level shown in Table 1, DSs are easier to comprehend than the PGNs, corresponding to a 8th and 9th grade education, respectively. Our results show that subjects' self-reported EHR note comprehension scores fall between 3 and 4 on a five-point scale. In contrast, all 59 of our subjects have a high school education and higher. The results suggest a gap between education level, readability and health literacy. The observation of such a literacy gap is consistent with other evaluation studies in health literacy [30]. As shown in Figure 4, our results show that text readability scores positively correlate with the comprehension scores, suggesting that our subjects' assignment of self-comprehension scoring is consistent with the readability assessment. Our results show that overall, the NoteAid systems improve comprehension. Of all four systems, the Wiki implementation on PGNs has the highest performance and statistical significance in improving EHN comprehension. In contrast, the consumer-driven authoritative resources of the UMLS and the MedlinePlus implementations yield relatively less improvement. The non-significant improvement in the comprehension of DS could be due to the fact that DSs are easier to comprehend than the PGNs. The self-comprehension scores are higher in DSs than in PGNs and therefore the difference in improvement is smaller. Content coverage may partially explain performance differences among the three external resources. As shown in Table 3, EHR notes link to more Wikipedia definitions than to UMLS. MedlinePlus has the least number of definitions available. While Wikipedia incorporates over 4 million topics and articles written in English, the content of MedlinePlus and UMLS are limited. For example, we found only 900 health topics in MedlinePlus. As a result, the NoteAid system that links EHR notes to Wikipedia yields the best performance.An illustrative example is shown in the following EHR note:

Example 1: "Her cardiac index is 3.6. She is off of drips. We will start on baseline Coreg. history of diabetes on 80 of Lantus a day. Would try to wean her off of the insulin infusion to a low level of Lantus with a sliding scale. No evidence of bleeding. Keep the chest tubes in place. We have started her Synthroid. From a respiratory standpoint, continue incentive spirometry, mobilization, and oral narcotics."

In this EHR note, Wikipedia covers 6 concepts—"cardiac index," "Coreg," "diabetes," "lantus," "bleeding," and "synthroid" and received an average comprehension score of 4.3. In contrast, the UMLS covers three concepts—"bleeding," "Synthroid," and "oral narcotics" and received an average comprehension score of 4. MedlinePlus covers only two concepts "diabetes" and "bleeding" and received the lowest average comprehension score of 2.5.

Furthermore, we found that the Wikipedia content is easier to read than the UMLS or the MedlinePlus content. An example is shown below.

Example 2: "The patient's bilirubin is 1.6. He is not coagulopathic."

The definition of "coagulopathic" is complex in the UMLS: "Hemorrhagic and thrombotic disorders that occur as a consequence of abnormalities in blood coagulation due to a variety of factors such as COAGULATION PROTEIN DISORDERS; BLOOD PLATELET DISORDERS; BLOOD PROTEIN DISORDERS or nutritional conditions" which has a Flesch-Kincaid grade level of 24. In contrast, its Wikipedia definition—"Coagulopathy is a condition in which the blood's ability to clot is impaired. This condition can cause prolonged or excessive bleeding, which may occur spontaneously or following an injury or medical and dental procedures. The normal clotting process depends on the interplay of various proteins in the blood," —has a Flesch-Kincaid grade level of 13 and is easier to comprehend than the UMLS definition. The evaluation results show that the NoteAid system that integrates all three resources did not perform as well as the Wikipedia system, although the integration outperformed both the UMLS and the MedlinePlus systems. This may be explained by the fact that the addition of less readable content from UMLS and MedlinePlus hurts performance.

Conclusion

NoteAid improved EHR note comprehension and linking EHR notes to Wikipedia was the best performing NoteAid system. Although MedlinePlus and UMLS are designed to facilitate consumer-oriented health information, they both need to improve their content coverage as well as readability.

Limitations: First, our NoteAid implementations link EHR notes to definitions only, not to other education materials that MedlinePlus additionally provides. Secondly, lay people performed our evaluation, not the patients who comprehend their own EHR notes. We evaluated only before and after the NoteAid system, so order-effect bias may have been introduced. Finally, we scored subjects' EHR note comprehension but did not evaluate to what extent they accurately comprehended the EHR note content.

Future Work: We plan to access and improve the effectiveness of the concept filtering and coverage to improve the performance of the system. In addition, we hope to evaluate the quality of the definitions provided by various educational resources and evaluate the system in a real health care setting — the next step towards building a clinical application.

Acknowledgments

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 represent the official views of the NIH.

References

[1] Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, Feldman HJ, Mejilla R, Ngo L, Ralston JD, Ross SE, Trivedi N, Vodicka E, Leveille SG. Inviting Patients to Read Their Doctors' Notes: A Quasi-experimental Study and a Look Ahead. Ann Intern Med. 2012 Oct 2;157(7):461–70.

[2] The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy [Internet]. 2006 [cited 2012 Oct 15]. Available from: http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006483

[3] Literacy C on H. Health Literacy: A Prescription to End Confusion. 1st ed. Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. National Academies Press; 2004.

[4] Rothman RL, DeWalt DA, Malone R, Bryant B, Shintani A, Crigler B, Weinberger M, Pignone M. Influence of patient literacy on the effectiveness of a primary care-based diabetes disease management program. JAMA. 2004 Oct 13;292(14):1711–6.

[5] DeWalt DA, Malone RM, Bryant ME, Kosnar MC, Corr KE, Rothman RL, Sueta CA, Pignone MP. A heart failure self-management program for patients of all literacy levels: a randomized, controlled trial [ISRCTN11535170]. BMC Health Serv Res. 2006;6:30.

[6] Schillinger D, Handley M, Wang F, Hammer H. Effects of self-management support on structure, process, and outcomes among vulnerable patients with diabetes: a three-arm practical clinical trial. Diabetes Care. 2009 Apr;32(4):559–66.

[7] McCray AT, Loane RF, Browne AC, Bangalore AK. Terminology issues in user access to Web-based medical information. Proc AMIA Symp. 1999;107–11.

[8] Smith B, Fellbaum C. Medical WordNet: a new methodology for the construction and validation of information resources for consumer health. Proceedings of the 20th international conference on Computational Linguistics [Internet]. 2004 [cited 2012 Nov 20]. page 371. Available from: http://dl.acm.org/citation.cfm?id=1220409

[9] Consumer Health Vocabulary http://www.consumerhealthvocab.org/.

[10] Zeng QT, Tse T, Crowell J, Divita G, Roth L, Browne AC. Identifying Consumer-Friendly Display (CFD) Names for Health Concepts. AMIA Annu Symp Proc. 2005;2005:859–63.

[11] Patrick TB, Monga HK, Sievert MC, Hall JH, Longo DR. Evaluation of Controlled Vocabulary Resources for Development of a Consumer Entry Vocabulary for Diabetes. Journal of Medical Internet Research. 2001 Aug 28;3(3):e24.

[12] Slaughter LA, Soergel D, Rindflesch TC. Semantic representation of consumer questions and physician answers. Int J Med Inform. 2006 Jul;75(7):513–29.

[13] Smith CA, Stavri PZ, Chapman WW. In their own words? A terminological analysis of e-mail to a cancer information service. Proc AMIA Symp. 2002;697–701.

[14] Smith CA, Wicks PJ. PatientsLikeMe: Consumer health vocabulary as a folksonomy. AMIA Annu Symp Proc. 2008;682–6.

[15] Rosemblat G, Logan R, Tse T, Graham L. Text Features and Readability: Expert Evaluation of Consumer Health Text. MEDNET.

[16] Elhadad N. Comprehending technical texts: predicting and defining unfamiliar terms. AMIA Annu Symp Proc. 2006;239–43.

[17] Zeng Q, Kim E, Crowell J, Tse T. A text corpora-based estimation of the familiarity of health terminology. Biological and Medical Data Analysis. 2005;184–92.

[18] Kandula S, Curtis D, Zeng-Treitler Q. A semantic and syntactic text simplification tool for health content. AMIA Annual Symposium Proceedings [Internet]. 2010 [cited 2012 Dec 2]. page 366. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041424/

[19] Leroy G, Endicott J, Mouradi O, Kauchak D, Just M. Improving perceived and actual text difficulty for health information consumers using semi-automated methods. AMIA Annu Symp Proc. 2012; pp. 522–31.

[20] Cimino JJ, Patel VL, Kushniruk AW. The patient clinical information system (PatCIS): technical solutions for and experience with giving patients access to their electronic medical records. Int J Med Inform. 2002 Dec 18;68(1-3):113–27.

[21] Cimino JJ, Patel VL, Kushniruk AW. What do patients do with access to their medical records? Stud Health Technol Inform. 2001;84(Pt 2):1440–4.

[22] Humphrey B, Lindberg DAB, Schoolman HM, Barnett GO. The Unified Medical Language System: An Informatics Research Collaboration. Journal of the American Medical Association. 1998;5:1–11.

[23] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp. 2001;17–21.

[24] (U.S.) NL of M. Fact SheetMedlinePlus® [Internet]. [cited 2012 Nov 19]. Available from: http://www.nlm.nih.gov/pubs/factsheets/medlineplus.html

[25] Heilman JM, Kemmann E, Bonert M, Chatterjee A, Ragar B, Beards GM, et al. Wikipedia: A Key Tool for Global Public Health Promotion. Journal of Medical Internet Research. 2011 Jan 31;13(1):e14.

[26] Laurent M, Vickers TJ. Seeking Health Information Online: Does Wikipedia Matter? J Am Med Inform Assoc. 2009 Jul 1;16(4):471–9.

[27] Proulx J, Kandula S, Hill B, Zeng-Treitler Q. Creating Consumer Friendly Health Content: Implementing and Testing a Readability Diagnosis and Enhancement Tool. 2012.

[28] Chapman W. University of Pittsburgh NLP Repository (http://www.dbmi.pitt.edu/nlpfront).

[29] Si L, Callan J. A statistical model for scientific readability. Proceedings of the tenth international conference on Information and knowledge management [Internet]. 2001 [cited 2012 Dec 11]. page 574–6. Available from: http://dl.acm.org/citation.cfm?id=502695

[30] Schapira MM, Fletcher KE, Hayes A, Eastwood D, Patterson L, Ertl K, Whittle J. The development and validation of the hypertension evaluation of lifestyle and management knowledge scale. J Clin Hypertens (Greenwich). 2012 Jul;14(7):461–6.

Address for correspondence

Dr. Hong Yu, University of Massachusetts Medical School
AS6-2071 ASC-QHS, 368 Plantation St., Worcester, MA 01655
Email: hong.yu@umassmed.edu