Assessing the readability of ClinicalTrials.gov

Danny TY Wu[1], David A Hanauer[1,2], Qiaozhu Mei[1,3], Patricia M Clark[4,5], Lawrence C An[5,6], Joshua Proulx[7], Qing T Zeng[7], VG Vinod Vydiswaran[1], Kevyn Collins-Thompson[1,3], Kai Zheng[1,8]

RESEARCH AND APPLICATIONS

ABSTRACT

Objective ClinicalTrials.gov serves critical functions of disseminating trial information to the public and helping the trials recruit participants. This study assessed the readability of trial descriptions at ClinicalTrials.gov using multiple quantitative measures.

Materials and Methods The analysis included all 165 988 trials registered at ClinicalTrials.gov as of April 30, 2014. To obtain benchmarks, the authors also analyzed 2 other medical corpora: (1) all 955 Health Topics articles from MedlinePlus and (2) a random sample of 100 000 clinician notes retrieved from an electronic health records system intended for conveying internal communication among medical professionals. The authors characterized each of the corpora using 4 surface metrics, and then applied 5 different scoring algorithms to assess their readability. The authors hypothesized that clinician notes would be most difficult to read, followed by trial descriptions and MedlinePlus Health Topics articles.

Results Trial descriptions have the longest average sentence length (26.1 words) across all corpora; 65% of their words used are not covered by a basic medical English dictionary. In comparison, average sentence length of MedlinePlus Health Topics articles is 61% shorter, vocabulary size is 95% smaller, and dictionary coverage is 46% higher. All 5 scoring algorithms consistently rated ClinicalTrials.gov trial descriptions the most difficult corpus to read, even harder than clinician notes. On average, it requires 18 years of education to properly understand these trial descriptions according to the results generated by the readability assessment algorithms.

Discussion and Conclusion Trial descriptions at ClinicalTrials.gov are extremely difficult to read. Significant work is warranted to improve their readability in order to achieve ClinicalTrials.gov's goal of facilitating information dissemination and subject recruitment.

....

Keywords: readability, comprehension, clinical trial, ClinicalTrials.gov, electronic health records, natural language processing

INTRODUCTION

Background and Significance

Clinical trials are the bedrock of research for a variety of medical interventions including drugs, devices, and therapies intended to improve treatment efficacy and patient outcomes. Today, many clinical trials must be registered in a US National Institutes of Health repository (http://ClinicalTrials.gov) as a means to improve their accessibility to the public and to enhance participant recruitment.[1,2] Although ClinicalTrials.gov does not include all trials ever conducted, recent regulatory requirements have led to the exponential growth of the number of studies registered, with a 10-fold increase occurring over the past decade.[3]

Each study registered at ClinicalTrials.gov is accompanied with a detailed description covering all aspects of the trial protocol including the disease(s) targeted, intervention under evaluation, and requirements for participant recruitment. The registry hence serves not only as a mechanism for ensuring the ethics and integrity of the trials through increased transparency, but also a credible source of information for patients who are interested in participating or in learning about the results of the trials. As described in the ClinicalTrials.gov mission statement, the website was established in part to fulfill the goal of "providing patients, their family members, health care professionals, researchers, and the public with easy access to information on publicly and privately supported clinical studies on a wide range of diseases and conditions."[4] Additionally, the US Congress Food and Drug Modernization Act, which led to the creation of ClinicalTrials.gov, requires that the details about all clinical trials registered must be "in a form that could be readily understood by the public."[5]

However, the registry's potential for facilitating information dissemination and participant recruitment could be limited if the public, with varying literacy levels, are unable to read and properly understand the descriptions of the trials. Poor readability can also be a source of self-selection bias undermining the broad applicability of study findings, as those who are able to better comprehend the trial protocols may be more likely to volunteer for study participation.[6] Thus, it is important to investigate the readability of trial descriptions available at ClinicalTrials.gov to ensure that the study information can be effectively conveyed to a wide audience with varying literacy.

Readability is known to affect the comprehensibility and communication effectiveness of text.[7] Developing readability measures and validating/applying them in different empirical settings have thus been of great interest to researchers and educators in a wide range of domains.[8–11] In this study, we evaluated the readability of ClinicalTrials.gov trial descriptions using 4 general-purpose readability scoring algorithms[12–16] in addition to a measure specifically developed to work with medical text.[17] The evaluation was conducted by comparing the readability of trial descriptions to the readability of 2 other related but distinct corpora: (1) Health Topics articles from MedlinePlus, a website created and maintained by the US National Library of Medicine to provide the general public high-quality information about diseases, conditions, and wellness;[18] and (2) clinician notes retrieved from the electronic health records (EHRs) system used at our institution that were created for conveying internal communication among medical professionals. We hypothesized that clinician notes

Correspondence to Kai Zheng PhD, Associate Professor, School of Public Health Department of Health Management and Policy, School of Information, University of Michigan, M3531 SPH II, 1415 Washington Heights, Ann Arbor, MI 48109, USA; kzheng@umich.edu; Tel: +1 (734) 936-6331; Fax: +1 (734) 964-4338.

would be most difficult to read, followed by clinical trial descriptions and then MedlinePlus Health Topics articles.

Previous Work

While we are not aware of any prior studies that have specifically looked at the readability of clinical trial descriptions, there has been empirical work to assess laypersons' ability to retell the trial descriptions they read, revealing considerable comprehension errors.[19,20] In addition, there have been studies investigating the readability of patient handouts and health education pamphlets,[21–23] online health content,[24–26] and informed consent forms.[11,27] These studies consistently found that patient and consumer health information resources tend to be difficult to read and require a literacy level higher than their intended audiences. For example, several studies demonstrated that patient consent forms for both patient care as well as clinical research were often written in very complex language,[28–32] with one study suggesting that surgical consent forms were written at the level of scientific journals.[33] Even Institutional Review Board consent form templates, which are intended to serve as the model for easy-to-understand text for laypersons, were deemed too complex for their proposed benchmarks (i.e., 5th- and 10th-grade reading level).[34]

Nevertheless, prior readability studies conducted in healthcare have several notable limitations. First, the sample size employed was often small (no more than a few hundred documents). Second, most of the studies applied readability scoring algorithms developed for general purposes that do not take into account the unique characteristics of healthcare text.[35] In this study, we attempted to address these limitations by analyzing a much larger dataset, consisting of all trials registered at ClinicalTrials.gov, all health articles from MedlinePlus, and 100 000 randomly selected clinician notes retrieved from an EHR system, using both general-purpose and medical specific readability assessment measures.

MATERIALS AND METHODS

Corpora and Text Features

We comparatively analyzed 3 corpora in this study. The first corpus contained all 165 988 clinical trial studies available at ClinicalTrials.gov as of April 30, 2014. Each trial provided a detailed description on the website about its study objectives, target patient population(s), and approaches in the following 4 structured sections: *Purpose*, *Eligibility*, *Contacts and Locations*, and *More Information*. Among them, the Purpose section often begins with a narrative introduction and ends with a "detailed description" subsection orienting readers to the most important facts about the study setting(s) and the overall research design (a sample is provided in Figure 1). Because these narratives serve as the entry point for readers to skim and decide whether the trial is of potential interest and worth exploring further, their readability is crucial. We therefore focused our analysis on these narratives extracted from the Purpose section. For convenience, we refer to this corpus as "Trial Description" in this paper. It contained approximately 1.5 million sentences and 33 million words.

In addition to "Trial Description," we analyzed 2 other corpora in order to obtain benchmarks to better interpret the results generated by readability scoring algorithms. The second corpus we analyzed consisted of all 955 "Health Topics" articles in English available at MedlinePlus as of April 30, 2014 (a sample is provided in Figure 2). Because these Health Topics articles are carefully curated by the US National Library of Medicine with the goal of disseminating high quality, easy-to-understand health information to the general public, they should be highly comprehensible by laypersons and thus should receive the best readability evaluation scores. This corpus is referred to

as "MedlinePlus" in this paper. It contained a total of 13 630 sentences and 136 032 words. Note that 3 other types of consumer-oriented materials available at MedlinePlus (a Medical Encyclopedia, Drug & Supplements information, and "Video & Cool Tools") were not included in our MedlinePlus corpus. This is because the Encyclopedia and the Drug & Supplements information are highly structured (i.e., a majority of this content is expressed via bullet points), and the "Video & Cool Tools" are mostly multimedia resources with little text for analysis.

The third corpus analyzed in this study consisted of 100 000 free-text narrative clinician notes (a sample is provided in Figure 3) randomly retrieved from the EHR system in use at the University of Michigan Health System, a tertiary care academic medical center with over 45 000 inpatient admissions and 1.9 million outpatient visits annually.[36] The homegrown EHR system, called CareWeb, allows clinicians to create notes via dictation/transcription or via typing.[37] These notes are generally unstructured, but clinicians could use simple, customizable text-based templates if desired. The corpus contained multiple document types retrieved from CareWeb generated in both inpatient and outpatient areas including admission notes, progress notes, radiology reports, and narrative assessments and plans. Because these clinician notes were composed by medical professionals and intended to be read by other medical professionals, we hypothesized that they would be most difficult to read across the 3 study corpora. This "EHR" corpus contained over 5 million sentences with about 56 million words.

For patient privacy protection reasons, all documents contained in the EHR corpus were first de-identified before they were used in this study. The identification was performed using the MITRE Identification Scrubber Toolkit,[38] and was based on a well-performing, locally developed model that we previously evaluated and reported in the literature.[39] Identifiable information including names, ages, and dates was replaced with standardized placeholders such as *[NAME]*, *[AGE]*, and *[DATE]*, with the majority of the clinical text remaining in its original form.

Surface Metrics

We first characterized each corpus using 4 surface metrics: average document length, average sentence length, vocabulary size, and vocabulary coverage. Vocabulary size is defined as the number of distinct words contained in the corpus. Vocabulary coverage is the percentage of distinct words that can be found in known dictionaries.

In the empirical study, we developed a medical English dictionary by combining entries extracted from an open-source English spell-check tool, GNU Aspell,[40] with medical terms extracted from an open-source medical spelling checker, OpenMedSpel.[41] In addition, we created a comprehensive dictionary of medical terminologies based on the content of the Unified Medical Language System (UMLS) 2013AB Metathesaurus, which includes more than 2.9 million concepts and 11.4 million unique concept names from over 160 source vocabularies.[42] We refer to the first dictionary as the Basic Medical English Dictionary ("Med-Dict") and the second as "UMLS."

Readability Measures

In this study, we applied 4 general-purpose readability scoring algorithms and one medical specific algorithm to measure the readability of each corpus. The 4 general-purpose measures have been popularly used in a wide range of domains, including healthcare, all of which produce a readability score in the form of the number of years of education required to comprehend the material under evaluation.[43] Below, we briefly summarize the mechanism underlying each of these

Figure 1: A sample *Purpose* section from ClinicalTrials.gov.

> **PURPOSE**
> We hypothesize that individuals with Alpha-1 Antitrypsin (AAT) deficiency have ongoing liver injury which is not detected by the usual blood tests used to look at liver function. This ongoing liver injury leads to cirrhosis in a significant number of adults with AAT deficiency.
>
> **Detailed Description:**
> Our overarching hypothesis is that liver disease in adults with AAT deficiency is the result of the accumulation of the abnormally folded protein within the endoplasmic reticulum of the hepatocyte. In some individuals, the intrinsic cellular mechanisms of the hepatocyte are sufficient to clear adequate amounts of the abnormally folded protein such that liver disease does not occur. In AAT deficient individuals who develop liver disease, environmental and other genetic factors stress the hepatocyte, and the normal cellular mechanisms that maintain homeostasis are disrupted, leading to liver disease.
>
> For this proposal, our hypothesis is that the prevalence of liver disease in adults with AAT is higher than previously reported because liver injury and fibrosis is not accurately detected by available routine liver testing. Testing this hypothesis will require an initial evaluation for liver disease with liver function testing and imaging, and then histologic confirmation by liver biopsy.

Figure 2: A sample MedlinePlus Health Topics article on "Aortic Aneurysm".

> **Aortic Aneurysm**
>
> An aneurysm is a bulge or "ballooning" in the wall of an artery. Arteries are blood vessels that carry oxygen-rich blood from the heart to other parts of the body. If an aneurysm grows large, it can burst and cause dangerous bleeding or even death.
>
> Most aneurysms are in the aorta, the main artery that runs from the heart through the chest and abdomen.
>
> There are two types of aortic aneurysm:
>
> - Thoracic aortic aneurysms - these occur in the part of the aorta running through the chest
> - Abdominal aortic aneurysms - these occur in the part of the aorta running through the abdomen
>
> Most aneurysms are found during tests done for other reasons. Some people are at high risk for aneurysms. It is important for them to get screening, because aneurysms can develop and become large before causing any symptoms Screening is recommended for people between the ages of 65 and 75 if they have a family history, or if they are men who have smoked. Doctors use imaging tests to find aneurysms. Medicines and surgery are the two main treatments.

general-purpose readability measures. More in-depth descriptions of these measures are provided in Supplementary Appendix A:

1. New Dale-Chall (NDC), computed based on the average number of words per sentence and the percentage of "unfamiliar" words not covered in a pre-defined dictionary[12,13];
2. Flesch-Kincaid Grade Level (FKGL), computed based on the weighted sum of the average number of words per sentence and the average number of syllables per word, and then adjusted by a baseline score[14];
3. Simple Measure of Gobbedygook (SMOG), computed based on the square root of the average number of syllables per word of words that have 3 or more syllables.[15]

4. Gunning-Fog Index (GFI), computed based on the average number of words per sentence and the complexity-syntax patterns of the words.[16]

The fifth readability scoring algorithm, called the Medical-Specific Readability Measure (MSRM), was developed by co-authors of this paper (Q.T.Z. and J.P.) specifically for assessing the readability of medical text.[17] Besides average sentence length and average word length, MSRM makes use of several additional text features such as average number of sentences per paragraph and parts of speech. Rather than producing an absolute score, MSRM estimates the relative distance of the features of the text being evaluated to those of a set of *easy-to-read* text samples and a set of *difficult-to-read* text samples. The

RESEARCH AND APPLICATIONS

Figure 3: A sample clinical note from University of Michigan Health System.

[NAME] is a [AGE] old female who underwent a composite resection of her mandible and mouth floor with local advancement flap and tracheostomy on [DATE] transferred to the SICU overnight for respiratory distress. She was initially diagnosed with an invasive squamous cell carcinoma of the mandibular alveolus. She was staged at a T3 N1 squamous cell carcinoma of the right mandibular alveolus. She underwent an uneventful surgery and post-operatively, she was doing well with a medicine team following for her multiple comorbidities including diabetes, COPD, HTN and GERD. She was though to be fluid positive over the last several days. She was receiving 200cc of fluid an hour over the 14 hrs prior to her decompensation last evening. Yesterday evening, she began to have increasing respiratory distress with increasing oxygen requirements on trach mask. She was placed on 80% and transferred to the SICU after receiving 60mg of Lasix. Upon arrival, she notes some difficulty with breathing, denies any chest pain, abdominal pain, cough.

---

easy-to-read samples consist of content extracted from online health education materials whereas the *difficult-to-read* samples consist of text extracted from scientific journal articles and medical textbooks. The scores produced by the algorithm range between −1 and 1, wherein 1 indicates the best readability. The mathematical underpinnings of MSRM can be found in the original publication[17] as well as in Supplementary Appendix A.

Analysis Procedures

The readability of each of the study corpora was independently evaluated using the 5 scoring algorithms. We also produced a composite score by averaging the grade level metrics generated by the 4 general-purpose measures. No stop words were removed prior to the analysis as it might change the text features and subsequently affect the readability scoring. Pairwise differences among the readability scores of the 3 corpora were conservatively tested using Analysis of variance (ANOVA) with Tukey's Honestly Significant Difference. All statistical analyses were performed in *R* version 3.0.2. The Institutional Review Board at the University of Michigan reviewed and approved the research protocol of this study.

RESULTS

Surface Metrics

The surface metrics of the 3 study corpora are reported in Table 1. Consistent with our hypothesis, MedlinePlus Health Topics articles appear to the easiest to read as they have the shortest average sentence length, smallest vocabulary size, and highest vocabulary coverage (97.1% by Med-Dict alone and 99.4% by Med-Dict and UMLS combined). In comparison, the average sentence length of the Trial Description corpus is more than twofold longer than that of MedlinePlus, and its vocabulary size is more than twenty times larger. Further, only one third of the words used in ClinicalTrials.gov descriptions are covered by Med-Dict. The EHR corpus, not surprisingly, has the largest vocabulary size and the least vocabulary coverage by known dictionaries, and is therefore likely most difficult to read.

Readability Scores

Table 2 reports the readability scores. All 5 measures consistently rated ClinicalTrials.gov trial descriptions as the most difficult corpus, which requires 15.8–21.1 years of education on average to be able to proficiently read and understand. The MedlinePlus Health Topics articles were consistently rated as the corpus that is easiest to read, requiring no more than a high school level of education (8.0–11.3 years). The scores of the EHR corpus always fell in the middle range.

The Tukey's Honestly Significant Difference test results showed that the differences across the mean readability scores of the 3 corpora are all statistically significant regardless of the readability measure used. In Table 3, we provide some sample narratives illustrating text that was rated easy to read vs. text that was rated difficult to read.

Figure 4 depicts the distribution of readability scores among the documents in each corpus. Figure 5 illustrates the variation among the scores. As shown in Figures 4 and 5, a majority of ClinicalTrials.gov trial descriptions were rated more difficult to read compared to the EHR and MedlinePlus corpora, which suggests that the findings of this study were robust and were not caused by a few outliners. Among the 3 corpora studied, the readability scores of MedlinePlus Health Topics articles have the least amount of variation, whereas the readability scores of ClinicalTrials.gov trial descriptions are most widely spread out.

DISCUSSION

While there have been studies assessing the readability of healthcare text such as patient education materials and informed consent forms,[11,21–23,27–33] to the best of our knowledge, no research has been conducted to date to evaluate the readability of trial descriptions at ClinicalTrials.gov. As this federal registry plays an important role in informing the general public about clinical trial studies, information published at the website should be prepared in a manner that can be easily read and understand by laypersons. Our evaluation however suggests concerning results. Every readability scoring algorithm we employed in the study rated ClinicalTrials.gov trial descriptions as the most difficult corpus, on average requiring 18 years of education, or a postgraduate level, to proficiently read and comprehend. While these readability algorithms do not provide precise measures of the number of years of education required to comprehend the material, our results strongly suggest that ClinicalTrials.gov trial descriptions have severe readability issues.

This study was not designed to provide concrete guidelines as to how to improve the readability of ClinicalTrials.gov. That said, the results do suggest several areas where potential improvements could be made. For example, the very long average sentence length of the Trial Description corpus adversely affected its readability scores. Breaking down long sentences into shorter ones can thus be a quick way to improve the readability of many trial descriptions. Further, about two thirds of the words used in ClinicalTrials.gov trial descriptions are not found in the basic medical English dictionary. Those submitting these descriptions could find the process of changing all these terms to be a significant burden, or they may view trial descriptions without complex

medical nomenclatures to be less scientifically rigorous. Thus, an alternative strategy might be to provide a consumer-oriented companion version of the description that uses plain layperson English; e.g., "chickenpox virus" instead of "varicella zoster virus," and "removal of kidney stone" instead of "nephrolithotomy."

In this study, we applied multiple measures to ensure the reliability of readability scoring. The fact that MedlinePlus Health Topics articles were consistently rated as the easiest-to-read corpus to a certain degree validates the readability measures we applied. The finding that clinician notes are generally easier to read than ClinicalTrials.gov trial descriptions is however surprising. This might be due to the fact that the available readability scoring algorithms are not best suited to evaluate the readability of clinician notes due to their unique characteristics. Prior research does show that machine-rated readability of

clinical notes often generates convoluted results: when evaluated at the lexical level, the readability of clinical notes tends to be comparable to that of easy-to-read documents (e.g., consumer-oriented education materials); while when evaluated at the syntactic and semantic level, it tends to be comparable to that of difficult-to-read documents (e.g., publications in scientific journals).[44] Another possible explanation is that clinician notes "speed-written" in time-sensitive patient care environments are often succinct, and complex medical words are often abbreviated or "acronymized," resulting in shorter sentences and shorter words with fewer syllables that may work in their favor when rated by readability scoring algorithms. This, however, could make the document much more difficult to read by patients.[45] With

Table 1: Surface metrics

Surface metrics	Trial Description	MedlinePlus	EHR
Average sentence length (number of words)	26.1 ✋	10.2 👍	12.3
Vocabulary size	147 978	6939 👍	307 750 ✋
Vocabulary covered by Med-Dict (%)	34.8	97.1 👍	15.3 ✋
Vocabulary covered by UMLS (%)	38.0	66.9 👍	17.6 ✋
Vocabulary covered by UMLS + Med-Dict (%)	53.7	99.4	24.6 ✋

👍 Likely associated with the best readability; ✋ Likely associated with the poorest readability.

Table 2: Readability scores

Scoring algorithm	Trial Description	MedlinePlus	EHR
NDC	15.8 ± 0.8 ✋	11.3 ± 2.3 👍	15.1 ± 1.7
FKGL	17.2 ± 4.2 ✋	8.0 ± 1.4 👍	9.1 ± 1.9
SMOG	17.9 ± 3.1 ✋	10.7 ± 1.3 👍	11.9 ± 1.5
GFI	21.1 ± 4.6 ✋	10.9 ± 1.9 👍	12.9 ± 2.2
Average of NDC, FKGL, SMOG, and GFI	18.0 ± 3.0 ✋	10.2 ± 1.5 👍	12.2 ± 1.5
MSRM	−0.44 ± 0.28 ✋	− 0.10 ± 0.23 👍	− 0.36 ± 0.18

👍 Best readability; ✋ Worst readability.
Abbreviations: NDC, New Dale-Chall; FKGL, Flesch-Kincaid Grade Level; SMOG, Simple Measure of Gobbedygook; GFI, Gunning-Fog Index; MSRM, Medical-Specific Readability Measure.

Table 3: Sample text from each of the study corpora

Corpus	Readability	Average Number of Years of Eduaction Requried for Proficient Reading[a]	Sample text
Trial Description	Easy	7.2	"This study plans to learn more about the immune system's response to breast cancer in young women."
	Hard	58.9	"The primary objectives of this study are: To evaluate the safety and tolerability of TH-302 monotherapy and in combination with bortezomib in subjects with relapsed/refractory multiple myeloma. To identify the dose-limiting toxicities and determine the maximum tolerated dose of TH-302 monotherapy and in combination with bortezomib in subjects with relapsed/refractory multiple myeloma. To identify a recommended Phase 2 dose for TH-302 and dexamethasone with or without bortezomib in subjects with relapsed/refractory multiple myeloma."
MedlinePlus	Easy	5.2	"Did you know that the average person has 5 million hairs? Hair grows all over your body except on your lips, palms and the soles of your feet. It takes about a month for healthy hair to grow half an inch. Most hairs grow for up to 6 years and then fall out. New hairs grow in their place."
	Hard	17.2	"Dupuytren's contracture: a hereditary thickening of the tough tissue that lies just below the skin of your palm, which causes the fingers to stiffen and bend."
EHR	Easy	4.6	"Will continue to follow and assess identified deficits and goals."
	Hard	28.6	"Right hydroureter confirmed by retrograde pyelogram prior to stent placement." "CHF with ischemic cardiac myopathy and ejection and an ejection fraction of 35%. PVOD with bilateral carotid stenosis."

[a]Average scores pooling the results generated by the 4 conventional measures.

Figure 4: Distributions of average readability scores of the 3 study corpora (average scores were computed by pooling the results generated by the 4 conventional readability measures).
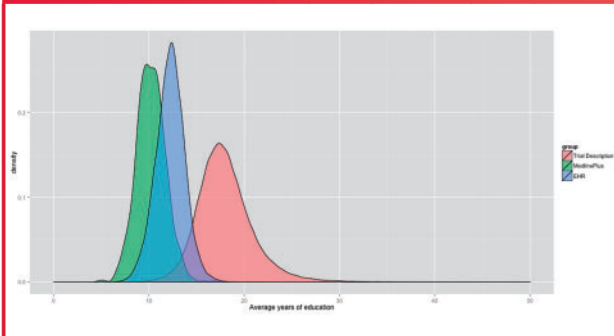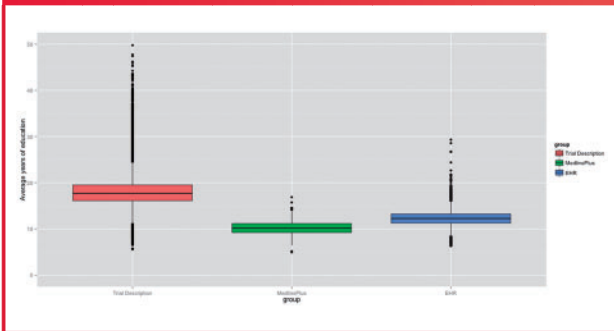
Figure 5: Variability of average readability scores (average scores were computed by pooling the results generated by the 4 conventional readability measures).

the US healthcare system now becoming increasingly "wired," this situation might improve, or it might deteriorate. On the one hand, modern and "meaningful use" certified EHRs discourage clinicians from writing or dictating unstructured, free-text notes in favor of generating such notes from structured templates or with text generated by computer algorithms. This change has the potential to improve readability as it may reduce the amount of abbreviations, acronyms, and non-standard medical language used in clinician notes. However, on the other hand, EHR-generated clinician notes populated from templates and structured data may appear absurd to human readers, and may lack sufficient context explaining the medical conditions described. Therefore, laypersons, perhaps also clinicians themselves, may find computer-generated notes more challenging to read and understand. To the best of our knowledge, the MSRM method used in this paper developed by Kim et al.[17] was the first and only attempt to develop custom readability scoring algorithms for clinician notes. Future work to develop better algorithms for assessing the readability of medical text, and for understanding and improving the readability of computer-generated clinician notes populated from templates and structured data in modern EHRs, is therefore critically needed.

This study has several limitations. First, even though we included a readability scoring algorithm tailored to evaluating medical documents, the general-purpose readability measures used in the study were not specifically designed to work with healthcare text. Therefore they may not be able to generate highly accurate results. Second, clinician notes that we analyzed were retrieved from a single EHR

system. The idiosyncrasies of the system, and of the local culture of clinical documentation, may also introduce biases into the study findings. Lastly, we only used computational methods to quantify the readability of the 3 study corpora, without engaging human readers who should ideally be drawn from a representative patient panel. Thus, the results of this study are only suggestive, not conclusive.

CONCLUSIONS

In this study, we used 5 different scoring algorithms to evaluate the readability of clinical trial descriptions available at ClinicalTrials.gov. The evaluation was conducted in comparison with MedlinePlus Health Topics articles and clinician notes retrieved from an EHR system. The results show that ClinicalTrials.gov trial descriptions are the most difficult corpus, on average requiring 18 years of education in order to proficiently read and comprehend. Because ClinicalTrials.gov serves critical functions of disseminating trial information to the general public and helping the trials recruit patient participants, there is a critical need to develop guidelines and strategies to improve the readability of ClinicalTrials.gov trial descriptions so they can be understood by laypersons.

FUNDING

COMPETING INTERESTS

The authors have no conflicts of interest to report.

CONTRIBUTORS

D.A.H., Q.M., K.C.T., and K.Z. designed the study. D.T.Y.W. collected the data, performed data analyses, and wrote the first draft of the manuscript, with the assistance of V.G.V.V., P.M.C., and L.C.A. J.P. and Q.T.Z. advised on computing the Medical-Specific Readability Measure. All authors reviewed and approved the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary material is available online at http://jamia.oxfordjournals.org/.

REFERENCES

1. Food and Drug Administration Amendments Act of 2007 (FDAAA), Section 801. 2007; PUBLIC LAW 110–85—SEPT. 27. http://clinicaltrials.gov/ct2/manage-recs/fdaaa; accessed November 15, 2014.
2. Angelis CD, Drazen JM, Frizelle FA, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. N Engl J Med. 2004;351(12):1250–1251.
3. Trends, Charts, and Maps. 2014. http://www.clinicaltrials.gov/ct2/resources/trends. Accessed November 15, 2014.
4. Background of ClinicalTrials.gov. http://www.clinicaltrials.gov/ct2/about-site/background. Accessed November 15, 2014.
5. Historical perspective on the development of ClinicalTrials.gov, and an overview of FDA's role in supporting the success of the database, and accessibility to clinical trials information by the public. http://www.fda.gov/ForConsumers/ByAudience/ForPatientAdvocates/ParticipatinginClinicalTrials/ucm143647.htm. Accessed November 15, 2014.
6. Jüni P, Altman DG, Egger M. Systematic reviews in health care: assessing the quality of controlled clinical trial. BMJ. 2001;323(7303):42–46.
7. Garner M, Ning Z, Francis J. A framework for the evaluation of patient information leaflets. Health Expect. 2012;15(3):283–294.

8. Rosemblat G, Logan R, Tse T, *et al.* Text features and readability: expert evaluation of consumer health text. Mednet 2006. http://www.mednetcongress.org/fullpapers/MEDNET-192_RosemblatGracielaA_e.pdf. Accessed November 15, 2014.

9. Zeng-Treitler Q, Kim H, Goryachev S, *et al.* Text characteristics of clinical reports and their implications on the readability of personal health records. *Stud Health Technol Inform.* 2007;129(Pt 2):1117–1121.

10. Adnan M, Warren J, Orr M. Assessing text characteristics of electronic discharge summaries and their implications for patient readability. In: *HIKM '10 Proceedings of the Fourth Australasian Workshop on Health Informatics and Knowledge Management.* 2010;108:77–84.

11. Terblanche M, Burgess L. Examining the readability of patient-informed consent forms. *J Clin Trials.* 2010;2:157–162.

12. Dale E, Chall JS. A formula for predicting readability: instructions. *Educ Res Bull.* 1948;27(2):37–54.

13. Chall JS, Dale E. *Manual for Use of the New Dale-Chall Readability Formula.* Brookline, MA: Brookline Books; 1995.

14. Kincaid JP, Fishburne RP, Rogers RL, *et al. Derivation of New Readability Formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy Enlisted Personnel.* Research Branch Report 8–75. Chief of Naval Technical Training: Naval Air Station Memphis; 1975.

15. McLaughlin GH. SMOG grading: a new readability formula. *J Reading.* 1969; 12:639–646.

16. Gunning R. *The Technique of Clear Writing.* New York, NY: McGraw-Hill International Book Co.; 1952.

17. Kim H, Goryachev S, Rosemblat G, *et al.* Beyond surface characteristics: a new health text-specific readability measurement. *AMIA Annu Symp Proc.* 2007;11:418–422.

18. About MedlinePlus. http://www.nlm.nih.gov/medlineplus/aboutmedlineplus.html. Accessed November 15, 2014.

19. Keselman A, Smith CA. A classification of errors in lay comprehension of medical documents. *J Biomed Inform.* 2012;45(6):1151–1163.

20. Smith CA, Hetzel S, Dalrymple P, *et al.* Beyond readability: investigating coherence of clinical text for consumers. *J Med Internet Res.* 2011;13(4): e104.

21. Yin HS, Gupta RS, Tomopoulos S, *et al.* Readability, suitability, and characteristics of asthma action plans: examination of factors that may impair understanding. *Pediatrics.* 2013;131(1):e116–e126.

22. Davis TC, Mayeaux EJ, Fredrickson D, *et al.* Reading ability of parents compared with reading level of pediatric patient education materials. *Pediatrics.* 1994;93(3);460–468.

23. Agarwal N, Hansberry DR, Sabourin V, *et al.* A comparative analysis of the quality of patient education materials from medical specialties. *JAMA Int Med.* 2013;173(13);1257–1259.

24. D'Alessandro DM, Kingsley P, Johnson-West J. The readability of pediatric patient education materials on the world wide web. *Arch Pediatr Adolesc Med.* 2001;155(7):807–812.

25. Risoldi Cochrane Z, Gregory P, Wilson A. Readability of consumer health information on the internet: a comparison of U.S. government-funded and commercially funded websites. *J Health Commun.* 2012;17(9);1003–1010.

26. Tian C, Champlin S, Mackert M, *et al.* Readability, suitability, and health content assessment of web-based patient education materials on colorectal cancer screening. *Gastrointest Endosc.* 2014;80(2);284–290.e2.

27. Langford AT, Resnicow K, Dimond EP, *et al.* Racial/ethnic differences in clinical trial enrollment, refusal rates, ineligibility, and reasons for decline among patients at sites in the National Cancer Institute's Community Cancer Centers Program. *Cancer.* 2014;120(6);877–884.

28. Baker MT, Taub HA. Readability of informed consent forms for research a veteran administration medical center. *JAMA.* 1983;250(19):2646–2648.

29. Tarnowski KJ, Allen DM, Mayhall C, *et al.* Readability of pediatric biomedical research informed consent forms. *Pediatrics.* 1990;85(1):58–62.

30. Meade CD, Howser DM. Consent forms: how to determine and improve their readability. *Oncol Nurs Forum.* 1992;19(10):1523–1528.

31. Hopper KD, TenHave TR, Hartzel J. How much do patients understand? Informed consent forms for clinical and research. *AJR Am J Roentgenol.* 1995;164(2):493–496.

32. Grossman SA, Piantadosi S, Covahey C. Are informed consent forms that describe clinical oncology research protocols readable by most patients and their families? *J Clin Oncol.* 1996;12(10):2211–2215.

33. Grundner TM. On the readability of surgical consent forms. *N Engl J Med.* 1980;302(16):900–902.

34. Paasche-Orlow MK, Taylor HA, Brancati FL. Readability standards for informed-consent forms as compared with actual readability. *N Engl J Med.* 2003;348(8):721–726.

35. Wu TY, Hanauer DA, Mei Q, *et al.* Applying multiple methods to assess the readability of a large corpus of medical documents. *Stud Health Technol Inform.* 2013;192:647–651.

36. Facts & Figures, University of Michigan Health System. http://www.uofmhealth.orgabout%20umhs/facts-figures. Accessed November 15, 2014.

37. Zheng K, Mei Q, Yang L, *et al.* Voice-dictated versus typed-in clinician notes: linguistic properties and the potential implications on natural language processing. *AMIA Annu Symp Proc.* 2011;2011:1630–1638. Epub 2011 October 22.

38. Aberdeen J, Bayer S, Yeniterzi R, *et al.* The MITRE Identification Scrubber Toolkit: design, training, and assessment. *Int J Med Inform.* 2010;79(12): 849–859.

39. Hanauer D, Aberdeen J, Bayer S, *et al.* Bootstrapping a de-identification system for narrative patient records: cost-performance tradeoffs. *Int J Med Inform.* 2013;82(9):821–831.

40. Atkinson K. GNU Aspell. http://aspell.net/. Accessed November 15, 2014..

41. e-MedTools. OpenMedSpel - Opensource Medical Spelling. http://e-medtools.com/openmedspel.html. Accessed November 15, 2014

42. Unified Medical Language System (UMLS) 2013AB Release Information. http://www.nlm.nih.gov/archive/20140415/research/umls/knowledge_sources/metathesaurus/release/index.html. Accessed November 15, 2014.

43. Ley P, Florio T. The use of readability formulas in health care. *Psychol, Health Med.* 1996;1(1);7–28.

44. Zeng-Treitler Q, Kim H, Goryachev S, *et al.* Text characteristics of clinical reports and their implications on the readability of personal health records. *Stud Health Technol Inform.* 2007;129(Pt 2):1117–1121.

45. Adnan M, Warren J, Orr M. Assessing text characteristics of electronic discharge summaries and their implications for patient readability. In: *HIKM '10 Proceedings of the Fourth Australasian Workshop on Health Informatics and Knowledge Management.* 2010;108:77–84.

AUTHOR AFFILIATIONS

[1]School of Information, University of Michigan, Ann Arbor, MI, USA

[2]Department of Pediatrics, University of Michigan, Ann Arbor, MI, USA

[3]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA

[4]School of Nursing, University of Michigan, Ann Arbor, MI, USA

[5]Center for Health Communication Research, University of Michigan, Ann Arbor, MI, USA

[6]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA

[7]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA

[8]School of Public Health Department of Health Management and Policy, University of Michigan, Ann Arbor, MI, USA

RESEARCH AND APPLICATIONS