Assessing the Readability of Online Patient Education Materials in Obstetrics and Gynecology using Traditional Measures: Comparative Analysis and Limitations

Anunita Nattam, Tripura Vithala, Tzu-Chun Wu, Shwetha Bindhu, Gregory Bond, Hexuan Liu, Amy Thompson, Danny Wu

Submitted to: Journal of Medical Internet Research
on: February 07, 2023

Table of Contents

Assessing the Readability of Online Patient Education Materials in Obstetrics and Gynecology using Traditional Measures: Comparative Analysis and Limitations

Anunita Nattam[1*]; Tripura Vithala[1*]; Tzu-Chun Wu[1] PhD; Shwetha Bindhu[1] BSc; Gregory Bond[2] MSc; Hexuan Liu[3] PhD; Amy Thompson[1] MD; Danny Wu[1, 2] PhD, MSI

[1]College of Medicine University of Cincinnati Cincinnati US
[2]Department of Computer Science University of Cincinnati Cincinnati US
[3]School of Criminal Justice University of Cincinnati Cincinnati US
[*]these authors contributed equally

Corresponding Author:
Danny Wu PhD, MSI
Department of Computer Science
University of Cincinnati
2610 University Circle
Cincinnati
US

Abstract

Background: Patient education materials (PEMs) can be vital sources of information for the general population. However, despite American Medical Association (AMA) and National Institute of Health (NIH) recommendations to make PEMs easier to read for patients with low health literacy, they often do not adhere to these recommendations. The readability of online PEMs in the Obstetrics and Gynecology (OB/GYN) field, in particular, has not been thoroughly investigated.

Objective: The study sampled online OB/GYN PEMs and aimed to examine agreeability across traditional readability measures (RMs), adherence of online PEMs to AMA and NIH recommendations, and whether readability level of online PEMs varied by publication source and medical topic.

Methods: A total of 1,576 online OB/GYN PEMs were collected via three major search engines. Ninety-three were excluded due to shorter content (less than 100 words), yielding 1,483 PEMs for analysis. Each PEM was scored by four traditional readability measures (TRMs), including Flesch-Kincaid Grade Level (FKGL), Gunning-Fog Index (GFI), Simple Measure of Gobbledygook (SMOG), and the Dale-Chall (DCL). The PEMs were categorized based on publication source and medical topic by two research team members. The readability scores of the categories were compared statistically.

Results: Results indicated that the four TRMs did not agree with each other, leading to the use of an averaged readability (composite) score for comparison. The composite scores across all online PEMs were not normally distributed and had a median at the 11th grade. Governmental PEMs had the lowest readability level amongst source categorization and PEMs about menstruation had the highest readability.

Conclusions: This study found that online OB/GYN PEMs did not meet the AMA and NIH readability recommendations and would be difficult to read and comprehend for patients with low health literacy. Both findings connected well to the literature. This study highlights the need to improve the readability of OB/GYN PEMs to help patients make informed decisions. While research has been done to create more sophisticated readability measures for medical and health documents. Once validated, these tools need to be utilized by online content creators of health education materials.

(JMIR Preprints 07/02/2023:46346)
DOI: https://doi.org/10.2196/preprints.46346

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Assessing the Readability of Online Patient Education Materials in Obstetrics and Gynecology using Traditional Measures: Comparative Analysis and Limitations

Abstract

Background: Patient education materials (PEMs) can be vital sources of information for the general population. However, despite American Medical Association (AMA) and National Institute of Health (NIH) recommendations to make PEMs easier to read for patients with low health literacy, they often do not adhere to these recommendations. The readability of online PEMs in the Obstetrics and Gynecology (OB/GYN) field, in particular, has not been thoroughly investigated.

Objective: The study sampled online OB/GYN PEMs and aimed to examine 1) agreeability across traditional readability measures (TRMs), 2) adherence of online PEMs to AMA and NIH recommendations, and 3) whether the readability level of online PEMs varied by online source and medical topic. This study is not a scoping review, rather, it focused on scoring the readability of OB/GYN PEMs using the traditional measures to add empirical evidence to the literature.

Methods: A total of 1,576 online OB/GYN PEMs were collected via three major search engines. Ninety-three were excluded due to shorter content (less than 100 words), yielding 1,483 PEMs for analysis. Each PEM was scored by four traditional readability measures (TRMs), including Flesch-Kincaid Grade Level (FKGL), Gunning-Fog Index (GFI), Simple Measure of Gobbledygook (SMOG), and the Dale-Chall (DCL). The PEMs were categorized based on publication source and medical topic by two research team members. The readability scores of the categories were compared statistically.

Results: Results indicated that the four TRMs did not agree with each other, leading to the use of an averaged readability (composite) score for comparison. The composite scores across all online PEMs were not normally distributed and had a median at the 11th grade. Governmental PEMs were the easiest to read amongst source categorizations and PEMs about menstruation were the hardest to read. However, the differences in the readability scores among the sources and the topics were small.

Conclusions: This study found that online OB/GYN PEMs did not meet the AMA and NIH readability recommendations and would be difficult to read and comprehend for patients with low health literacy. Both findings connected well to the literature. This study highlights the need to improve the readability of OB/GYN PEMs to help patients make informed decisions. Research has been done to create more sophisticated readability measures for medical and health documents. Once validated, these tools need to be utilized by online content creators of health education materials.

Keywords: Obstetrics and Gynecology; Online Patient Education Materials; Readability

Introduction

Readability assessment of online health information (OHI) provides valuable insight into how much information is easily understandable, to effectively empower laypersons to make informed health decisions[1]. Specifically, recognizing the impact of the situational reading level demands of the population and implementing that knowledge to simplify OHI, particularly materials geared towards patient education, can act as a mediating factor in health and clinical decision-making for patients by simplifying clinician-patient communication and reducing the complexities of the healthcare system for the patient[2]. The aim of online patient education materials (PEMs) should be to support interactive health literacy within an individual, giving them the ability to extract health information and derive meaning from different sources. This ability grants patients the opportunity to engage in interactions with healthcare professionals, fostering greater understanding and shared decision making. Therefore, using the online population's reading level and literacy capacity as a metric to guide PEMs can increase the usability and effectiveness of PEMs[3]. Furthermore, as the internet now serves as the primary source of information in modern society, and the prevalence of smart phones has allowed the expansion of the internet to a wider population, improving the quality of OHI and PEMs can reach a diversity of populations, and can decrease the burdens of frontline professional support and patients alike[4]..

Current recommendations made by the American Medical Association (AMA) and the National Institutes of Health (NIH) suggest that all PEMs should be at a 6th-8th grade reading level or lower[5,6]. However, previous studies conducted broadly throughout clinical specialties indicate these guidelines are not regularly followed[7]. Moreover, our preliminary scoping review conducted prior to this study showed that there is a lack of readability assessment studies in the field of Obstetrics and Gynecology (OB/GYN). Since OB/GYN PEMs found online are often accessed to seek guidance on a wide range of symptoms, diagnoses, and treatments, it provides patients with information on a spectrum of preventable and curable gynecological diseases. One important public health example is Human papillomavirus (HPV) and cervical cancer, for which preventative care via vaccination and early detection through pap smear screening are available. Given that OB/GYN is a primary healthcare field, lack of access to appropriate online PEMs may be a limiting factor in overcoming poor interactive health literacy and its associated outcomes[8].

Over recent years, popular traditional readability measures (TRMs), such as the Flesch-Kincaid Grade Level (FKGL)[9], Gunning-Fog Index (GFI)[10], Simple Measure of Gobbledygook (SMOG) [11], and the Dale-Chall (DCL)[12] formulas generate a grade-level score associating readability with the grade level of education needed to understand a document. For example, a document with a readability of the 5th-grade level is easier to read and understand than a document with a 10th-grade reading level. The TRMs used similar textual features in their formula (Table 1), including the average number of words per sentence, the average number of syllables per word, the average number of sentences, and custom easy versus difficult word lists.

Previous research done to assess the readability of OB/GYN PEMs, although scarce, has shown that most existing OB/GYN PEMs are written between the 9th and 12th grade level[13–18]. However, several of these studies focus solely on PEMs published by academic sources, limiting the study to a specific subtype of PEMs that patients are likely to encounter online[13,14]. Moreover, in these studies where TRMs were used, scores generated by each measure were not always consistent, indicating a limitation related to the validity of these TRMs used to widely assess patient readability and understanding[19].

Therefore, to address the gaps in existing research and further address the issue of readability in online OB/GYN PEMs, four research questions were produced. First, do the scores generated by four traditional readability measures agree with each other? (RQ1) This research question was prompted by previous research indicating that certain readability measures do not agree with each other, particularly FKGL, SMOG and GFI. [20–22]. Second, do the PEMs found in the field of OB/GYN follow the 6th-8th grade level readability recommendation by the AMA and NIH? (RQ2) Third, are there differences in readability level by sources (government agencies, nonprofit organizations, educational websites, or commercial entities)? (RQ3) Finally, will the readability level change when discussing different gynecological processes and topics, such as menstruation, pregnancy, cancer, general disease, and procedural information? (RQ4)

Methods

Data Collection and Coding

The PEMs were selected by searching the keywords “OB/GYN” AND “Patient Education Materials” in the three most used search engines (i.e., Google, Yahoo, and Bing!), and selecting the results on the first page (top 10 results) from each. This is not considered as a scoping review. Patient search behavior and patterns were mimicked in order to generate the most realistic search results. After removing repetitions and broken links, 15 different website sources from varying categories were identified containing 1,576 PEMs (Supplement 1). The PEMs were crawled by using the Python Selenium library, which automatically visited the individual web pages and extracted the full text for each article with formatting retained. In terms of coding, two reviewers (TV and SP) surveyed the texts together and coded a small sample to determine the codebook by sources and topics. This process is considered as a bottom-up approach as opposed to a top-down approach using pre-defined categories. Then, the two reviewers coded all the PEMs independently to produce the source and topic categories. Any discrepancy was resolved by the third reviewer (AN). Of note, each PEM might have multiple topics. The primary topic of each PEMs was determined and used for the subsequent statistical analysis. The process of data collection and screening is shown in Figure 1.

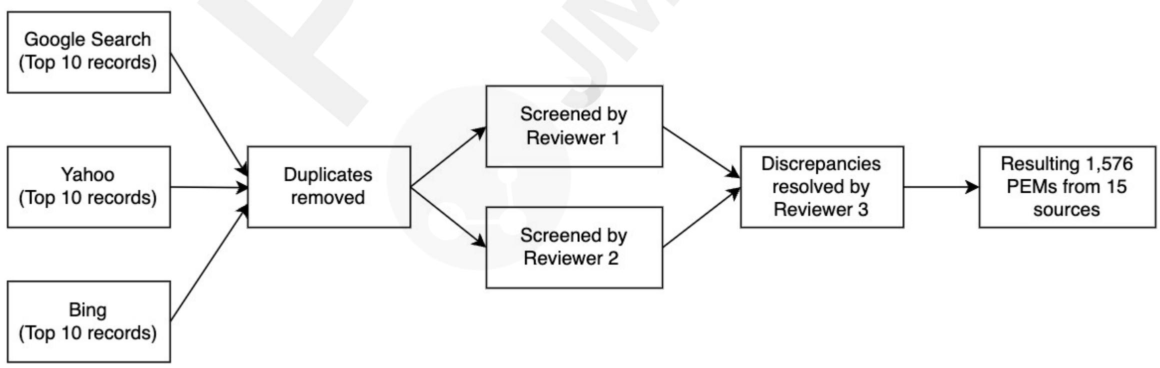


Figure 1. Data collection and screening process

Readability Scoring

Among the 1,576 PEMs crawled, 93 of them did not have enough number of words for GFI scoring (minimum 100 words). These PEMs were removed in the readability scoring and subsequent analysis, resulting in a total of 1,483 OB/GYN PEMs analyzed. To obtain readability scores, the text of the collected PEMs was first cleaned, primarily by removing the title and any sections unrelated to the health material, such as mentions of other articles and advertisements, and then scored based on the TRM formulas. Specifically, each PEM was broken down into sentences, words, and syllables. Sentences were derived by splitting the text by common end-of-sentence punctuation, while words were found by using a space delimiter for each sentence; syllables and base words were found mainly through the NLTK Python library. The Natural Language Tool Kit (NLTK) Python library is a collection of modules to process text and perform Natural Language Processing (NLP) tasks[23]. Surface metrics of the text, such as the average number of words per sentence, were then calculated based on the four TRMs to generate the scores. Of note, the DCL raw scores were transformed into grade levels using its reading grade level formula (Table 1). Only the integer part of the result from the mathematical equations will be preserved as the reading grade level. A random sample of 50 PEMs was selected and hand calculations were conducted and compared to the Python-generated values to ensure validity of our methods. Of note, the decision to use these four TRMs (FKGL, GFI, SMOG, and DC) was based on their use in previous studies assessing the readability of online PEMs, although not particularly in the field of OB/GYN[24–27].

Table 1: Mathematical Equation Used for Each TRM.

Formula	Equation
Flesch-Kincaid Grade Level (FKGL) [9]	Reading grade level = (0.39 × average no. of words per sentence) + (11.8 × average no. of syllables per word) − 15.59
Gunning Fog Index (GFI) [10]*	Reading grade level = 0.4 (average no. of words per sentence + no. of words with 3 or more syllables × [100 / no. of words])
Simple Measure of Gobbledygook (SMOG) [11]	If number of sentences >= 30: Reading Grade Level = 3 + square root of polysyllable count to the nearest perfect square. Else if number of sentences is between 1 and 29: Raw score = average no. of polysyllable per sentence * ratio of sentences) + no. of polysyllable Round the Raw score to the integer. Reading grade level = $\begin{cases} 5 & if\ Raw\ score \in [1,6] \\ R & if\ Raw\ score \in [R^2 - 7*R + 13, (R-2)*(R-3)] \\ 18 & if\ Raw\ score \geq 211 \end{cases}$ where $R = 6,7,\cdots,17$
Dale-Chall (DCL) [12]	Raw score = (0.0496 × average no. of words per sentence) + (15.79 × no. of words not found on a word list) / (no. of words) + 3.6365 Reading grade level = $\begin{cases} 4 & if\ Raw\ score < 5 \\ 2*Raw\ score - 5 & if\ Raw\ score \in ¿ \\ 3*Raw\ score - 14 & if\ Raw\ score \in ¿ \\ 16 & if\ Raw\ score \geq 10 \end{cases}$

*Requires at least 100 words in the input document

Data Analysis

To address our research questions, respective hypotheses were proposed. Firstly (RQ1), it was hypothesized that the TRMs would have acceptable agreeability (H1) since they used similar textual features and determined readability as grade levels. To test this hypothesis, each readability score of a document was converted to a grade level from Grade 4 to 16 due to the grade level mapping scale of DCL and treated as categorical data. In addition, the study tried to improve the agreement by combining the categories, e.g., less than 6th grade, 6th-8th grade, and above 8th grade. The Fleiss's Kappa was calculated to measure the degree of agreement between the different readability measures over that which would be expected by chance[28] The given kappa value k can be a negative number up to 1.00, with negative numbers indicating poor agreement, 0.00-0.20 as slight agreement, 0.21-0.40 as fair agreement, 0.41-0.60 as moderate agreement, 0.61-0.80 as substantial agreement, and 0.81-1.00 being almost perfect to perfect agreement. When H1 is held, meaning all four TRMs agree with each other substantially, FKGL will be used to represent the readability of a document since it is widely used (e.g., FKGL is implemented in Microsoft Word). Otherwise, a composite score (average or median of all four TRMs) will be created to represent the readability of a document.

Second (RQ2), we hypothesized that the PEMs would not follow the recommended guidelines of 6th-8th grade level readability (H2) since literature of readability assessment has shown this trend. To test this hypothesis, the composite scores were used to generate both categorical and numerical data. The normality of the composite scores was examined using the Shapiro-Wilk test. The null hypothesis of this test is to assume that the population of data is normally distributed. Since the significant value of the Shapiro-Wilk test on our data was less than 0.05, the population of data deviates significantly from a normal distribution. Therefore, Wilcoxon Sign test was used to examine if the composite scores overall are above the recommended level (8th grade).

Lastly, the study explored the possibility of differences in readability scores for PEMs coming from varying sources (RQ3) and topics (RQ4). It was hypothesized that the government sourced PEMs would have the lowest readability level (H3), making them the easiest to understand, since the government agencies claimed that their online PEMs have been curated and suitable for the public. On the other hand, it was hypothesized that the readability level of PEMs with different topics was similar (H4). To test these two hypotheses, the PEMs were manually categorized into four source categories (i.e., government, commercial, nonprofit, and educational) and five topic categories (General Disease, Pregnancy, Menstruation, Procedure, and Cancer). General Disease referred to any articles referencing diseases that do not affect women exclusively, cannot be placed into any of the other categories, and appeared on these OB/GYN PEM websites. The Kruskal-Wallis test was applied to detect any differences between medians in each category (main effect), followed by the pairwise Mann-Whitney test with Bonferroni Correction as the post-hoc test to assess whether the median of the composite scores of the different categories were statistically different. The median of the composite score was used because the composite scores were not normally distributed. Table 2 shows distribution of the PEMs analyzed in the present study by the source and topic categorization.

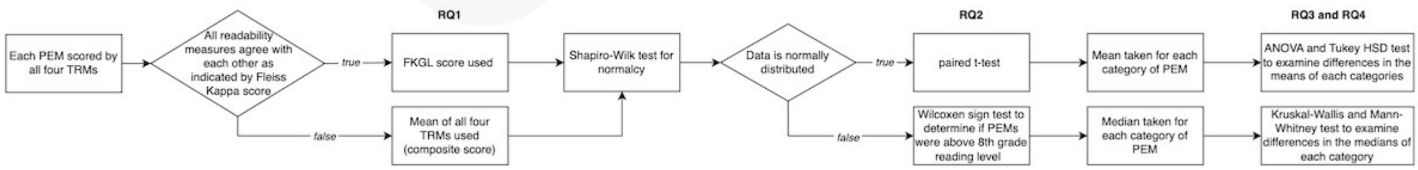


Figure 2. Process of data analysis

Table 2. Contingency Table of Online OB/GYN PEMs sources Based on Topics.

Topic \ Source	Educational	Non-Profit	Commercial	Government	**Total**
Pregnancy	553	531	34	27	**1,145**

General Disease	39	13	57	28	**137**
Procedure	42	22	24	10	**98**
Menstruation	24	3	22	6	**55**
Cancer	4	1	31	12	**48**
Total	**662**	**570**	**168**	**83**	**1,483**

Results

Hypothesis 1: Placing the PEMs into 13 categories from 4 to 16, that is one per grade level, resulted in a very low Fleiss Kappa value (K=.0025), which indicates slight agreement. Placing the readability scores into three categories (less than 6th grade, 6th to 8th grade, and greater than 8th grade) only resulted in a slight increase in the Fleiss Kappa (p = .033) and placing them into just two categories (less than or equal to 8 and greater than 8) resulted in a significant increase (p = .076). However, the agreement is still considered slight. As a result, the four TRMs did not perform fair agreement in our dataset. Then, the composite score of a document was generated by averaging the four TRMs. Figure 1 shows the distribution of the composite scores using both the mean and median. As shown in Figure 1, taking the average of the readability scores can be more conservative since the average considered a variety of responses in each TRM. On the other hand, the median involved only two out of four responses in our case, which resulted in the loss of half the information.

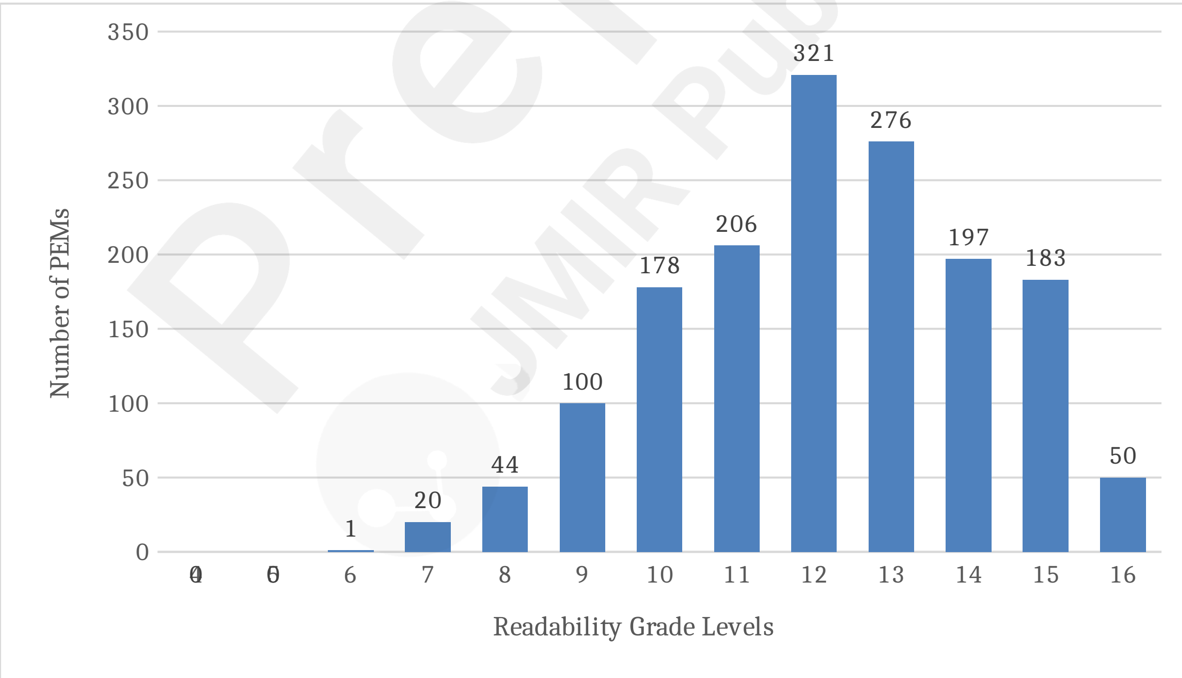


Figure 3. Distribution of the composite scores (average of the four TRMs) (RQ2)

Hypothesis 2: Using the composite score of the TRMs, the PEMs were found to be at the 11th grade reading level, which is significantly higher than the recommended 8th grade level ($p < .001$) using the Wilcoxon Sign test. The median only was used in these calculations, as the mean was shown to be skewed by extreme values.

Hypothesis 3: The Kruskal-Wallis Test showed that at least one median is significantly different from others (p < .001) among the four documentation sources. The Pairwise Mann-Whitney test indicated that the median readability level of the government source (11.25) was significantly lower than the commercial (13.00), educational (12.75), and Non-Profit (12.75). Meanwhile, the median readability level of the commercial source (13.00) was significantly higher than the Non-Profit source (12.75), but not different from the educational source (12.75). None of the categories had a median readability level less than the recommended 8th grade level.

Table 3: Results from Pairwise Mann-Whitney tests (* < 0.008) (RQ3).

Source Category Comparison		Median of Mean			
Category 1	Category 2	Median 1	Median 2	W	p-value
Commercial	Government	13.00	11.25	11223	2.02e-14*
Commercial	Educational	13.00	12.75	62131	0.01865
Commercial	Non-Profit	13.00	12.75	54775	0.00397*
Government	Educational	11.25	12.75	16356	7.40e-10*
Government	Non-Profit	11.25	12.75	15189	6.67e-08*
Educational	Non-Profit	12.75	12.75	192935	0.45960

Hypothesis 4: The Kruskal-Wallis Test showed there is at least one median which is significantly different from others (p=.001) amongst the five medical topics. The Pairwise Mann-Whitney Test indicated that the median readability levels of many groups were similar to each other. However, the median readability level of Menstruation (13.5) and Pregnancy (12.75) are significantly higher than General Disease (12.25).

Table 4: Results from Pairwise Mann-Whitney tests (* < 0.005) (RQ4).

Topic Category Comparison		Composite Scores (Mean)			
Category 1	Category 2	Median 1	Median 2	W	p-value
Menstruation	Procedure	13.5	12.75	3369	0.01034
Menstruation	General Disease	13.5	12.25	5271.5	1.54e-05*
Menstruation	Cancer	13.5	12.5	1720.5	0.01512
Menstruation	Pregnancy	13.5	12.75	38180	0.00734
Procedure	General Disease	12.75	12.25	7886.5	0.02232
Procedure	Cancer	12.75	12.5	2422.5	0.93110
Procedure	Pregnancy	12.75	12.75	54094	0.56460
General Disease	Cancer	12.25	12.5	2747.5	0.05963
General Disease	Pregnancy	12.25	12.75	62984	0.00017*
Cancer	Pregnancy	12.5	12.75	26927	0.64080

Discussion

Principal Findings

This study applied four TRMs to score the readability of OB/GYN online PEMs and compared the scores of the TRMs, which is the first to conduct such comparative analysis and provide empirical evidence to show the limitations of the TRMs. The findings showed that the four TRMs lacked agreement with one another, indicating the need for developing modern readability measures for PEMs and general OHI. Using the average score of TRMs as a proxy to the readability level, our study found that most OB/GYN PEMs required high readability levels despite the recommendation for online PEMs to be at the 6th-8th grade reading level. In addition, the study found that the online PEMs from the government source were slightly easier to read than other sources. However, they still require 11th grade level. It is important to provide health information in a consistent language so that OHI can be disseminated to all, regardless of source. Lastly, the comparison of the topic categories showed that the Menstruation and Pregnancy PEMs are harder to read than the General Disease. This opens new research questions and research investigations.

Implications

The finding where the four TRMs did not agree with one another connects well with recent research concerning TRMs. For example, one study evaluating the readability of Diabetes-related PEMs found that certain readability measures consistently indicated higher levels of readability than others[22]. This indicates that the existing measures do not reflect completely the readability level of health texts[34]. Moving away from syntax and surface language features, recent research has identified other factors, such as sentence complexity, use of passive voice, grammar frequency, and the patient’s familiarity with vocabulary, as additional components that need to be considered in measuring health text readability[36–38]. Passive voice, in particular, was shown to distinguish very complex texts from very simple texts, but was not able to distinguish intermediate complexities, making it an important language feature to study further [39]. Additionally, the evaluation of readability measures is critical, especially using human annotations since the readability measures may not reflect the actual readability and understandability of a piece text. Previous studies have found such discrepancy between the perceived document difficulty and actual difficulty[19,40]. Our study highlights the need to improve the readability of OB/GYN PEMs and ensure the validity and reliability of readability measures.

When looking at readability assessment studies done in medical specialties other than OB/GYN, required readability levels are consistently very high[29–31], showing a need for remediation across specialties like ophthalmology, cardiology, geriatric care, and beyond. For example, in one study specifically using TRM to analyze ophthalmology related PEMS, average readability was found to be at the 10th grade reading level[29]. Even with inconsistent TRM levels, human annotations have rated PEMs in several medical fields to be too difficult to understand[32]. Such studies have been repeated in various specialties over several years, highlighting the issue of readability in health education as a pertinent issue for some time.

Over the years, some studies have moved away from the use of TRM, most likely due to their poor ability to accurately assess health information. Recently developed readability assessment methods have adopted machine learning, artificial intelligence (AI), and natural language processing techniques (NLP), which have been shown to better predict readability when compared to human annotations[33–35]. However, these methods have not been widely used or promoted in the research community. With further advancement of AI and NLP, such as pre-trained large language models, it may become easier to assess readability, and even revise and generate easier-to-read health information automatically.

Limitations

This study has at least six limitations. First, the PEM search may not be comprehensive; it may miss other online PEMs in the field of OB/GYN. However, we used modern search engines to collect PEMs that were most likely to be accessed by laypersons. Therefore, the PEMs included in the present study should be frequently seen by laypersons. Since the study only included the results on the first page of each search engine, relevant websites may be listed on the 2nd page or after. Second, the PEMs were scored by only four TRMs. There are other TRMs (such as Fry readability graph[41] and FORCAST[42]) that could potentially be more useful and better adept at identifying the true readability of online PEMs. Moreover, there have been other readability measures developed recently for health information[19,34,37]. However, since many of the new readability assessment methods have not been widely used by the research community, they were excluded in the present study. Third, in the present study, the texts for all PEMs were taken without any content analysis, which can limit our understanding of the actual readability level. Fourth, although statistical significance was found in the results, there is not sufficient evidence to prove that small differences in grade level above grade 12 are meaningful since most readability measures are calibrated for school grades. Next, although the PEMs were cleaned, the title and some sections of the PEMs were removed, which may affect the readability score distribution[43] Lastly, the search results might have been slightly altered due to the presence of cache and other confounders in Google Search[44].

Future Work

Future work includes surveying the literature and summarizing the current advances in readability formula development, given the fact that although research has been conducted to create new readability formulas, very few are promoted and widely used to replace TRMs in the research community. Implementing these health information specific readability measures may provide some more insight into what must be done to improve the readability of online PEMs. Additionally, there will be further content analysis conducted on PEMs to create writing guidelines to support clinicians and patients alike, especially in the context of OB/GYN. Patient stakeholders or standardized patients can also inform the creation of new PEM content for validation of healthcare-specific readability tools. Lastly, a simple English thesaurus for OB/GYN PEMs can be developed to help those writing PEMs to simplify their materials.

Conclusion

The present study examined the readability of the PEMs in OB/GYN. While the OB/GYN PEMs were collected using search engines, which may introduce biases, the study found that most of the PEMs were hard to read, requiring high school or college level of education to read the content. More research is needed to evaluate the PEMs in OB/GYN in a more comprehensive manner and create writing guidelines to improve the readability, understandability, and actionable of these PEMs.

Acknowledgments

We would like to thank Mr. Mostafa Ghaderian for his effort in statistical analysis. We also thank Ms. Somya Pandey for her effort in document categorization and copy-editing.

Conflicts of Interest

None declared.

Abbreviations

AMA: American Medical Association
DCL: Dale-Chall
FKGL: Flesch-Kincaid Grade Level
GFI: Gunning-Fog Index
HPV: Human papillomavirus
NIH: National Institute of Health

OB/GYN: Obstetrics and Gynecology
OHI: Online health information
PEMs: Patient education materials
SMOG: Simple Measure of Gobbledygook
TRM: Traditional readability measures

References

1. Edmunds MR, Barry RJ, Denniston AK. Readability assessment of online ophthalmic patient information. JAMA Ophthalmol 2013 Dec;131(12):1610–1616. PMID:24178035

2. Elwyn G, Laitner S, Coulter A, Walker E, Watson P, Thomson R. Implementing shared decision making in the NHS. BMJ 2010 Oct 14;341:c5146. PMID:20947577

3. Nutbeam D, Lloyd JE. Understanding and Responding to Health Literacy as a Social Determinant of Health. Annu Rev Public Health 2021 Apr 1;42:159–173. PMID:33035427

4. Friis K, Lasgaard M, Osborne RH, Maindal HT. Gaps in understanding health and engagement with healthcare providers across common long-term conditions: a population survey of health literacy in 29,473 Danish citizens. BMJ Open 2016 Jan 14;6(1):e009627. PMID:26769783

5. Abu-Heija AA, Shatta M, Ajam M, Abu-Heija U, Imran N, Levine D. Quantitative Readability Assessment of the Internal Medicine Online Patient Information on Annals.org. Cureus 2019 Mar 6; doi: 10.7759/cureus.4184

6. Weiss BD. Health Literacy: A Manual for Clinicians. 2003. Available from: http://lib.ncfh.org/pdfs/6617.pdf [accessed Jun 5, 2023]

7. Readability of Patient Education Materials From High-Impact Medical Journals: A 20-Year Analysis - Michael K Rooney, Gaia Santiago, Subha Perni, David P Horowitz, Anne R McCall, Andrew J Einstein, Reshma Jagsi, Daniel W Golden, 2021. Available from: https://journals.sagepub.com/doi/10.1177/2374373521998847 [accessed Jun 6, 2023]

8. Cudjoe J, Budhathoki C, Roter D, Gallo JJ, Sharps P, Han H-R. Exploring Health Literacy and the Correlates of Pap Testing Among African Immigrant Women: Findings from the AfroPap Study. J Cancer Educ Off J Am Assoc Cancer Educ 2021 Jun;36(3):441–451. PMID:32410109

9. Flesch R. A new readability yardstick. J Appl Psychol 1948;32(3):221–233. doi: 10.1037/h0057532

10. Gunning R. The technique of clear writing. [S.l.]: McGraw-Hill; 1968. ISBN:0-07-025206-8

11. MC LAUGHLIN GH. SMOG Grading —– a New Readability Formula. J Read 1969;12:639–646.

12. Chall JS, Dale E. Manual for use of the new Dale-Chall readability formula. Cambridge, Mass.: Brookline Books; 1995. ISBN:1571290125

13. Fahimuddin FZ, Sidhu S, Agrawal A. Reading Level of Online Patient Education Materials From Major Obstetrics and Gynecology Societies. Obstet Gynecol 2019 May;133(5):987–993. PMID:30969212

14. Okuhara T, Okada H, Goto E, Kiuchi T. Readability Assessment of HPV Vaccination and Cervical Cancer Information: A Systematic Scoping Review. Healthc Basel Switz 2021 Sep 22;9(10):1246. PMID:34682926

15. Lange EMS, Shah AM, Braithwaite BA, You WB, Wong CA, Grobman WA, Toledo P. Readability, content, and quality of online patient education materials on preeclampsia. Hypertens Pregnancy 2015;34(3):383–390. PMID:26153628

16. Yee LM, Niznik CM, Simon MA. Examining the Role of Health Literacy in Optimizing the Care of Pregnant Women with Diabetes. Am J Perinatol 2016 Nov;33(13):1242–1249. PMID:27322666

17. AlKhalili R, Shukla PA, Patel RH, Sanghvi S, Hubbi B. Readability assessment of internet-based patient education materials related to mammography for breast cancer screening. Acad Radiol 2015 Mar;22(3):290–295. PMID:25488695

18. Freda MC, Damus K, Merkatz IR. Evaluation of the readability of ACOG patient education pamphlets. The American College of Obstetricians and Gynecologists. Obstet Gynecol 1999 May;93(5 Pt 1):771–774. PMID:10912984

19. Kauchak D, Leroy G. Moving Beyond Readability Metrics for Health-Related Text Simplification. IT Prof 2016 Jun;18(3):45–51. PMID:27698611

20. Wu DTY, Hanauer DA, Mei Q, Clark PM, An LC, Proulx J, Zeng QT, Vydiswaran VGV, Collins-Thompson K, Zheng K. Assessing the readability of ClinicalTrials.gov. J Am Med Inform Assoc JAMIA 2016 Mar;23(2):269–275. PMID:26269536

21. Mailloux S, Johnson M, Fisher D, Pettibone T. How reliable is computerized assessment of readability? Comput Nurs 1994 Nov 30;13:221–5.

22. Lipari M, Berlie H, Saleh Y, Hang P, Moser L. Understandability, actionability, and readability of online patient education materials about diabetes mellitus. Am J Health Syst Pharm 2019 Jan 25;76(3):182–186. doi: 10.1093/ajhp/zxy021

23. Lobur M, Romanyuk A, Romanyshyn M. Using NLTK for educational and scientific purposes. 2011 11th Int Conf Exp Des Appl CAD Syst Microelectron CADSM 2011. p. 426–428.

24. Cheng BT, Kim AB, Tanna AP. Readability of Online Patient Education Materials for Glaucoma. J Glaucoma 2022 Jun;31(6):438–442. doi: 10.1097/IJG.0000000000002012

25. Pearson K, Ngo S, Ekpo E, Sarraju A, Baird G, Knowles J, Rodriguez F. Online Patient Education Materials Related to Lipoprotein(a): Readability Assessment. J Med Internet Res 2022 Jan 11;24(1):e31284. PMID:35014955

26. Fortuna J, Riddering A, Shuster L, Lopez-Jeng C. Assessment of online patient education materials designed for people with age-related macular degeneration. BMC Ophthalmol 2020 Oct 2;20(1):391. PMID:33008367

27. Martin CA, Khan S, Lee R, Do AT, Sridhar J, Crowell EL, Bowden EC. Readability and Suitability of Online Patient Education Materials for Glaucoma. Ophthalmol Glaucoma 2022 Sep;5(5):525–530. doi: 10.1016/j.ogla.2022.03.004

28. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull US: American Psychological Association; 1971;76:378–382. doi: 10.1037/h0031619

29. Williams AM, Muir KW, Rosdahl JA. Readability of patient education materials in ophthalmology: a single-institution study and systematic review. BMC Ophthalmol 2016 Aug 3;16:133. PMID:27487960

30. Avra TD, Le M, Hernandez S, Thure K, Ulloa JG. Readability assessment of online peripheral artery disease education materials. J Vasc Surg 2022 Dec;76(6):1728–1732. PMID:35931399

31. Wang E, Kalloniatis M, Ly A. Assessment of patient education materials for age-related macular degeneration. Ophthalmic Physiol Opt J Br Coll Ophthalmic Opt Optom 2022 Jul;42(4):839–848. PMID:35521818

32. Alshehri A, Alghofaili N, ALshunaiber R, Alkadi L. Quality and Readability Assessment of Internet-Based Information on Common Prosthodontic Treatments. Int J Prosthodont 2022;35(1):62–67. PMID:33651024

33. Sung Y-T, Chen J-L, Cha J-H, Tseng H-C, Chang T-H, Chang K-E. Constructing and validating readability models: the method of integrating multilevel linguistic features with machine learning. Behav Res Methods 2015 Jun;47(2):340–354. PMID:24687843

34. Kim H, Goryachev S, Rosemblat G, Browne A, Keselman A, Zeng-Treitler Q. Beyond surface characteristics: a new health text-specific readability measurement. AMIA Annu Symp Proc AMIA Symp 2007 Oct 11;2007:418–422. PMID:18693870

35. Mukherjee P, Leroy G, Kauchak D. Using Lexical Chains to Identify Text Difficulty: A Corpus Statistics and Classification Study. IEEE J Biomed Health Inform 2019 Sep;23(5):2164–2173. PMID:30530380

36. Leroy G, Endicott JE, Kauchak D, Mouradi O, Just M. User Evaluation of the Effects of a Text Simplification Algorithm Using Term Familiarity on Perception, Understanding, Learning, and Information Retention. J Med Internet Res 2013 Jul 31;15(7):e2569. doi: 10.2196/jmir.2569

37. Zheng J, Yu H. Assessing the Readability of Medical Documents: A Ranking Approach. JMIR Med Inform 2018 Mar 23;6(1):e17. PMID:29572199

38. Ji M, Liu Y, Hao T. Predicting Health Material Accessibility: Development of Machine Learning Algorithms. JMIR Med Inform 2021 Sep 1;9(9):e29175. PMID:34468321

39. Ownby RL. Influence of Vocabulary and Sentence Complexity and Passive Voice on the Readability of Consumer-Oriented Mental Health Information on the Internet. AMIA Annu Symp Proc 2005;2005:585–588. PMID:16779107

40. Kauchak D, Leroy G, Hogue A. Measuring Text Difficulty Using Parse-Tree Frequency. J Assoc Inf Sci Technol 2017 Sep;68(9):2088–2100. PMID:29057293

41. Fry E. A Readability Formula That Saves Time. J Read 1968 Apr;11(7):513–516, 575–578.

42. Caylor JS, others. Methodologies for Determining Reading Requirements of Military Occupational Specialties. ERIC; 1973;

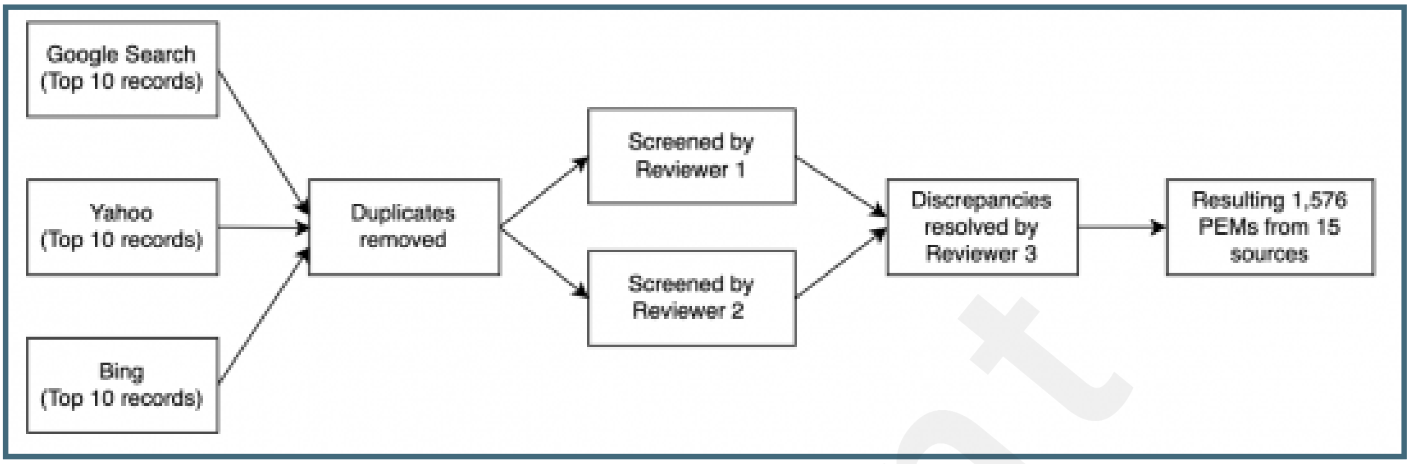
43. Comparison of Readability Scores for Written Health Information Across Formulas Using Automated vs Manual Measures | Public Health | JAMA Network Open | JAMA Network. Available from: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2799377 [accessed Jun 5, 2023]

44. Fajardo MA, Weir KR, Bonner C, Gnjidic D, Jansen J. Availability and readability of patient education materials for deprescribing: An environmental scan. Br J Clin Pharmacol 2019;85(7):1396–1406. doi: 10.1111/bcp.13912

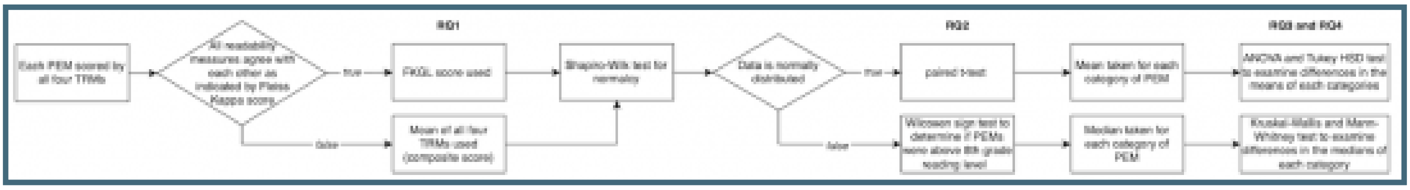
Supplementary Files

Figures

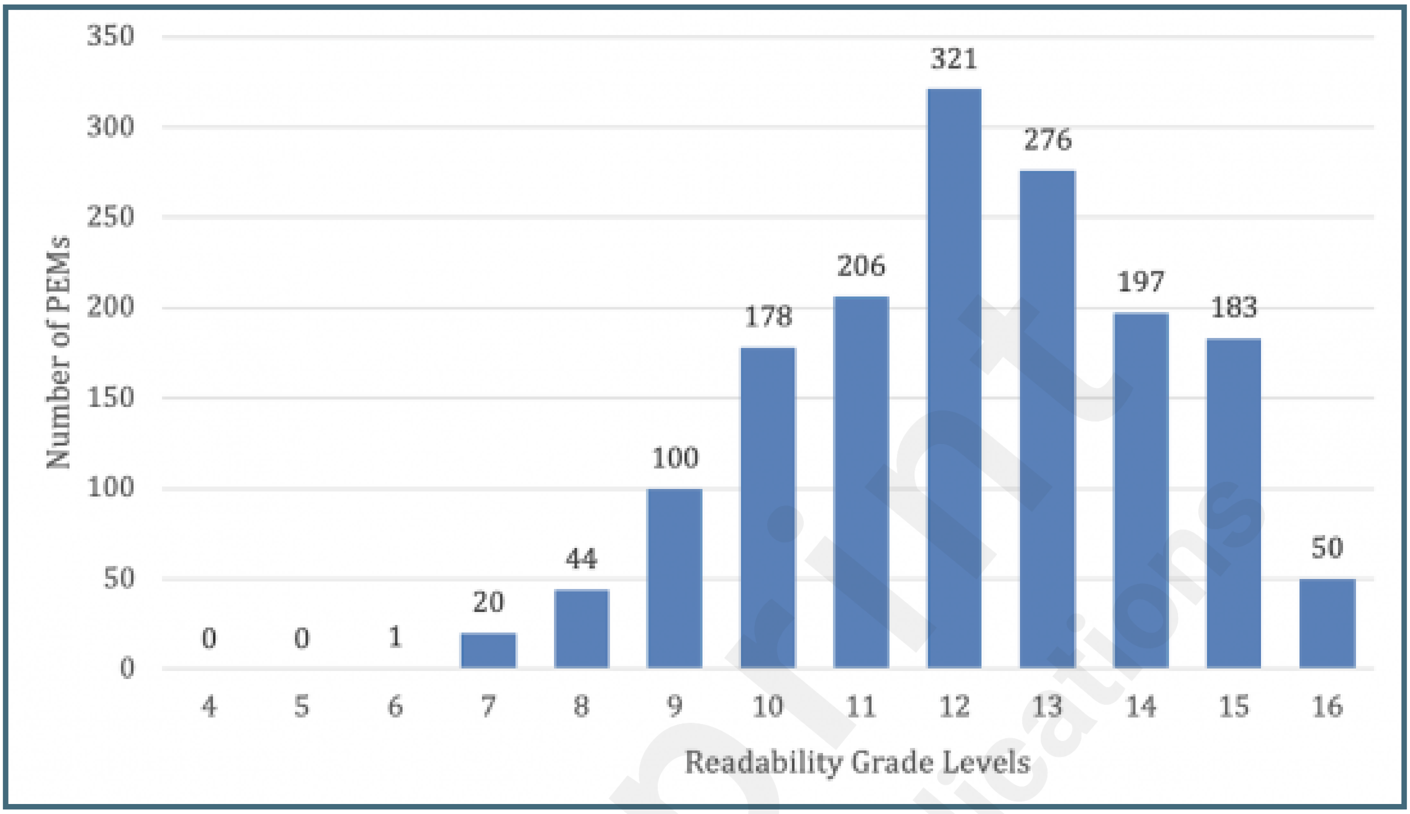
The process and data collection and screening.



The process of data analysis.



The distribution of the composite scores (average of the four TRMs).



Multimedia Appendixes

The PEM website source information and categorization.
URL: http://asset.jmir.pub/assets/af0375de4a1fe355c60e0b9553a75c03.docx