Original Paper

The Readability of Electronic Cigarette Health Information and Advice: A Quantitative Analysis of Web-Based Information

Albert Park[1], PhD; Shu-Hong Zhu[2], PhD; Mike Conway[1], PhD

[1]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, United States
[2]Department of Family Medicine & Public Health, University of California, San Diego, La Jolla, CA, United States

Corresponding Author:
Albert Park, PhD
Department of Biomedical Informatics
University of Utah
421 Wakara Way #140
Salt Lake City, UT, 84108
United States
Phone: 1 206 743 7843
Fax: 1 801 581 4297
Email: alpark1216@gmail.com

Abstract

Background: The popularity and use of electronic cigarettes (e-cigarettes) has increased across all demographic groups in recent years. However, little is currently known about the readability of health information and advice aimed at the general public regarding the use of e-cigarettes.

Objective: The objective of our study was to examine the readability of publicly available health information as well as advice on e-cigarettes. We compared information and advice available from US government agencies, nongovernment organizations, English speaking government agencies outside the United States, and for-profit entities.

Methods: A systematic search for health information and advice on e-cigarettes was conducted using search engines. We manually verified search results and converted to plain text for analysis. We then assessed readability of the collected documents using 4 readability metrics followed by pairwise comparisons of groups with adjustment for multiple comparisons.

Results: A total of 54 documents were collected for this study. All 4 readability metrics indicate that all information and advice on e-cigarette use is written at a level higher than that recommended for the general public by National Institutes of Health (NIH) communication guidelines. However, health information and advice written by for-profit entities, many of which were promoting e-cigarettes, were significantly easier to read.

Conclusions: A substantial proportion of potential and current e-cigarette users are likely to have difficulty in fully comprehending Web-based health information regarding e-cigarettes, potentially hindering effective health-seeking behaviors. To comply with NIH communication guidelines, government entities and nongovernment organizations would benefit from improving the readability of e-cigarettes information and advice.

Introduction

The popularity and use of electronic cigarettes (e-cigarettes) has rapidly increased across all demographic groups in recent years [1]. In fact, there is a continuing increase in not only Web-based promotional messages for e-cigarette brands and flavors [2], but also the use of e-cigarettes by non or former-smokers[1] and youth [3]. Despite inconclusive and contested evidence regarding their safety and effectiveness in helping smoking cessation [4,5], many e-cigarette users believe that they have better health, including improved breathing, less coughing, and lesser chance of getting a sore throat when compared with combustible cigarette users [3]. Thus, analyzing readability (ie, how difficult a text is to understand) of easily accessible e-cigarette related health information and advice (EHIA) is a much needed step toward understanding available

EHIA and identifying opportunities to enhance health advice practices for specific target populations.

The Internet has become a prominent source of text-based health information for consumers [6]. Meanwhile, health information is only productive if it is understood by its audience. The average American adults' reading level is estimated to be at the 8th grade [7]. Thus, the US Department of Health and Human Services (HHS) [8] and the National Institutes of Health (NIH) [9] recommend health information to be written at 6th to 7th grade level, which is the expected reading level for age 10 to 13 years in the US education system. These recommendations are made to ensure the understandability of health information and reduce health information deficits in the general population.

A number of studies have investigated the readability of health-related content on the Internet. Across these studies, researchers consistently found empirical evidence that text-based consumer health information resources were too complex for the recommended 6th to 7th grade reading level [8,9]. For instance, smoking education materials [10], warnings on alcohol and tobacco products [11], Web-based patient education materials [12-14], informed consent documents used in clinical trial research [15], government endorsed written action-plan handouts [16], and commercially available health information [17] were found to require higher literacy levels than that recommended by the NIH and HHS. Moreover, health information available from commercially funded sources was significantly more difficult to read than information available from government-funded sources [18]. This complexity often led to comprehension errors [19,20] for average Americans. We believe that this study is the first study that examines the readability of EHIA available on the Internet.

Methods

A systematic search of EHIA was conducted using 3 search engines (ie, Google, Yahoo, and Bing) in January of 2016. We simulated the behavior of general consumers using various combinations of search terms: *advice*, *cig*, *cigarette*, *e*, *electronic*, *health*, and *information*. Then for comparison purposes, we specifically searched for EHIA from various US public health agencies (eg, HHS), other English speaking nations' public health agencies (United Kingdom, Australia, New Zealand, Canada), popular consumer health information sites (eg, WebMD), as well as nongovernment organizations (eg, Wikipedia).

In this study, data was only gathered from the first page of search results for each search engine, as most users rarely investigate past the first page of search results [21], and so our focus with this work is the analysis of the most frequently accessed EHIA, rather than a comprehensive study of all EHIA. We manually verified search results and retained those webpages that included any EHIA. We excluded articles published in peer-reviewed journals since general consumers are unlikely to read them. Any figures, such as pictorial descriptions, were removed and the webpages were converted to plain text for analysis.

Organization types were determined by the affiliations, funding sources, and available classification information for each organization. Several websites had no explicit indication of their affiliations or funding sources. We assumed that they were for-profit entities due to their informational advertising style content. Moreover, several documents formed part of a bigger document (eg, Wikipedia), in which case we only included sections on EHIA in this study (see Multimedia Appendix 1).

To assess readability (ie, the estimated US grade level that is required to comprehend a text), we used *Flesch-Kincaid grade level* [22], *Simple Measure Of Gobbledygook (SMOG) Index* [23], *Coleman and Liau Index* [24], and *automated readability index* [25], which are widely used metrics in previously mentioned readability studies [10-16,18]. To perform the automated readability analysis, we used the open-source Python *textstat* package [26]. In order to increase the reliability of our readability metrics, and given that different readability metrics can generate a range of results, our analysis was based on the mean of the 4 readability metrics. We then conducted pairwise independent sample *t* tests to compare readability scores among different groups (ie, for-profit entities, nongovernment organizations, non-US government entities, the US government, the US government entities written for teens) followed by *P* value adjustments using the prespecified Hommel procedure [27] to adjust for multiple comparison. The research reported in this study was exempted from review by the University of Utah Institutional Review Board (ethics committee) (IRB_00076188).

Results

We collected a total of 54 documents for this study including materials from 27 US government entities (eg, HHS), 10 for-profit entities (eg, Consumer Affair), 7 non-US government entities (eg, Ministry of Health New Zealand), 7 nongovernment organizations (eg, Mayo Clinic), and 3 documents that were specifically written for teens by US government entities (eg, National Institute on Drug Abuse).

Complete readability scores for each document are presented in Multimedia Appendix 1. On average, the following grade reading levels (standard error) were required to understand the materials from these organizations (see Multimedia Appendix 2):

- for-profit entities: 10.46 (0.55)
- nongovernment organizations: 14.30 (0.86)
- non-US government entities: 14.44 (0.58)
- the US government: 13.48 (0.33)
- the US government entities written for teens: 10.71 (1.15)

The overall comparisons of different groups are shown in Table 1, and the details of comparison results using individual metrics are available in Multimedia Appendices 3-6. Content from for-profit entities was found to be significantly easier to read when compared with materials from all other entities except for materials written for teens by the US government. The differences among all other groups were not found to be significant (Table 1).

Table 1. Pairwise *t* test using average scores of 4 metrics.

Organization Type	Organization Type	*t* value	*P* value	Adjusted *P* value (Hommel)
Versus for-profit entities	Nongovernment organizations	−3.97	.001	.01
	Non-US government entities	−4.90	<.001	.002
	US government	−4.76	<.001	<.001
	US government (teen)	−0.21	.84	.89
Versus nongovernment organizations	Non-US government entities	−0.14	.84	.89
	US government	1.07	.29	.88
	US government (teen)	2.36	.05	.23
Versus non-US government entities	US government	1.36	.18	.59
	US government (teen)	3.27	.01	.07
Versus US government	US government (teen)	2.64	.01	.08

Discussion

Principal Findings

In this study, we used 4 different readability metrics to evaluate the readability of EHIA from 54 sources gathered on the Internet. All 4 metrics indicate that all located EHIA are written at a higher level than the recommended level for the general public. Moreover, EHIA written by for-profit entities, many of whom were advocating e-cigarettes, were significantly easier to read than materials written by nongovernment organizations, non-US government entities, and the US government. Our results contrast with the results of a previous readability study comparing health information written by commercially funded sources and government-funded sources [18]. However, both studies found that the readability of health information was generally too difficult for the public. One encouraging finding in this study is that materials written specifically for teens by US government entities were easier to read than other materials generated by US government entities aimed at the general population, although the difference was found significant for only 1 metric— *Coleman and Liau Index* (see Multimedia Appendix 5).

Limitations

We recognize various limitations of this study. First, individuals accessing EHIA on the Web may not be representative of the general population. However, given that the Internet has become an increasingly popular resource for gathering health information in recent years [6,28], it is likely that a substantial proportion of those potential and current e-cigarettes users seeking EHIA on the Web would have experienced difficulties in fully comprehending "official" health advice, potentially hindering effective health-seeking behaviors. Second, we acknowledge that readability measures alone may not be a perfect representation of reading level [29]. For instance, EHIA could contain pictorial information, which has been shown to be more effective than text-only messages in conveying health warnings on tobacco packages [30]. In this study, we focused on textual information as text remains the primary medium for health communication and information dissemination on the Internet [31]. Third, we used general purpose readability metrics that measure rudimentary lexical features of text. Although these metrics may not be able to accurately assess the complexity of a text [32], a recent study shows lexical features are more important in estimating readability than the complexity of sentences [33]. Finally, our analysis, although systematic, is not exhaustive. A large number of EHIA exist that were not included in our study. Moreover, we limited our search to English language materials. However, we evaluated materials from key official websites that are easily accessible via widely used search engines.

Conclusions

The results of this study suggest that EHIA generated by the for-profit sector is easier to read than EHIA generated by government entities. In order to comply with communication guidelines of the NIH and HHS, government entities and nongovernment organizations would benefit from improving the readability of EHIA.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Individual scores.

[[PDF File (Adobe PDF File), 85KB](#)-[Multimedia Appendix 1](#)]

Multimedia Appendix 2

Mean (SE) of each organization type.

[[PDF File (Adobe PDF File), 23KB](#)-[Multimedia Appendix 2](#)]

Multimedia Appendix 3

Pairwise *t* test of Flesch Kincaid Grade.

[[PDF File (Adobe PDF File), 15KB](#)-[Multimedia Appendix 3](#)]

Multimedia Appendix 4

Pairwise *t* test of SMOG Index.

[[PDF File (Adobe PDF File), 15KB](#)-[Multimedia Appendix 4](#)]

Multimedia Appendix 5

Pairwise *t* test of Coleman Liau Index.

[[PDF File (Adobe PDF File), 15KB](#)-[Multimedia Appendix 5](#)]

Multimedia Appendix 6

Pairwise *t* test of Automated Readability Index.

[[PDF File (Adobe PDF File), 15KB](#)-[Multimedia Appendix 6](#)]

References

1. McMillen R, Gottlieb M, Shaefer R, Winickoff J, Klein J. Trends in Electronic Cigarette Use Among U.S. Adults: Use is Increasing in Both Smokers and Nonsmokers. Nicotine Tob Res. Oct 2015;17(10):1195-1202. [doi: [10.1093/ntr/ntu213](#)] [Medline: [25381306](#)]
2. Zhu SH, Sun J, Bonnevie E, Cummins S, Gamst A, Yin L, et al. Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. Tob Control. Jul 2014;23 Suppl 3:iii3-iii9. [[FREE Full text](#)] [doi: [10.1136/tobaccocontrol-2014-051670](#)] [Medline: [24935895](#)]
3. Pepper JK, Brewer NT. Electronic nicotine delivery system (electronic cigarette) awareness, use, reactions and beliefs: a systematic review. Tob Control. Sep 2014;23(5):375-384. [[FREE Full text](#)] [doi: [10.1136/tobaccocontrol-2013-051122](#)] [Medline: [24259045](#)]
4. Henningfield JE, Zaatari GS. Electronic nicotine delivery systems: emerging science foundation for policy. Tob Control. Apr 2010;19(2):89-90. [doi: [10.1136/tc.2009.035279](#)] [Medline: [20378582](#)]
5. Etter J, Bullen C, Flouris A, Laugesen M, Eissenberg T. Electronic nicotine delivery systems: a research agenda. Tob Control. May 1, 2011;20(3):243-248. [[FREE Full text](#)] [doi: [10.1136/tc.2010.042168](#)] [Medline: [21415064](#)]
6. Fox S. URL: [http://www.pewinternet.org/files/old-media/Files/Reports/2011/PIP_Health_Topics.pdf](#)[WebCite Cache ID 6gbUGjFS5]
7. Cotugna N, Vickery C, Carpenter-Haefele K. Evaluation of literacy level of patient education pages in health-related journals. J Commun Health. Jun 2005;30(3):213-219. [doi: [10.1007/s10900-004-1959-x](#)] [Medline: [15847246](#)]
8. Walsh TM, Volsko TA. Readability assessment of internet-based consumer health information. Respir Care. Oct 2008;53(10):1310-1315. [[FREE Full text](#)] [Medline: [18811992](#)]
9. National Institutes of Health. URL: [https://www.nlm.nih.gov/medlineplus/etr.html](#) [accessed 2016-01-29] [WebCite Cache ID 6etJOZbZM]
10. Meade CD, Byrd JC. Patient literacy and the readability of smoking education literature. Am J Public Health. Feb 1989;79(2):204-206. [Medline: [2913844](#)]
11. Malouff J, Gabrilowitz D, Schutte N. Readability of health warnings on alcohol and tobacco products. Am J Public Health. Mar 1992;82(3):464. [Medline: [1536374](#)]

XSL•FO
RenderX

12. Tian C, Champlin S, Mackert M, Lazard A, Agrawal D. Readability, suitability, and health content assessment of web-based patient education materials on colorectal cancer screening. Gastrointest Endosc. Aug 2014;80(2):284-290. [doi: 10.1016/j.gie.2014.01.034] [Medline: 24674352]

13. D'Alessandro DM, Kingsley P, Johnson-West J. The Readability of Pediatric Patient Education Materials on the World Wide Web. Arch Pediatr Adolesc Med. Jul 01, 2001;155(7):807-812. [doi: 10.1001/archpedi.155.7.807] [Medline: 11434848]

14. Agarwal N, Hansberry DR, Sabourin V, Tomei KL, Prestigiacomo CJ. A comparative analysis of the quality of patient education materials from medical specialties. JAMA Intern Med. Jul 08, 2013;173(13):1257-1259. [doi: 10.1001/jamainternmed.2013.6060] [Medline: 23689468]

15. Terblanche M. Examining the readability of patient-informed consent forms. Open Access J Clin Trials. Oct 19, 2010;2010:2:157-162. [FREE Full text] [doi: 10.2147/OAJCT.S13608]

16. Yin H, Gupta R, Tomopoulos S, Wolf M, Mendelsohn A, Antler L, et al. Readability, suitability, and characteristics of asthma action plans: examination of factors that may impair understanding. Pediatrics. Jan 2013;131(1):e116-e126. [FREE Full text] [doi: 10.1542/peds.2012-0612] [Medline: 23209106]

17. Davis TC, Mayeaux EJ, Fredrickson D, Bocchini JA, Jackson RH, Murphy PW. Reading ability of parents compared with reading level of pediatric patient education materials. Pediatrics. Mar 1994;93(3):460-468. [Medline: 8115206]

18. 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. Apr 18, 2012;17(9):1003-1010. [doi: 10.1080/10810730.2011.650823] [Medline: 22512714]

19. Keselman A, Smith C. A classification of errors in lay comprehension of medical documents. J Biomed Inform. Dec 2012;45(6):1151-1163. [FREE Full text] [doi: 10.1016/j.jbi.2012.07.012] [Medline: 22925723]

20. Smith CA, Hetzel S, Dalrymple P, Keselman A. Beyond readability: investigating coherence of clinical text for consumers. J Med Internet Res. Dec 02, 2011;13(4):e104. [FREE Full text] [doi: 10.2196/jmir.1842] [Medline: 22138127]

21. van Deursen AJ, van Dijk JA. Using the Internet: Skill related problems in users' online behavior. Interact Comput. Dec 13, 2009;21(5-6):393-402. [FREE Full text] [doi: 10.1016/j.intcom.2009.06.005]

22. Kincaid J, Fishburne R, Rogers R, Chissom B. Derivation of new readability formulas (automated readability index, fog count, and Flesch reading ease formula) for navy enlisted personnel. In: Research Branch Report. Millington, TN. Chief of Naval Technical Training; 1975:8-75.

23. McLaughlin GH. SMOG grading: A new readability formula. J Reading. 1969;12(8):639-646. [FREE Full text]

24. Coleman M, Liau T. A computer readability formula designed for machine scoring. J Appl Psychol. 1975;60(2):283-284. [doi: 10.1037/h0076540]

25. Smith EA, Senter RJ. Automated readability index. AMRL TR. May 1967:1-14. [Medline: 5302480]

26. Bansal S, Aggarwal C. URL: https://pypi.python.org/pypi/textstat [accessed 2016-09-23] [WebCite Cache ID 6kjFlWIwD]

27. Wright S. Adjusted P-Values for Simultaneous Inference. Biometrics. Dec 1992;48(4):1005-1006. [doi: 10.2307/2532694]

28. Fox S, Duggan M. URL: http://www.pewinternet.org/2013/01/15/health-online-2013/[WebCite Cache ID 6Y5K9a3JH]

29. Meade C, Smith C. Readability formulas: Cautions and criteria. Patient Educ Couns. Apr 1991;17(2):153-158. [FREE Full text] [doi: 10.1016/0738-3991(91)90017-Y]

30. Hammond D. Health warning messages on tobacco products: a review. Tob Control. Sep 2011;20(5):327-337. [doi: 10.1136/tc.2010.037630] [Medline: 21606180]

31. Himelboim I, McCreery S. New technology, old practices: Examining news websites from a professional perspective. Convergence: The International Journal of Research into New Media Technologies. Feb 07, 2012;18(4):427-444. [FREE Full text] [doi: 10.1177/1354856511429648]

32. Davison A, Kantor R. On the failure of readability formulas to define readable texts: a case study from adaptations. Read Res Q. 1982;17(2):187-209. [FREE Full text] [doi: 10.2307/747483]

33. Kang T, Elhadad N, Weng C. Initial Readability Assessment of Clinical Trial Eligibility Criteria. AMIA Annu Symp Proc. 2015;2015:687-696. [FREE Full text] [Medline: 26958204]

Abbreviations

e-cigarettes: electronic cigarettes
EHIA:: e-cigarette related health information and advice
HHS: US Department of Health and Human Services
NIH: National Institutes of Health
SMOG: Simple Measure of Gobbledygook

XSL•FO
RenderX