5-23-2022

Exploring the Readability of Ingredients Lists of Food Labels with Existing Metrics

Kathryn Dempsey Cooper
University of Nebraska at Omaha, kdempsey@unomaha.edu

William Gasper

Ricky Flores
rickyflores@unomaha.edu

Martina Clarke
University of Nebraska Medical Center

A. Erin Bass
University of Nebraska at Omaha, aebass@unomaha.edu

See next page for additional authors

Exploring the Readability of Ingredients Lists of Food Labels with Existing Metrics

Kathryn Cooper, PhD[1*], William Gasper, BS[1] , Ricky Flores, BS[1], Martina Clarke, MPH, PhD[1], Erin Bass, MBA, PhD, Leslie Evans, MS, RD, LMNT[2], Jana Ponce, PhD, RD, LMNT[3]

[1]University of Nebraska at Omaha, Omaha, NE, USA; [2]Nebraska Medicine, Omaha, NE, USA, [3]University of Nebraska Medical Center, Omaha, NE, USA

Abstract

Healthy diet and dietary behaviors are key components in prevention of chronic disease and management of chronic illness. Nutritional literacy has been associated with dietary behaviors and consumer choice of healthy foods. Nutritional literacy can be measured, for example, by examining consumer food label use, but current research focuses largely on the Nutrition Facts panel of a food product. Ingredients lists are critical for communicating food composition but are relatively unstudied in existing literature. The goal of this work is to measure the readability of ingredients lists on branded food products in the United States using existing metrics. We examined ingredients lists for all 495,646 products listed in the USDA Food Data Central database using four existing readability measures for text written in natural language. Each of these indices approximates the grade level that would be expected to comprehend a text; comparatively, patient consent forms are considered acceptable at an 8th grade reading level or lower. We report a broad variability for in readability using different metrics: ingredients lists recorded at a 9th grade reading level or higher to comprehend are found at rates of 16.5% (Automated Reading Index) to 74.9% (Gunning-Fog Index). Ingredients lists recorded at a 10th grade reading level or higher to comprehend are found at rates of 84.2% (using FRE Index). These results demonstrate the need to further explore how ingredients lists can be measured for readability, both for the purposes of consumer understanding as well as for supporting future nutrition research involving text mining.

Introduction

Nutrition is a known factor in the prevention and management of chronic diseases and impacts both short- and long-term health outcomes. Food label use is an important facet for many consumers when making nutritional choices for themselves and their families. In the United States, a food product label or food label is required to have 5 components, including: (1) product name or identity, (2) the package contents, (3) contact information for the food manufacturer, (4) the Nutrition Facts panel (NFP), and (5) a list of ingredients sorted in order of decreasing composite weight of the product.

Over 75% of consumers report that they read the ingredients list on the product label "sometimes" or "often" when they purchase a product for the first time per the 2019 Food Safety and Nutrition Survey Report (FSNSR). One major reason consumers read food labels is to identify unfavorable and potentially harmful ingredients. For example, the 2019 FSNSR indicates that consumers largely want to detect the presence of artificial ingredients in a food product: 47% of consumers surveyed were either very or extremely concerned about the use of artificial ingredients. Other reasons for focus on the ingredients list is to avoid or reduce intake of certain nutrients or ingredients, such as those with food allergies, individuals with cardiac disease managing their sodium, or a parent nursing an infant with food intolerances.

Poor nutrition is one of the key underlying causes of heart disease and improving nutritional literacy has been suggested as a way for individuals with heart disease improve their condition [1]. A study using a Healthy Eating Index (HEI) on US community-dwelling adults aged 70-79, found that 79.7% had a diet ranked as "poor" or "needs improvement" [2]. Approximately 25% of participants in this study later developed malnutrition. A more recent 2019 study found that 19% of US adults reported a food allergy, and around 11% of US adults are reported have a physician-diagnosed food allergy. 51.1% of physician-diagnosed food allergies were classified as severe, with 24% of individuals with physician-diagnosed food allergies having epinephrine pen prescriptions [3]. It is critically important for food labels to be understandable to the consumer at the point of first purchase.

Nutrition literacy can be generally described as the capacity an individual has for gathering and synthesizing nutrition information to make healthy decisions in their daily lives [4]. Higher nutrition literacy is associated with positive health behaviors, including making healthier dietary choices (i.e. eating more fruits and vegetables), and increasing one's daily activity. Nutritional literacy falls under the larger umbrella term of health literacy, or one's ability to make decisions impacting their overall health (Figure 1). Lower health literacy has been demonstrated to be associated with increases in emergency room visits and hospitalization, as well as higher mortality rates, and also with lower rates of preventative screening, medication adherence, label interpretation, and health messaging comprehension [5].

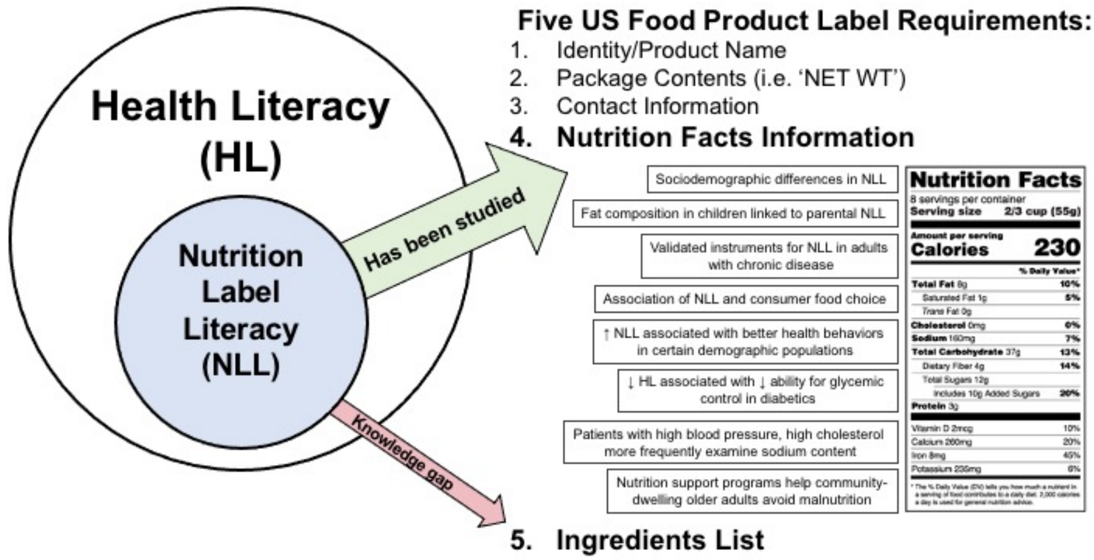

Figure 1. An overview of the relationship between health literacy, nutrition label literacy, and how they are measured against the 5 required components of a food product label in the US.

One important measure of nutritional literacy is usage of the food product label, or the frequency with which one uses a food label. A 2015 review of the effects of nutritional literacy on food label usage found that consumers who are familiar with how to read a food label are more likely to use its information to make healthy decisions [6]. Food label usage predicted dietary quality in 18- to 29-year-old persons (n=103), with increased usage expected to improve health outcomes such as dietary quality [7]. Women are known to be more frequent users of food labels [8], and there is some evidence to suggest that consumers identifying as Hispanic are frequent users as well [9]. Food label use is especially important for those trying to avoid or manage the incorporation of a nutrient or ingredient in their diet: Older adults are also more likely to avoid foods associated with known personal health issues than younger adults [9]. Low food label usage was found to be associated with difficulty following gluten-free diets for individuals with diagnosed Celiac Disease or Gluten Sensitivity [10]. The 2019 FSNSR notes that consumers read the food label upon first purchase, indicating that label readability is an important facet of consumer behavior [11]. Eighty-three percent of consumers checked the ingredient list upon first purchase of a product [11]. Therefore, it is critically important for food labels to be **readable at the point of purchase** so that consumers will be encouraged to use them.

In the United States, a food product label or food label is required to have 5 components, including: (1) product name or identity, (2) the package contents, (3) contact information for the food manufacturer, (4) the Nutrition Facts panel (NFP), and (5) a list of ingredients sorted in order of decreasing composite weight of the product. Consumer understanding of the composite food product label has been found to be correlated with income and education, but even highly educated subjects display difficulty understanding a food label (n=100) [12]. The NFP is definitively the most studied component of a food label in current literature, and was recently updated in 2016 to help consumers better understand the dietary implications of a product. Recent studies on the NFP have found evidence that supports a relationship between NFP use and positive dietary behaviors in young adults [8], in prediabetic adults [13], and in Latinx adults diagnosed with Type II diabetes [14]. However, readability of ingredient lists on a food label has not been extensively studied. The main purpose of this work is to explore the way ingredients are currently presented to consumers on packaged food products in the United States using existing measures of readability.

Methods

A flow diagram of the overall approach used in this method is shown in **Error! Reference source not found.**. The October 2020 release of the Branded Foods dataset from the USDA Food Data Central database [9] was downloaded via https://fdc.nal.usda.gov/download-datasets.html as a .CSV file. Information for a total of 498,182 products is contained in the Branded Foods dataset. Ingredients lists that were empty (n = 2,536, 0.51%) were removed, leaving a total of 495,646 food products to be analyzed. Ingredients lists for these products were pulled from the CSV file, tokenized, and analyzed using the readability [1] and Natural Language Toolkit [2] libraries with Python version 3.6.10. The full process and Python scripts used to perform these tasks, along with data availability, can be found at https://github.com/kmcooper/il_readability_existing_measures. Ingredients lists were unaltered other than tokenization for the purpose of readability analysis. The goal was to analyze the readability of an ingredients list as it would be observed by the consumer without modification.

Readability Measures

Readability measures are used to determine the difficulty with which one can expect when reading a selection of text written in natural language. A number of readability metrics exist already; most of these incorporate factors such as word length, word complexity, number of syllables, text length, and others to determine readability, and are applied to documents such as manuals, textbooks, and patient consent forms, where an individual would be reading the document for comprehension. For context, in healthcare it is generally accepted that documents should be written at an 8th grade reading level or lower to be considered acceptable for documents needing to be written in plain language, such as consent forms [15].

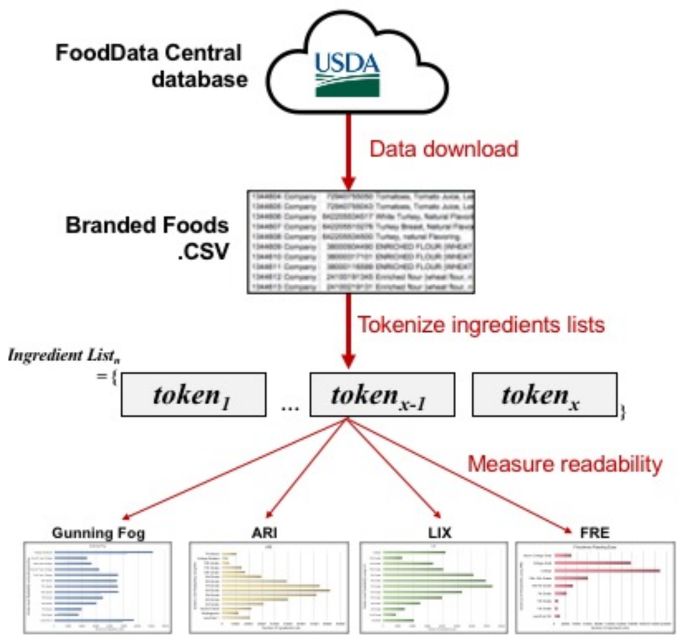

Figure 2. A high level overview of the methods used to measure readability. All preprocessing and analysis was performed using GNU bash and Python 3.6.10.

In this analysis we examined ingredients lists of 496,646 distinct food products listed in the Branded Foods dataset from the USDA FoodData central database using four existing readability measures: the Flesh-Kincaid Reading Ease, Gunning-Fog, LIX, and the Automated Readability Index (ARI). Each of these indices approximates the grade level that would be expected to comprehend a text. After readability measures were calculated for each ingredients list, metrics were aggregated into grade levels according to the scale given by each measure; for example, the ARI measure ranks from Kindergarten to College. Each of those ranks are described below.

The Automated Readability Index (ARI): The Automated Readability Index or ARI was developed in 1967 by Smith and Senter [16]and has been used to measure readability of online websites for consumers on topics related to epilepsy[17], otolaryngology [18], breast lesions [19], hip surgery [20], and privacy policies [21]. The ARI measure is based on the number of words per sentence in text written in natural language, as well as characters per word to approximate word complexity [16]. The ARI index reports measures that align with readability from Kindergarten through College Students and Professorial Levels [16].

The Flesch-Kincaid Reading Ease Index (FRE): The Flesch-Kincaid Reading Ease Index, or FRE, was developed in 1975 to help the US Navy author technical documents and ranks texts from 5th grade up through College levels (College, College Graduate, and above) [22]. It is a popular index for measuring the readability of consent forms, with notable use in validating consent form reading levels for vulnerable individuals [23], for those participating in DNA sequencing analyses [24], and HIPAA compliant consenting materials [15], [25]-[27].

The Lasbarhetsindex Swedish Readability Formula (LIX): The Lasbarhetsindex Swedish Readability formula, or LIX, was developed in 1983 for measuring readability of newspapers and is applied to rank texts from 1st grade up through 12th grade, in addition to a final College reading level [28]. The LIX score is based on the number of words, and periods in a text, as well as word length (i.e. a "long word" is sometimes defined as one having 6 or more characters). It is noted for its applicability in readability analyses that are not necessarily rooted in syllable count [29] and has been used to measure readability in consent forms and online health information [30], [31], similar to other measures presented here.

***The Gunning-Fog Readability Index*:** The Gunning-Fog Readability formula was developed in 1952 for measuring readability of newspapers upon first exposure to the text and is applied to ranks texts from 6th grade up through College Graduate level [32]. The Gunning-Fog index is based on word count, sentence count, and complex word count in a text, where complex words contain >= 3 syllables. Similar to the other measures reported here, it is a popular measure of readability for health information presented to consumers and patients alike [33]-[37].

Mean and median scores as well as standard deviation for all readability metrics are automatically reported by the *readability* Python library. Conversion of readability scores to grade level is metric-specific and was performed using Python. Code to perform conversions for the measures reported here is also available on our Github repository for this project at https://github.com/kmcooper/il_readability_existing_measures.

Results

The mean and median scores for the four readability metrics used in our analyses are reported in **Table 1.** Mean, median scores for the 4 readability metrics, including mean reading level according to each metric, and the percent total of 495,696 ingredients lists analyzed falling at or below an 8th or 9th grade reading level.Table 1. Broadly, the median score reported by the four indices ranges between the 6th grade reading level and a college reading level, with little consensus between metrics. The Automated Reading Index (ARI) is by far the most magnanimous measure, reporting that ingredients lists for 83.5% products fall at or below an 8th grade reading level. By contrast, the Flesch-Kincaid Reading Ease metric is by far the most conservative measure, reporting that ingredients lists for 84.23% of products fall at or below a 10th grade reading level (FRE groups 8th and 9th grade together). At large, three of the four measures reported describe a mean reading level of 8th grade or higher for ingredients lists on branded foods.

Table 1. Mean, median scores for the 4 readability metrics, including mean reading level according to each metric, and the percent total of 495,696 ingredients lists analyzed falling at or below an 8th or 9th grade reading level.

	Mean Score	Mean Reading Level	Median Score	Std. Dev	% of IL at/below 8th Grade Level	% of IL above 8th Grade Level
Gunning-Fog Index	11.983	11-12th Grade	11.916	5.611	25.14%	74.86%
ARI	6.539	6th-7th Grade	6.474	3.357	83.50%	16.50%
LIX	37.295	8th Grade	36.694	15.558	61.57%	38.43%
	Mean Score	**Mean Reading Level**	**Median Score**	**Std. Dev**	**% of IL at/below 9th Grade Level**	**% of IL above 10th Grade Level**
Flesch- Kincaid Reading Ease	37.931	College Level	37.026	26.889	15.77%	84.23%

The reading level distributions for each of the reported metrics are shown below in Figures 2-6. These distributions are presented as a total count of ingredients lists reported at each reading level using the given metric. It is not effective to compare metrics by their distributions due to their different reading level categorizations, however, visualizing their distributions in this way highlights the disagreement between metrics. For example, both the LIX and the Gunning-Fog indices capture ingredients lists at higher-than-normal-levels at the extreme ends of their distributions, and the Flesch-Kincaid Reading Ease is skewed toward the college ranked reading levels.

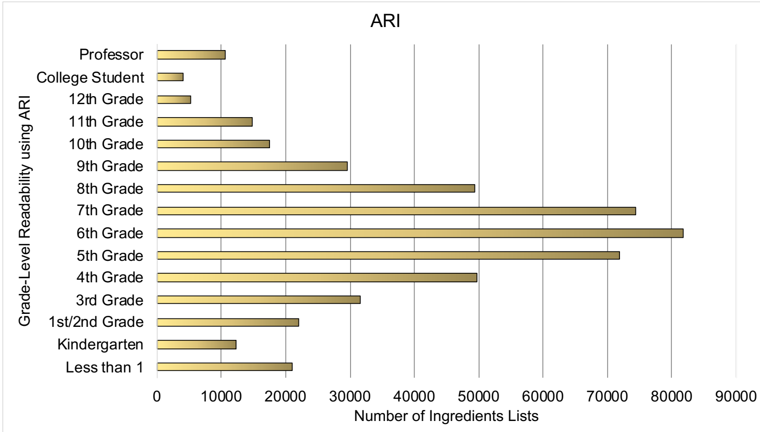

Figure 3. Ingredient list reading level distribution using the ARI measure. All ILs were analyzed using the ARI measure and categorized into grade levels between Kindergarten and Professor using the ARI. According to this measure, 83.5% of ingredients lists are presented at an 8th grade or lower reading level.

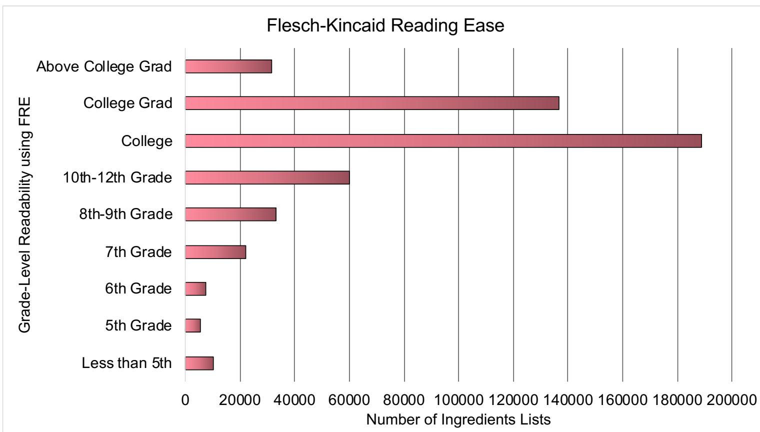

Figure 4. Ingredient list reading level distribution assessed using the FRE measure. All ILs were analyzed using the FRE measure and categorized into grade levels between 5th Grade and College Graduate. According to this measure, 15.8% of ingredients lists are presented at an 10th grade or lower reading level.

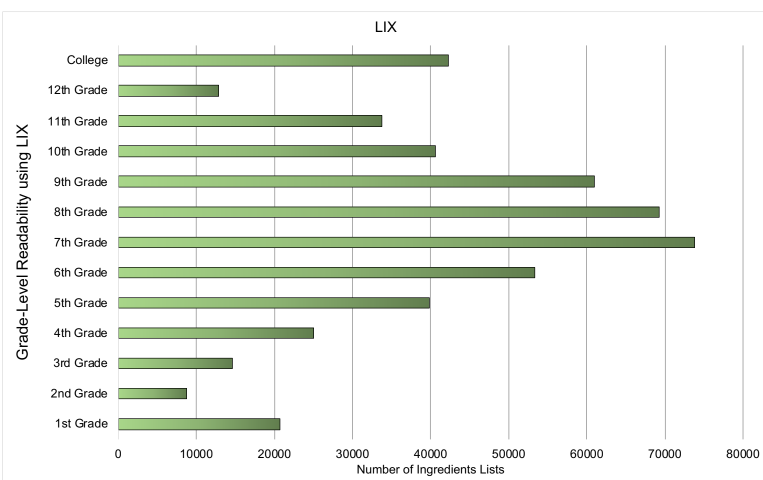

Figure 5. Ingredient list reading level distribution assessed using the LIX measure. All ILs were analyzed using the LIX measure and categorized into grade levels between 1st Grade and College using the LIX. Using LIX, 61.6% of ILs are presented at an 8th grade or lower reading level.

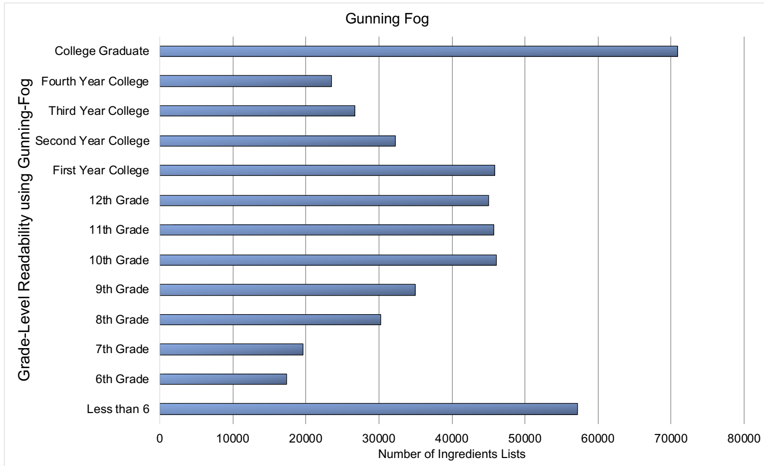

Figure 6. Ingredient list reading level distribution assessed using Gunning-Fog measure (FOG). ILs were analyzed using FOG and categorized into grade levels between 6th Grade and College Graduate. According to this measure, 25.1% of ILs are at an 8th grade or lower reading level.

Discussion

Specifically, this work examines the readability of ingredients lists using existing readability indices, where readability can be generally defined to measure the level of education needed for an individual to understand a list of ingredients. There are limited studies on the readability of food products labels and few studies examining the readability of other commercial products, such as cosmetics [38] and dietary supplements [39]. This study highlights the documented challenges of analyzing the readability of ingredients lists using existing metrics; generally, text mining challenges such as these are already somewhat recognized in food composition research and information systems [40]-[43]. The disagreement between these measures used in this research suggests that existing readability metrics may not be sufficient to infer readability of ingredients lists, which are not traditionally written as natural language. Better metrics for measuring readability of ingredients lists can be developed to more accurately reflect the text structure of ingredients lists versus text written in natural language.

Another challenge directly impacting the consumer is that food production companies differ, sometimes widely, in the terms they use to present ingredients on a food product label, the preparation of those ingredients, and the purpose for which an ingredient is used. For example, soy lecithin is sometimes used as an emulsifier, in other foods it is used as a flavor protection agent. Additionally, a label might say “chocolate”, “milk chocolate”, or “chopped chocolate”, which could all be the same product, but prepared and or presented in different ways. A 2017 study on nutrition modeling recommended that food labels should denote (1) the ingredients in the food product itself and (2) how the ingredient was *prepared* to be used in the product, such as “chopped”, “raw”, or pureed” [44]. There is building evidence that food preparation affects the gut microbiome[45], [46], which has implications for health outcomes. As the body of research in consumer access and use of nutritional information continues to grow, it is expected that there will be consumer-demand for information on ingredient preparation and provenance in food labeling policy. A 2014 review on challenges facing food science acknowledges the need for multidisciplinary teams to address these and other challenges, incorporating the fields of computer science, text mining, and informatics [47]. Resources and interdisciplinary teams are necessary to create consumer-centered information systems.

Conclusions

More research is needed to fully understand the consumer experience with ingredients lists on the food products label. With the digitization and aggregation of information on food products made and distributed in the United States, it is possible to apply informatics approaches that will support consumers in the pursuit of healthy dietary behaviors, like natural language processing. The study of the readability and consumer experience of ingredients lists can reveal insights into how consumers use and experience food products in their daily lives, but remains an under-utilized tool for improving health.

Limitations

This work uses existing metrics for measuring readability but it is not readily clear when the application of readability metrics becomes inappropriate, such as in the application of the metrics to a list of terms. There are a number of readability metrics that were not used in this work because they count sentences and punctuation typically used in narrative text [48, 49] that would not be appropriate in application to an ingredients list. The four metrics used here were chosen as they focus largely on word and character counts which are applicable in this context, however, future work could focus on the development of a readability metric specific to ingredients lists.

Acknowledgements

This work was supported with a Graduate Research and Creative Activity (GRACA) grant from the University of Nebraska at Omaha as well as a pilot award from the Business, Informatics, and Gerontology Big Ideas Award.

References

1. (October 7,). *Heart Disease and Stroke*. Available: https://www.cdc.gov/chronicdisease/resources/publications/factsheets/heart-disease-stroke.htm.
2. L. M. Hengeveld *et al*, "Prospective associations of poor diet quality with long-term incidence of protein-energy malnutrition in community-dwelling older adults: the Health, Aging, and Body Composition (Health ABC) Study," *Am. J. Clin. Nutr.,* vol. 107, *(2),* pp. 155-164, 2018.
3. R. S. Gupta *et al*, "Prevalence and severity of food allergies among US adults," *JAMA Network Open,* vol. 2, *(1),* pp. e185630, 2019.
4. K. J. Silk *et al*, "Increasing nutrition literacy: testing the effectiveness of print, web site, and game modalities," *Journal of Nutrition Education and Behavior,* vol. 40, *(1),* pp. 3-10, 2008.
5. N. D. Berkman *et al*, "Low health literacy and health outcomes: an updated systematic review," *Ann. Intern. Med.,* vol. 155, *(2),* pp. 97-107, 2011.
6. L. M. S. Miller and D. L. Cassady, "The effects of nutrition knowledge on food label use. A review of the literature," *Appetite,* vol. 92, pp. 207-216, 2015.
7. E. Cha *et al*, "Health literacy, self-efficacy, food label use, and diet in young adults," *Am. J. Health Behav.,* vol. 38, *(3),* pp. 331-339, 2014.
8. M. J. Christoph *et al*, "Nutrition facts panels: who uses them, what do they use, and how does use relate to dietary intake?" *Journal of the Academy of Nutrition and Dietetics,* vol. 118, *(2),* pp. 217-228, 2018.
9. Y. Zhang *et al*, "How Do Consumers Use Nutrition Labels on Food Products in the United States?" *Topics in Clinical Nutrition,* vol. 32, *(2),* pp. 161-171, 2017.
10. L. Verrill, Y. Zhang and R. Kane, "Food label usage and reported difficulty with following a gluten-free diet among individuals in the USA with coeliac disease and those with noncoeliac gluten sensitivity," *Journal of Human Nutrition and Dietetics,* vol. 26, *(5),* pp. 479-487, 2013.
11. A. Lando, L. Verril and F. Wu, "2019 food safety and nutrition survey report," March. 2021.
12. R. L. Rothman *et al*, "Patient understanding of food labels: the role of literacy and numeracy," *Am. J. Prev. Med.,* vol. 31, *(5),* pp. 391-398, 2006.
13. G. Kollannoor-Samuel *et al*, "Nutrition facts panel use is associated with higher diet quality and lower glycated hemoglobin concentrations in US adults with undiagnosed prediabetes," *Am. J. Clin. Nutr.,* vol. 104, *(6),* pp. 1639-1646, 2016.
14. G. Kollannoor-Samuel *et al*, "Nutrition Facts Panel use is associated with diet quality and dietary patterns among Latinos with type 2 diabetes," *Public Health Nutr.,* vol. 20, *(16),* pp. 2909-2919, 2017.
15. M. K. Paasche-Orlow, H. A. Taylor and F. L. Brancati, "Readability standards for informed-consent forms as compared with actual readability," *N. Engl. J. Med.,* vol. 348, *(8),* pp. 721-726, 2003.
16. E. A. Smith and R. J. Senter, "Automated readability index." *AMRL-TR.Aerospace Medical Research Laboratories (US),* pp. 1-14, 1967.
17. F. Brigo *et al*, "Clearly written, easily comprehended? The readability of websites providing information on epilepsy," *Epilepsy & Behavior,* vol. 44, pp. 35-39, 2015.
18. K. Wong and J. R. Levi, "Readability trends of online information by the American Academy of Otolaryngology—Head and Neck Surgery Foundation," *Otolaryngology–Head and Neck Surgery,* vol. 156, *(1),* pp. 96-102, 2017.
19. R. C. Miles *et al*, "Readability of online patient educational materials related to breast lesions requiring surgery," *Radiology,* vol. 291, *(1),* pp. 112-118, 2019.

20. M. P. Mehta *et al*, "Assessing the readability of online information about hip arthroscopy," *Arthroscopy: The Journal of Arthroscopic & Related Surgery,* vol. 34, *(7),* pp. 2142-2149, 2018.
21. T. Ermakova, H. Krasnova and B. Fabian, "EXPLORING THE IMPACT OF READABILITY OF PRIVACY POLICIES ON USERS'TRUST," 2016.
22. J. P. Kincaid *et al*, "No title," *Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel,* 1975.
23. R. L. Sudore *et al*, "Use of a modified informed consent process among vulnerable patients: a descriptive study," *Journal of General Internal Medicine,* vol. 21, *(8),* pp. 867-873, 2006.
24. E. Niemiec *et al*, "Readability of informed consent forms for whole-exome and whole-genome sequencing," *Journal of Community Genetics,* vol. 9, *(2),* pp. 143-151, 2018.
25. P. Fortun *et al*, "Recall of informed consent information by healthy volunteers in clinical trials," *QJM: An International Journal of Medicine,* vol. 101, *(8),* pp. 625-629, 2008.
26. P. Breese and W. Burman, "Readability of notice of privacy forms used by major health care institutions," *Jama,* vol. 293, *(13),* pp. 1588-1594, 2005.
27. H. J. Silverman *et al*, "Recommendations for informed consent forms for critical care clinical trials," *Crit. Care Med.,* vol. 33, *(4),* pp. 867-882, 2005.
28. C. Björnsson, "Readability of newspapers in 11 languages," *Reading Research Quarterly,* pp. 480-497, 1983.
29. M. Hochhauser, "Consent form readability in the SUPPORT study," *Journal of Clinical Research Best Practices,* vol. 9, *(6),* pp. 1-5, 2013.
30. Y. Liu *et al*, "Combining Readability Formulas and Machine Learning for Reader-oriented Evaluation of Online Health Resources," *IEEE Access,* 2021.
31. J. Palotti, G. Zuccon and A. Hanbury, "Consumer health search on the web: study of web page understandability and its integration in ranking algorithms," *Journal of Medical Internet Research,* vol. 21, *(1),* pp. e10986, 2019.
32. R. Gunning, "Technique of clear writing," 1952.
33. D. R. Hansberry, N. Agarwal and S. R. Baker, "Health literacy and online educational resources: an opportunity to educate patients," *Am. J. Roentgenol.,* vol. 204, *(1),* pp. 111-116, 2015.
34. M. I. Cajita *et al*, "Quality and health literacy demand of online heart failure information," *J. Cardiovasc. Nurs.,* vol. 32, *(2),* pp. 156, 2017.
35. G. Zuccon *et al*, "The IR Task at the CLEF eHealth evaluation lab 2016: user-centred health information retrieval," 2016.
36. A. Park and M. Conway, "Harnessing Reddit to understand the written-communication challenges experienced by individuals with mental health disorders: analysis of texts from mental health communities," *Journal of Medical Internet Research,* vol. 20, *(4),* pp. e8219, 2018.
37. K. Roberts and D. Demner-Fushman, "Interactive use of online health resources: a comparison of consumer and professional questions," *Journal of the American Medical Informatics Association,* vol. 23, *(4),* pp. 802-811, 2016.
38. K. Yazar *et al*, "Readability of product ingredient labels can be improved by simple means: an experimental study," *Contact Derm.,* vol. 71, *(4),* pp. 233-241, 2014.
39. K. A. Clauson, Q. Zeng-Treitler and S. Kandula, "Readability of patient and health care professional targeted dietary supplement leaflets used for diabetes and chronic fatigue syndrome," *The Journal of Alternative and Complementary Medicine,* vol. 16, *(1),* pp. 119-124, 2010.
40. M. Franz and L. Sampson, "Challenges in developing a whole grain database: Definitions, methods and quantification," *Journal of Food Composition and Analysis,* vol. 19, pp. S38-S44, 2006.
41. G. Ros, E. M. De Victoria and A. Farran, "Spanish food composition database: A challenge for a consensus," *Food Chem.,* vol. 113, *(3),* pp. 789-794, 2009.
42. J. T. Spence, "Challenges related to the composition of functional foods," *Journal of Food Composition and Analysis,* vol. 19, pp. S4-S6, 2006.
43. S. Westenbrink, K. Brunt and J. van der Kamp, "Dietary fibre: Challenges in production and use of food composition data," *Food Chem.,* vol. 140, *(3),* pp. 562-567, 2013.
44. H. Schfer *et al*, "User nutrition modelling and recommendation: Balancing simplicity and complexity," in *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization,* 2017, pp. 93-96.
45. R. N. Carmody *et al*, "Cooking shapes the structure and function of the gut microbiome," *Nature Microbiology,* vol. 4, *(12),* pp. 2052-2063, 2019.

46. M. Kreuzer and W. Hardt, "How Food Affects Colonization Resistance Against Enteropathogenic Bacteria," *Annu. Rev. Microbiol.,* vol. 74, pp. 787-813, 2020.
47. H. G. Van Mil *et al*, "A complex system approach to address world challenges in food and agriculture," *Trends Food Sci. Technol.,* vol. 40, *(1),* pp. 20-32, 2014.
48. Zhou, S., Jeong, H., & Green, P. A. (2017). How consistent are the best-known readability equations in estimating the readability of design standards?. IEEE Transactions on Professional Communication, 60(1), 97-111.
49. Redish, J. C. (1981). Understanding the limitations of readability formulas. IEEE transactions on professional communication, (1), 46-48.