

# Online Readability of COVID-19 Health Information

Amy P Worrall (✉ [worralap@tcd.ie](mailto:worralap@tcd.ie))

Beaumont Hospital <https://orcid.org/0000-0002-7417-1844>

Mary J Connolly

Beaumont Hospital

Aine O'Neill

Beaumont Hospital

Murray O'Doherty

Beaumont Hospital

Kenneth P Thornton

Beaumont Hospital

Cora McNally

Beaumont Hospital

Samuel J McConkey

Beaumont Hospital

Eoghan de Barra

Beaumont Hospital

---

## Research article

**Keywords:** COVID-19, coronavirus pandemic, health information, readability, health literacy

**Posted Date:** May 27th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-30124/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on November 13th, 2020. See the published version at <https://doi.org/10.1186/s12889-020-09710-5>.

# Abstract

**Introduction:** The internet is now the first line source of health information for many people worldwide. In the current Coronavirus Disease 2019 (COVID-19) global pandemic, health information is being produced, revised, updated and disseminated at an increasingly rapid rate. The general public are faced with a plethora of misinformation regarding COVID-19 and the readability of online information has an impact on their understanding of the disease. The accessibility of online healthcare information relating to COVID-19 is unknown.

**Methods:** The Google® search engine was used to collate the first twenty webpage URLs for three individual searches for 'COVID', 'COVID-19', and 'coronavirus' from Ireland, the United Kingdom, Canada and the United States. The Gunning Fog Index (GFI), Flesch-Kincaid Grade (FKG) Score, Flesch Reading Ease Score (FRES), Simple Measure of Gobbledygook (SMOG) score were calculated to assess the readability.

**Results:** There were poor levels of readability webpages reviewed, with only 17.2% of webpages at a universally readable level. There was a significant difference in readability between the different webpages based on their information source ( $p < 0.01$ ). Public Health organisations and Government organisations provided the most readable COVID-19 material, while digital media sources were significantly less readable. There were no significant differences in readability between regions.

**Conclusion:** Much of the general public have relied on online information during the pandemic. Information on COVID-19 should be made more readable, and those writing webpages and information tools should ensure universal accessibility is considered in their production. Governments and healthcare practitioners should have an awareness of the online sources of information available, and ensure that readability of our own productions is at a universally readable level which will increase understanding and adherence to health guidelines.

## Introduction

The Coronavirus Disease 2019 (COVID-19) pandemic has led to an expected increase in the level of online searches on the condition. Internet users are now frequently searching for health related information and as a tool to answer questions about symptoms, diagnoses and treatment<sup>1</sup>. Social distancing, lockdowns and self-isolation policies worldwide have also meant patients' access to their primary care has decreased and reliance on online information has increased. This is reflected in the rise in Google® Trends searches for 'coronavirus', 'COVID' and 'COVID-19' in recent months<sup>2</sup>.

The internet as a source of health information is unregulated and the quality, reliability, and accessibility to the reader is variable. Many webpages provide inaccurate or questionable information and this can be harmful<sup>3</sup>. A small number of studies have already reported on the quality of COVID-19 related health information<sup>4</sup>, and indeed the misinformation that has appeared on webpages and in particular on social

media in recent months<sup>3,5</sup>. The quality of information relating to COVID-19 assessed found that there are often discrepancies between health information issued by public health organisation and general information available on other digital media<sup>6</sup>.

Several tools are available to assess the readability of information, such as the Gunning Fox Index (GFI), the Flesch Reading Ease Score (FRES), the Flesch-Kincaid Grade (FKG) and the Simple Measure of Gobbledygook (SMOG) score<sup>7</sup>. These tools are established validated readability tools and are validated in health information studies and the English language<sup>8</sup>. The readability of health information related to COVID-19 has not been published. We sought to evaluate the readability of online information relating to COVID-19 in four English speaking regions: Ireland, the United Kingdom, Canada and the United States.

## Methods

### *Webpage Search and Identification*

The Google search engine was used to collate the first twenty webpage URLs for three individual searches for 'COVID', 'COVID-19', and 'coronavirus'. The searches were conducted from geolocation search engine settings to reflect the webpages found in Ireland, the United Kingdom, Canada and the United States. All searches were conducted on 17<sup>th</sup> April 2020. When searching for information on the internet users typically will pick one of the first five search results, and will typically rephrase their search criteria instead of proceeding to the second page (or further)<sup>9</sup>.

### *Readability Assessment Tools*

Four scores were used to calculate readability of the webpages; the Gunning Fog Index (GFI), the Flesch Kincaid Index (FKG) Score and Flesch Reading Ease Score (FRES) and the Simple Measure of Gobbledygook (SMOG) Index (Appendix 1). To ensure consistency and avoid human error the readability tests were done using an online readability calculator to provide FRES, FKG, GFI and SMOG scores<sup>10</sup>. All webpages were screened by the readability tool and hyperlinks, non-standard text, abbreviations and author names were not included in the analysis to prevent low-skewing of results.

### *Statistical Analysis*

Descriptive statistics were calculated for SMOG, FRES, FKG and GFI scores. Shapiro-Wilk test determined data distribution. Mean (SD) was used for normally distributed data, while median (range) was used for skewed data. Spearman's (for skewed) and Pearson's correlation coefficient (for normally distributed) was used to assess association between readability scores. ANOVAs and Kruskal-Wallis tests were used to compare differences between the results for univariate group analysis. A 5% level of significance was used for all statistical tests. All statistical analysis was performed using GraphPad Prism software Version 8 (La Jolla, CA, USA, 2020), SPSS Statistics Version 26 (IBM, 2020) and Microsoft Office Excel Version 16 (USA, 2018).

# Results

The searches were performed using the keywords: coronavirus, COVID, COVID-19. The first 20 webpages were collated from each search and the search was conducted geolocated to Ireland, the United Kingdom, Canada and the United States, totalling 240 webpages (Appendix 2). Of the 240 webpages analysed 53% (n=127) were government organisations or public health organisation webpages, 29% (n=69) were digital or social media webpages, 5% (n=11) were from scientific or educational institutions and 14% (n=33) were from other sources (Table 1a). There was a significant differences between regional spread of webpage sources (ANOVA,  $p < 0.03$ ), and there was a matching inverse correlation between webpage sources between countries (Spearman correlation  $p < 0.07$ ,  $-0.2$ ) and between continents ( $p < 0.03$ ,  $-0.2$ ).

FRES results were parametric, while FKG, GFI and SMOG scores were all non-parametric. Only 17.2% (n=165) of all the readability scores analysed demonstrated a universally readable level. 19% (n=45) of FRES scores were at a universally readable level ( $>60$ ), 32% (n=77) of FKG scores (target  $<8$ ), 37% (n=88) of GFI scores (target  $<8$ ), and only 30% (n=73) of SMOG scores were at a universally readable levels ( $<10$ ). The mean readability scores for webpages searched from all regions were below the standard universal readability levels, and there were no significant differences comparably between regions (Table 1b).

There were significant differences between the readability of webpages depending on the information source for all readability scores FRES ( $p < 0.02$ ), FKG ( $p < 0.01$ ), GFI ( $p < 0.0004$ ), and SMOG ( $p < 0.0001$ ) (Table 1c). The most readable sources across the majority of the scores were webpages issued by government and public health organisations.

The association between readability scores and rankings were moderately correlated, in particular relating to SMOG scores (Spearman correlation  $p < 0.05$ ,  $0.225$ ). All four readability scores (FRES, FKG, GFI and SMOG) correlated with each other significantly (Spearman's correlation  $p < 0.01$ ,  $0.409$  to  $0.927$ ). There was a positive association between source of information category and ranking of the webpage on the search engine results (Spearman's correlation  $p < 0.04$ ,  $0.2$ ).

**Table 1. A)** Webpages tabulated by country and source type. **B)** Readability scores for webpages by region. Statistical analysis showed no significant difference between countries or continents. FRES target score  $>60$ , FKG target  $<8$ , GFI target  $<8$  and SMOG target score  $<10$ . **C)** Readability scores for webpages by source type. ANOVA for normally distributed FRES with mean (SD), and Kruskal-Wallis for non-parametric FKG, GFI and SMOG scores with median (range).

<b>A</b>	<b>Government and Public Health Organisations</b>	<b>Educational or Scientific Institution</b>	<b>Digital Media</b>	<b>Other</b>
Ireland	29	2	16	13
United Kingdom	22	7	24	7
Canada	39	1	13	9
United States	37	1	16	4
<b>Total</b>	<b>127</b>	<b>11</b>	<b>69</b>	<b>33</b>
<b>B</b>	<b>FRES</b>	<b>FKG</b>	<b>GFI</b>	<b>SMOG</b>
	<b>Mean (SD)</b>	<b>Median (range)</b>	<b>Median (range)</b>	<b>Median (range)</b>
Ireland	48.93 (12.45)	8.7 (5-14)	8.7 (3-16)	10.4 (7-17)
United Kingdom	52.21 (13.04)	8.7 (4-20)	9.25 (4-18)	10.9 (9-18)
Canada	47.89 (9.57)	8.8 (4-14)	8.8 (2-15)	11.0 (7-15)
United States	49.58 (10.25)	8.8 (5-20)	8.6 (3-18)	11.0 (7-18)
<b>Total</b>	<b>49.65 (11.4)</b>	<b>8.8 (4-20)</b>	<b>8.8 (2-18)</b>	<b>10.9 (7-18)</b>
<b>C</b>	<b>FRES</b>	<b>FKG</b>	<b>GFI</b>	<b>SMOG</b>
	<b>Mean (SD)</b>	<b>Median (range)</b>	<b>Median (range)</b>	<b>Median (range)</b>
<b>Government and Public Health Organisations</b>	48.5 (13.1)	8.7 (4-20)	8.7 (3-18)	10.4 (7-18)
<b>Educational or Scientific Institution</b>	44.8 (10.1)	10.4 (8-14)	11.1 (8-15)	11.7 (10-15)
<b>Digital Media</b>	53.3 (10.6)	9.4 (4-14)	9.7 (2-15)	11.2 (8-14)
<b>Other</b>	47.52 (10.89)	8.6 (5-12)	8.0 (3-14)	11.1 (9-14)
<b>Univariate Analysis (p-value)</b>	<b>&lt;0.02*</b>	<b>&lt;0.01*</b>	<b>&lt;0.0004***</b>	<b>&lt;0.0001**</b>

## Discussion

Health literacy is an important barrier for communication by health professionals, public health bodies and government institutions with the public<sup>8</sup>. The COVID-19 pandemic presents a number of health literacy obstacles which include the rapid publication of information, the frequently evolving and fluctuating nature of public guidelines and health information, the lack of clear treatments for COVID-19 pneumonia, and the inconsistent and sometimes dangerous information and misinformation that is occurring online, in particular on social media<sup>11</sup>. A fundamental necessity to understanding and engaging with health information is the accessibility and readability of the information and while there is a pressure and immediacy to publish information at short notice, readability should be considered when producing health literature and information<sup>11</sup>.

The webpages analysed were mostly below a universal level for readability. That the best performing readability score found only 37% of webpages readable to a universal audience does not reflect well for the health information produced and disseminated online. This poor readability level affects understanding of the health information; resulting in poor adherence to hygiene measures, social-distancing measures, and further public health recommendations<sup>5</sup>.

Webpages most likely to be viewed are webpages on the first page of search results<sup>9</sup>, making website rankings an important factor for consideration<sup>12</sup>. The moderately positive correlation between source type and ranking of webpages on the search results is reassuring as the majority of webpages were published by public health organisations or government bodies.

Search engines have the ability to manipulate ranking settings, and sponsored search results can often tamper with what audiences see first<sup>12</sup>. Google®, has been making corporate decisions to artificially rank high-profile health information from respectable prevalence such as the World Health Organisation since early March 2020<sup>13</sup>. This might explain why Government and Public Health bodies account for 53% (n=127) of search results, and while this is reassuring because readability tends to be higher from information from those origins, the mean readability scores in this study remain poor. These differences seen between countries and continents in both the type of source information available is worth considering, given that there is a clear difference in readability between sources.

The correlations between the various readability scores was reassuring and showed that while there are some differences that the trend in detecting poor readability was similar between tests. While much has been published in the last few weeks on the quality of health information and the misinformation relating to COVID-19, this is the first assessment of readability of online information on COVID-19.

We acknowledge the limitations of this study. There are a number of weaknesses associated with each of the readability scores<sup>14,15</sup>. The tests rely on numbers of words in sentences, or syllables in words, which may not always reflect the reading level. The scores do not consider layout, infographics or figures that often help accessibility and understanding of accompanying literature. Like all infodemiology research the nature of researching online health information is limited by the constantly changing, revising and updating of online material. This study may have different results if repeated at another time.

# Conclusion

The majority of webpages relating to COVID-19 are not at a universal reading level in four major English speaking regions. However, reassuringly most webpages originated from public health organisations and government bodies. While there is an urgency in a global pandemic to publish guidance and health information, there is an onus on publishers from all information sources to publish information that is readable for all levels of comprehension, which will in turn lead to better levels of education and adherence to guidance.

# Declarations

**Ethical Approval or Consent** None

**Funding or Sponsorship** None

**Conflict of Interest** None

**Data Repository** Original data can be provided

## Author Contributions

APW conducted data collection, data analysis, manuscript writing and contributed to study design. MJC conducted data analysis and paper writing, AO'N and MO'D were involved with manuscript writing, KPT did data analysis and collection, CM, SJM and EdB conceived study design and wrote the manuscript, with final oversight.

# References

1. Hesse BW, Nelson DE, Kreps GL, Croyle RT, Arora NK, Rimer BK, et al. Trust and Sources of Health Information. *Arch Intern Med.* 2005;
2. Google. Google Trends [Internet]. Google Trends. 2020 [cited 2020 Apr 12]. Available from: <https://trends.google.com/trends/>
3. Europarl. Disinformation: How to recognise and tackle Covid-19 myths. European Parliament. 2020.
4. Song P, Karako T. COVID-19: Real-time dissemination of scientific information to fight a public health emergency of international concern. *Bioscience trends.* 2020.
5. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. *Cureus.* 2020;
6. Hernández-García I, Giménez-Júlvez T. Assessment of health information about the prevention of COVID-19 on the Internet. *JMIR public Heal Surveill.* 2020;
7. Garner M, Ning Z, Francis J. A framework for the evaluation of patient information leaflets. *Heal Expect.* 2012;

8. McInnes N, Haglund BJA. Readability of online health information: Implications for health literacy. *Informatics Heal Soc Care*. 2011;
9. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *Br Med J*. 2002;
10. Added Bytes. Readable [Internet]. Brighton, UK; 2020. Available from: <https://readable.io>
11. Abel T, McQueen D. Critical health literacy and the COVID-19 crisis. *Health Promot Int*. 2020;
12. Fu LY, Zook K, Spoehr-Labutta Z, Hu P, Joseph JG. Search Engine Ranking, Quality, and Content of Web Pages That Are Critical Versus Noncritical of Human Papillomavirus Vaccine. *J Adolesc Heal*. 2016;
13. Google. Google Company Announcement: Coronavirus: How we're helping [Internet]. Google Blog. 2020. Available from: <https://www.blog.google/inside-google/company-announcements/coronavirus-covid19-response/>
14. Keogh CJ, McHugh SM, Clarke Moloney M, Hannigan A, Healy DA, Burke PE, et al. Assessing the quality of online information for patients with carotid disease. *Int J Surg*. 2014;
15. Walsh TM, Volsko TA. Readability assessment of Internet-based consumer health information. *Respir Care*. 2008;
16. Hansberry DR, Agarwal N, Baker SR. Health literacy and online educational resources: An opportunity to educate patients. *Am J Roentgenol*. 2015;

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Appendix1.docx](#)