Original Paper

Assessing COVID-19 Health Information on Google Using the Quality Evaluation Scoring Tool (QUEST): Cross-sectional and Readability Analysis

Vismaya S Bachu[1], BS; Heba Mahjoub[1], BS; Albert E Holler[1], BA; Tudor Crihalmeanu[2], MHS; Dheevena M Bachu[3]; Varun Ayyaswami[4], MD; Pearman D Parker[5], MPH, PhD; Arpan V Prabhu[6], MD

[1]School of Medicine, Johns Hopkins University, Baltimore, MD, United States

[2]School of Medicine, West Virginia University, Morgantown, WV, United States

[3]Department of Neuroscience, University of Colorado, Boulder, CO, United States

[4]Department of Internal Medicine, University of Massachusetts Medical School, Worcester, MA, United States

[5]College of Nursing, University of Arkansas for Medical Sciences, Little Rock, AR, United States

[6]Department of Radiation Oncology, University of Arkansas for Medical Sciences, Winthrop P Rockefeller Cancer Institute, Little Rock, AR, United States

Corresponding Author:
Arpan V Prabhu, MD
Department of Radiation Oncology
University of Arkansas for Medical Sciences
Winthrop P Rockefeller Cancer Institute
4130 Shuffield Dr
Little Rock, AR, 72205
United States
Phone: 1 5016644568
Email: avprabhu@uams.edu

Abstract

Background: The COVID-19 pandemic spurred an increase in online information regarding disease spread and symptomatology.

Objective: Our purpose is to systematically assess the quality and readability of articles resulting from frequently Google-searched COVID-19 terms in the United States.

Methods: We used Google Trends to determine the 25 most commonly searched health-related phrases between February 29 and April 30, 2020. The first 30 search results for each term were collected, and articles were analyzed using the Quality Evaluation Scoring Tool (QUEST). Three raters scored each article in authorship, attribution, conflict of interest, currency, complementarity, and tone. A readability analysis was conducted.

Results: Exactly 709 articles were screened, and 195 fulfilled inclusion criteria. The mean article score was 18.4 (SD 2.6) of 28, with 7% (14/189) scoring in the top quartile. National news outlets published the largest share (70/189, 36%) of articles. Peer-reviewed journals attained the highest average QUEST score compared to national/regional news outlets, national/state government sites, and global health organizations (all $P<.05$). The average reading level was 11.7 (SD 1.9, range 5.4-16.9). Only 3 (1.6%) articles were written at the recommended sixth grade level.

Conclusions: COVID-19–related articles are vastly varied in their attributes and levels of bias, and would benefit from revisions for increased readability.

(JMIR Form Res 2022;6(2):e32443) doi: 10.2196/32443

KEYWORDS

COVID-19; COVID-19 pandemic; health literacy; readability; QUEST; online health information; cross-sectional; trend; internet; spread; symptom; quality; United States

Introduction

Since the onset of the COVID-19 pandemic, new information is released daily, if not hourly, regarding disease spread, symptomatology, and health and economic consequences. In some cases, news has been rapidly spread only to be contradicted days later. For example, at the beginning of the pandemic, hydroxychloroquine was regularly discussed in lay news and scientific journals alike. Some articles touted a 90% chance of benefit to patients with COVID-19 [1], while peer-reviewed journals soon thereafter released a lack of clinical improvement with use of the drug [2,3]. Given varying accuracy levels of innumerable sources, there is a clear need for standardized quality control of online health information especially in light of current vaccination and other public health campaigns [4].

There was a disjointed public health response, partly due to contradicting information. For example, we now know that universal masking is of the utmost importance in preventing disease transmission, but earlier in the pandemic, it was only recommended for health care professionals [1,3]. These conflicting messages may have left many consumers confused, frustrated, and unsure of what broadcast news channels and online health information to trust. The burden of sorting through the flood of information fell on the consumer and, in many instances, left the consumer feeling paralyzed with information overload and overconcern from frequent use of social media [5]. Furthermore, an analysis of online health information prior to February 6, 2020, showed low quality information relative to several different quality scoring systems, including HONcode, the JAMA benchmark, and the DISCERN instrument [6]. With the prevalence of low-quality information and sudden influx of new conflicting information and associated overwhelming emotion, we felt compelled to analyze the information being consumed by the public.

Google Trends (GT) was used to identify popular COVID-19 search terms and produce a list of related online health articles, after which the Quality Evaluation Scoring Tool (QUEST) was applied to assess validity. QUEST is a verified metric created to assess online health information, or any information available online that patients may read to learn more about their health, in a quantifiable way. It consists of seven questions that numerically measure quality of authorship, attribution, conflict of interest, currency, complementarity, and tone [7]. According to QUEST, a high-quality article is deemed trustworthy and credible, and displays an appropriate level of tone for the reader. We opted for this tool, as opposed to another scale such as DISCERN, because it provides clear guidelines on scoring, with example statements clarifying which articles should receive a score of 0 to 3. Furthermore, the scores are weighted, emphasizing the importance of attribution, conflict of interest, and tone in assessing quality. Though there are many unique tools to analyze online health information, we valued QUEST for its unambiguous scoring and similarity to the US National Library of Medicine's "Medline Plus Guide" in individually judging legitimacy of online health material [8].

In addition to systematically assessing the quality of articles using QUEST, we sought to evaluate the readability of articles resulting from the most frequently Google-searched health-related COVID-19 terms in the United States. Because it is additionally important to recognize the varying degrees of literacy within the public, a readability analysis was performed on each article to compare against the recommended sixth grade reading level for patient health communication materials [9]. Although the production of accurate health information for patients to consume is important, it is equally important for the information to be presented in an understandable manner [10]. We hypothesized that the reading levels of popularly searched health phrases would be too difficult for the average American to understand and that the public was consuming low quality online information regarding COVID-19.

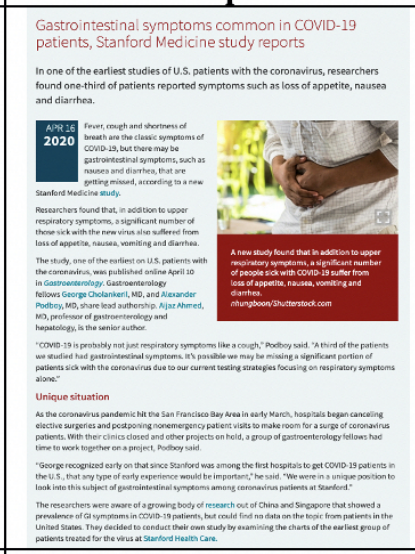
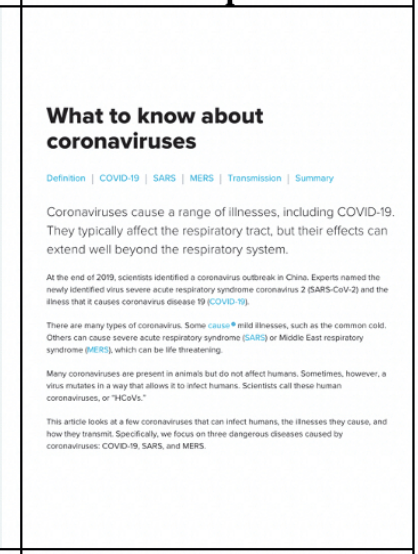
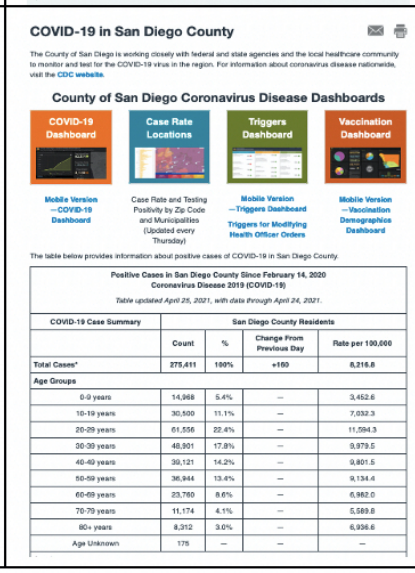
Methods

Article Selection

Institutional review board approval was not required for this study since all information was freely available online. For the purposes of this study, we defined an "article" as being any piece of published writing excluding personal blogs, editorials, and commentaries (Figure 1).

GT has been widely used to capture the most popular queries searched by the public, providing important information regarding emerging patterns. Prior studies have supported the use of GT to monitor COVID-19 incidence and public attention, especially within countries lacking proper diagnostic tools [11,12]. To prevent location bias, online articles were collected using a location-disabled search on Google.com/ncr on April 30, 2020. Using GT, we used the "Explore" option and applied the parameters "United States," "Custom date range 2/29/20 to 4/30/20," "Health," and "Web search." A start date of February 29, 2020, was chosen because this was associated with the first Centers for Disease Control and Prevention–reported death from COVID-19 in the United States [13] and marks an increase in Google searches for the word "coronavirus." From here, we sorted the search queries by "Top" and then collected the top five search queries that had an increased level of Google search frequency (Figure 2.1). These were "coronavirus," "corona," "corona virus," "symptoms,"and "coronavirus update." We then used each of these terms and searched them on GT using the aforementioned methodology. Along with the original five terms, we collected the top five related search queries (including the initially searched word) and excluded any repeat search queries. This resulted in 25 unique health-related search phrases (Figure 2.2). Of note, multi-word keywords were searched using quotations marks, and a comparative analysis of search volume without quotes was not performed.

Figure 1. Description and examples of basic inclusion (article) vs exclusion (nonarticle) criteria. Article, Example 1: “Gastrointestinal Symptoms Common in COVID-19 Patients, Stanford Medicine Study Reports.” Article, Example 2: “What to Know about Coronaviruses.” Nonarticle, Example 1: “COVID-19 in San Diego County.” Nonarticle, Example 2: “Covid-19 Coronavirus Pandemic."

	Description	Example 1	Example 2
Article	We defined “article” as a search result that included a cohesive section of writing that constituted its own separate section on a published page (as opposed to subtitle, caption, etc).	 Gastrointestinal symptoms common in COVID-19 patients, Stanford Medicine study reports In one of the earliest studies of U.S. patients with the coronavirus, researchers found one-third of patients reported symptoms such as loss of appetite, nausea and diarrhea. APR 16 2020 Fever, cough and shortness of breath are the classic symptoms of COVID-19, but there may be gastrointestinal symptoms, such as nausea and diarrhea, that are getting missed, according to a new Stanford Medicine study. Researchers found that, in addition to upper respiratory symptoms, a significant number of those sick with the new virus also suffered from loss of appetite, nausea, vomiting and diarrhea. The study, one of the earliest on U.S. patients with the coronavirus, was published online April 10 in Gastroenterology. Gastroenterology fellows George Cholankeril, MD, and Alexander Podboy, MD, share lead authorship. Aijaz Ahmed, MD, professor of gastroenterology and hepatology, is the senior author. A new study found that in addition to upper respiratory symptoms, a significant number of people sick with COVID-19 suffer from loss of appetite, nausea, vomiting and diarrhea. nhungboon/Shutterstock.com “COVID-19 is probably not just respiratory symptoms like a cough,” Podboy said. “A third of the patients we studied had gastrointestinal symptoms. It’s possible we may be missing a significant portion of patients sick with the coronavirus due to our current testing strategies focusing on respiratory symptoms alone.” Unique situation As the coronavirus pandemic hit the San Francisco Bay Area in early March, hospitals began canceling elective surgeries and postponing nonemergency patient visits to make room for a surge of coronavirus patients. With their clinics closed and other projects on hold, a group of gastroenterology fellows had time to work together on a project, Podboy said. “George recognized early on that since Stanford was among the first hospitals to get COVID-19 patients in the U.S., that any type of early experience would be important,” he said. “We were in a unique position to look into this subject of gastrointestinal symptoms among coronavirus patients at Stanford.” The researchers were aware of a growing body of research out of China and Singapore that showed a prevalence of GI symptoms in COVID-19 patients, but could find no data on the topic from patients in the United States. They decided to conduct their own study by examining the charts of the earliest group of patients treated for the virus at Stanford Health Care.	 What to know about coronaviruses Definition | COVID-19 | SARS | MERS | Transmission | Summary Coronaviruses cause a range of illnesses, including COVID-19. They typically affect the respiratory tract, but their effects can extend well beyond the respiratory system. At the end of 2019, scientists identified a coronavirus outbreak in China. Experts named the newly identified virus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the illness that it causes coronavirus disease 19 (COVID-19). There are many types of coronavirus. Some cause mild illnesses, such as the common cold. Others can cause severe acute respiratory syndrome (SARS) or Middle East respiratory syndrome (MERS), which can be life threatening. Many coronaviruses are present in animals but do not affect humans. Sometimes, however, a virus mutates in a way that allows it to infect humans. Scientists call these human coronaviruses, or “HCoVs.” This article looks at a few coronaviruses that can infect humans, the illnesses they cause, and how they transmit. Specifically, we focus on three dangerous diseases caused by coronaviruses: COVID-19, SARS, and MERS.
Non article	We defined “non article” as a search result which did not include a section of cohesive writing. Most non articles were either information dashboards, numerical dashboards, or main menu navigation pages.	 COVID-19 in San Diego County The County of San Diego is working closely with federal and state agencies and the local healthcare community to monitor and test for the COVID-19 virus in the region. For information about coronavirus disease nationwide, visit the CDC website. County of San Diego Coronavirus Disease Dashboards COVID-19 Dashboard Case Rate Locations Triggers Dashboard Vaccination Dashboard Mobile Version —COVID-19 Dashboard Case Rate and Testing Positivity by Zip Code and Municipalities (Updated every Thursday) Mobile Version —Triggers Dashboard Triggers for Modifying Health Officer Orders Mobile Version —Vaccination Demographics Dashboard The table below provides information about positive cases of COVID-19 in San Diego County. Positive Cases in San Diego County Since February 14, 2020 Coronavirus Disease 2019 (COVID-19) Table updated April 25, 2021, with data through April 24, 2021. COVID-19 Case Summary | San Diego County Residents Count | % | Change From Previous Day | Rate per 100,000 Total Cases* 275,411 100% +160 8,216.8 Age Groups 0-9 years 14,968 5.4% — 3,452.6 10-19 years 30,500 11.1% — 7,032.3 20-29 years 61,556 22.4% — 11,594.3 30-39 years 48,901 17.8% — 9,979.5 40-49 years 39,121 14.2% — 9,801.5 50-59 years 36,944 13.4% — 9,134.4 60-69 years 23,760 8.6% — 6,982.0 70-79 years 11,174 4.1% — 5,569.8 80+ years 8,312 3.0% — 6,936.6 Age Unknown 175 — — —	 worldometer Coronavirus Population COVID-19 CORONAVIRUS PANDEMIC Coronavirus Cases: 147,810,736 Deaths: 3,123,003 Recovered: 125,375,246 ACTIVE CASES CLOSED CASES Daily New Cases

Next, between April 30 and May 2, 2020, we searched each keyword phrase and collected all articles (writing that includes more than 100 words) from the first 3 pages of the Google search; this resulted in approximately 10 articles per page (Figure 3). Prior research has shown that internet users tend not to view past these first 3 pages on Google [14]. When queries yielded articles that overlapped, we excluded the repeated articles from analysis. For each article, we collected the Google page number, order on the page, article link, website name, category of website, article title, author, date of publication, and number of references.

Figure 2. (1) Google search words/phrases used to collect most viewed articles. (2) Increase in Google search popularity of the five most commonly searched health-related phrases in the United States between February 29 to April 30, 2020.

1

Health-Related Search Word or Phrase	
Coronavirus	Covid symptoms
Coronavirus cases	Symptoms covid 19
Coronavirus update	Flu symptoms
Coronavirus symptoms	Symptoms of coronavirus
Coronavirus map	Corona virus US
US coronavirus	Corona virus map
Corona	Coronavirus updates
Corona virus	Coronavirus update USA
The corona virus	Coronavirus map update
Corona virus cases	Coronavirus cases update
Corona update	Update on coronavirus
Corona virus update	US coronavirus update
Symptoms	

2

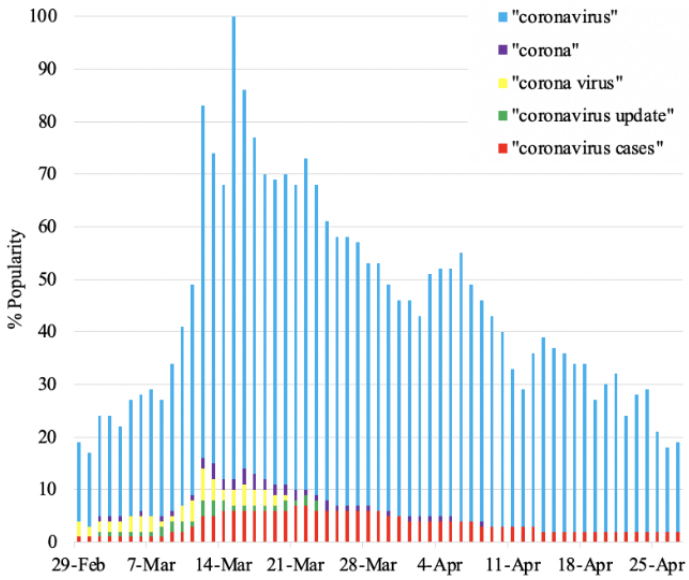
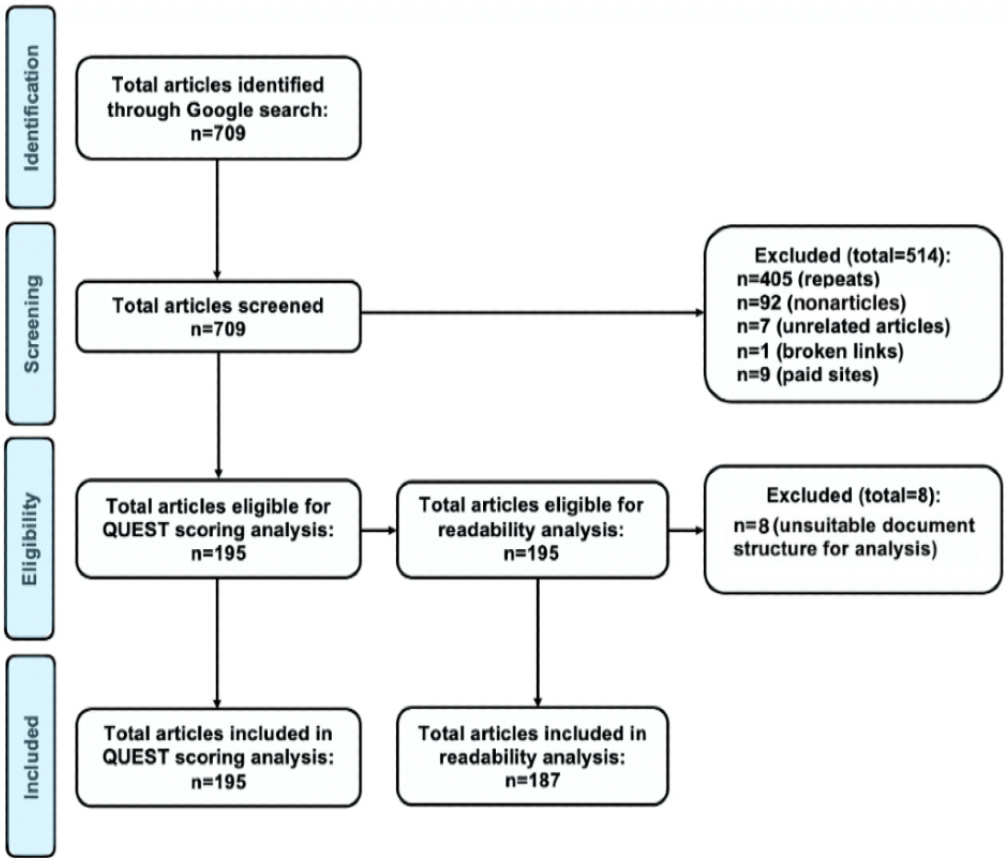


Figure 3. CONSORT (Consolidated Standards of Reporting Trials) diagram depicting article flow and eventual sample size. QUEST: Quality Evaluation Scoring Tool.



QUEST Scoring

Three separate authors scored all articles individually using each of the 7 QUEST questions and associated point values (Textbox 1). Each article's individual sections were then combined into a score between 0 and 28, where 28 was the highest quality article possible. The final score for each article was an average of the three independent scorers' analyses. Interrater consensus was determined using Fleiss's kappa metric.

Textbox 1. The Quality Evaluation Scoring Tool scoring rubric.

Authorship (score x1)

- 0—No indication of authorship or username
- 1—All other indications of authorship
- 2—Author's name and qualification clearly stated

Attribution (score x3)

- 0—No sources
- 1—Mention of expert source, research findings (though with insufficient information to identify the specific studies), links to various sites, advocacy body, or other
- 2—Reference to at least one identifiable scientific study, regardless of format (eg, information in text or reference list)
- 3—Reference to mainly identifiable scientific studies, regardless of format (in >50% of claims)
    - For all articles scoring 2 or 3 on attribution: type of study (score x 1): 0—in vitro, animal models, or editorials; 1—all observational work; 2—meta-analyses, randomized controlled trials, clinical studies

Conflict of interest (score x3)

- 0—Endorsement or promotion of intervention designed to prevent or treat condition (eg, supplements, brain training games, or foods) within the article
- 1—Endorsement or promotion of educational products and services (eg, books or care home services)
- 2—Unbiased information

Currency (score x1)

- 0—No date present
- 1—Article is dated but 5 years or older
- 2—Article is dated within the last 5 years

Complementarity (score x1)

- 0—No support of the patient-physician relationship
- 1—Support of the patient-physician relationship

Tone (includes title; score x3)

- 0—Fully supported (authors fully and unequivocally support the claims, strong vocabulary; eg, "cure," "guarantee," and "easy"; mostly use of nonconditional verb tenses ["can," "will"], no discussion of limitations)
- 1—Mainly supported (authors mainly support their claims but with more cautious vocabulary; eg, "can reduce your risk" or "may help prevent"; no discussion of limitations
- 2—Balanced/cautious support (authors' claims are balanced by caution, includes statements limitations or contrasting findings)

Statistical Analysis

Excel (Microsoft Corporation) was used to conduct the statistical analysis and generate figures for this study. In addition to determining general descriptive metrics (mean, median, etc), we coined the metric search order product to examine if there was any statistical difference in quality between articles toward the beginning and end of the results. Equal to the results page number on Google multiplied by the order of article on that page, the search order product encompasses the article's hierarchy in search results. For example, if an article was second on the third page of the results on Google, its search order product would be 6. A low search order product indicates an earlier appearing article once its associated phrase is searched (meaning increased public exposure), while a high search order product is characteristic of a later-appearing article (decreased public exposure). Comparative *t* tests and Pearson correlation analyses were conducted to stratify scores by variables. The Benjamini-Hochberg false detection rate correction for multiplicity was applied to appropriately adjust P values.

Readability Analysis

Readability analysis was performed using Readability Studio Professional Edition Version 2015 (Oleander Software, Ltd), applying nine validated formulas to quantify article readability: Coleman-Liau Index [15], Flesch-Kincaid Grade Level [16], FORCAST formula [17], Fry graph [18], Gunning Fog Index [19], New Dale-Chall [20], New Fog Count [16], Raygor Reading Estimate [21], and SMOG (Simple Measure of Gobbledygook) [22]. Nine different readability scales were used

to minimize the bias that comes with using only one scale. We subsequently calculated the reading level for each article by averaging estimates derived from all nine scales. These were then compared to the American Medical Association–recommended reading level of sixth grade for health education materials [9]. A 10th readability formula, the Flesch Reading Ease (FRE) [23], was applied separately as it calculates reading level on a different scale. FRE scores of 0 to 30 indicate very difficult, 30 to 50 difficult, 50 to 60 fairly difficult, 60 to 70 standard, 70 to 80 fairly easy, 80 to 90 easy, and 90 to 100 very easy.

Results

QUEST Analysis

A total of 709 Google results listings were initially examined. After excluding repeated articles (n=405), nonarticles (n=92), unrelated articles (n=7), broken links (n=1), and paid sites (n=9) from the analysis, 195 individual articles were scored using QUEST (Figure 3).

The mean article score was 18.4 (SD 2.6) of 28, with only 7% (14/189) of articles in the top score quartile and 89% (173/189) of articles in the top half of scores. National news outlets published the largest share (70/189, 36%) of the analyzed articles, followed by private health-focused entities (45/189, 23%) and regional news outlets (29/189, 15%; Table 1).

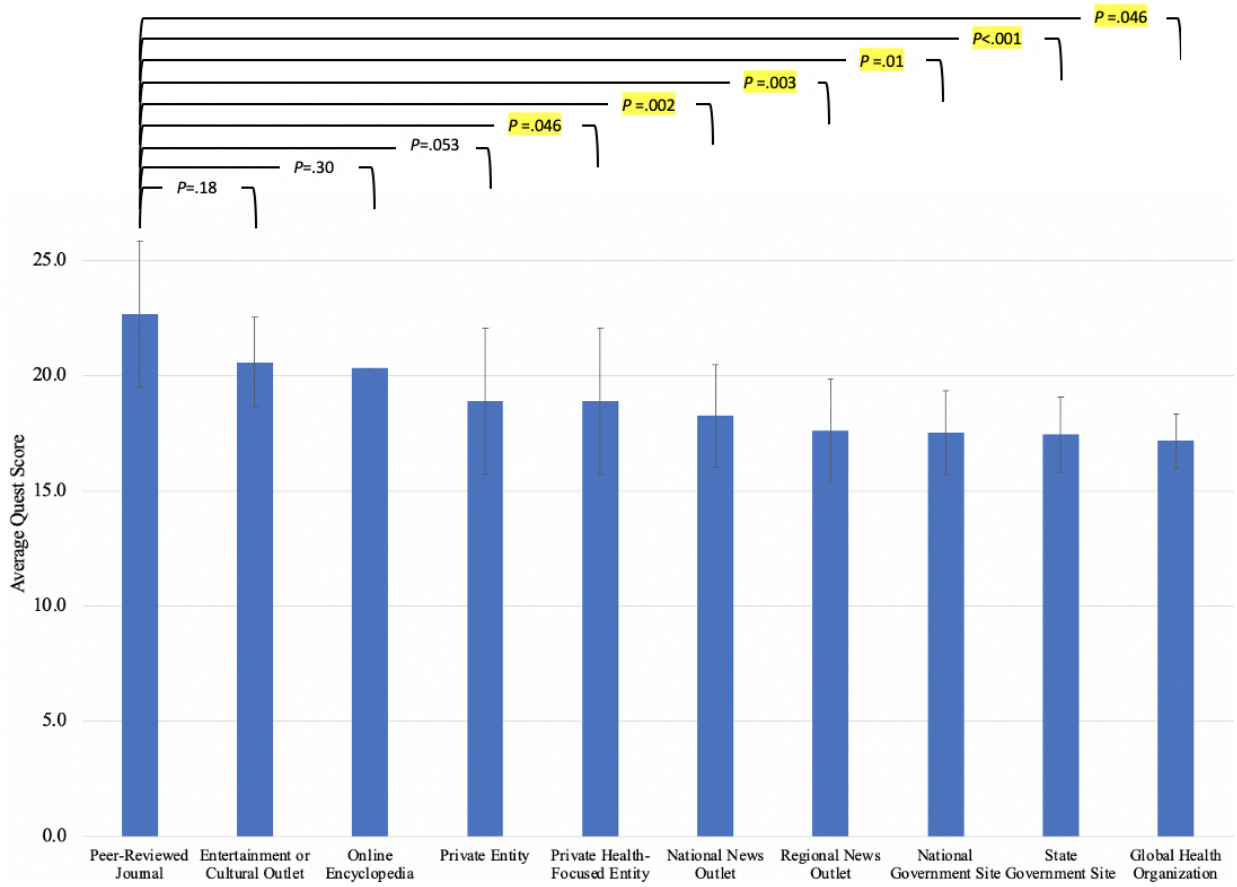
Categorically, global health organization sites had the lowest average score (mean 17.2, SD 1.2) and least dispersive data set (σ^2=1.39; Figure 4). Peer-reviewed journals had the most dispersive data set (σ^2=10.1) and the highest average QUEST score (mean 22.7, SD 3.18), with significantly higher quality averages as compared to national news outlets (mean 18.3, SD 2.20; *P*=.002), regional news outlets (mean 17.6, SD 2.24; *P*=.003), national government sites (mean 17.5, SD 1.81; *P*=.001), state government sites (mean 17.4, SD 1.65, *P*=.002), and global health organizations (mean 17.2, SD 1.18; *P*=.046). In addition, entertainment and cultural outlets (mean 20.6, SD 1.95) also had a significantly higher quality score than regional news outlets (mean 17.6, SD 2.24; *P*=.009) and state government sites (mean 17.4, SD 1.65; *P*=.002).

Table 1. Basic descriptive statistics regarding analyzed articles.

Descriptor	Articles, n (%)
QUEST[a] scoring fraction	
0%-25%	0 (0)
25%-50%	22 (11)
50%-75%	159 (82)
75%-100%	14 (7)
Category	
National news outlet	70 (36)
Private health-focused entity	45 (23)
Regional news outlet	29 (15)
Private entity	18 (9)
State government site	16 (8)
National government site	7 (4)
Peer-reviewed journal	3 (2)
Entertainment or cultural outlet	4 (2)
Global health organization	2 (1)
Online encyclopedia	1 (<1)
Google page	
1	47 (24)
2	76 (39)
3	72 (37)
Order on page	
0-5	106 (54)
6-10	89 (46)

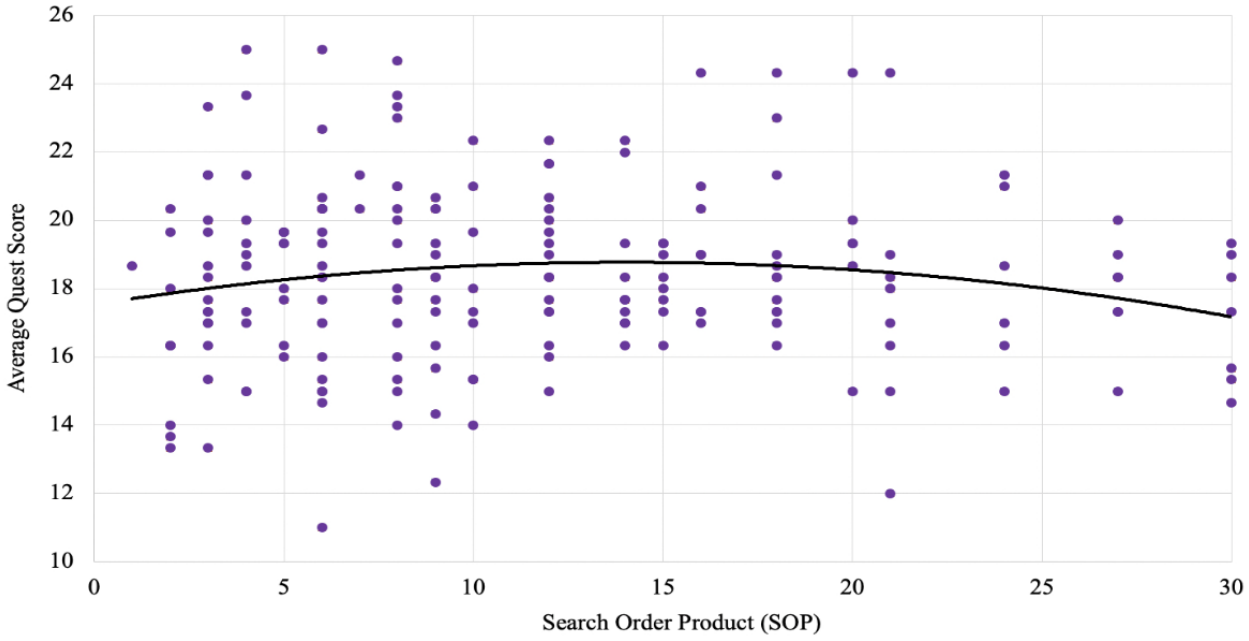
[a]QUEST: Quality Evaluation Scoring Tool.

Figure 4. QUEST score by categorization of 195 articles into the article type. Peer-reviewed journal (PRJ; mean 22.7, SD 2.20); entertainment or cultural outlet (mean 20.6, SD 1.95); online encyclopedia (mean 20.3, SD not applicable); private health-focused entity (mean 18.9, SD 3.17); private entity (mean 18.9, SD 3.17); national news outlet (NNO; mean 18.3, SD 2.2); regional news outlet (RNO; mean 17.6, SD 2.24); national government site (NGS; mean 17.5, SD 1.81); state government site (SGS; mean 17.4, SD 1.65); global health organization (GHO; mean 17.2, SD 1.18). PRJ quality score was significantly higher than NNO (*P*=.002), RNO (*P*=.003), NGS (*P*=.001), SGS (*P*<.001), and GHO (*P*=.046) scores. QUEST: Quality Evaluation Scoring Tool.



Analysis of the QUEST scores by the search order product showed no significant trends (*R*=–0.16; *P*=.75; Figure 5), discrediting any hierarchy by listing order within the sample set. A significant positive correlation (*R*=0.25; *P*<.001) existed between the number of references (≥1) in an article and the QUEST diagnostic score. Due to QUEST already allocating points in a binary fashion for containing references, only articles with ≥1 reference were considered.

Figure 5. QUEST score stratified by search order product (R=–0.16; P>.05); there is no hierarchy in score based on the order of search results. QUEST: Quality Evaluation Scoring Tool.

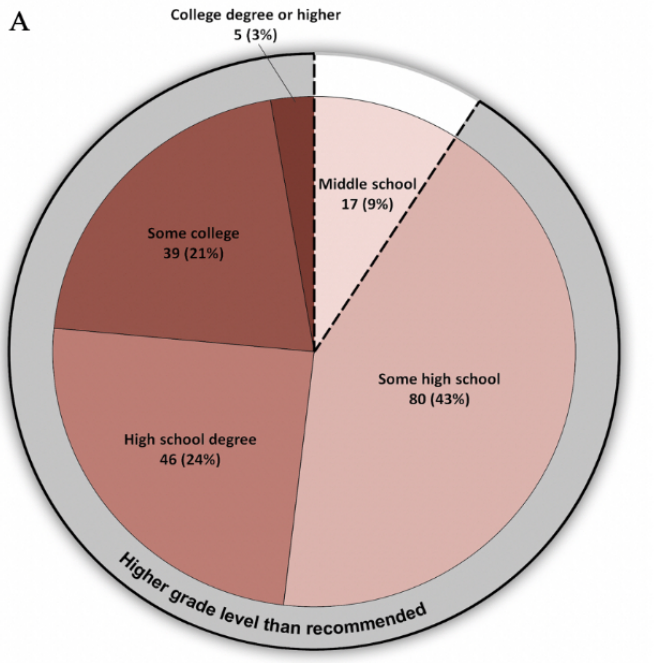


Readability Analysis

The readability levels for 187 of 195 articles were collected; 8 articles containing document structure unsuitable for analysis were excluded. The average reading level across all 187 articles was 11.7 (SD 1.9), ranging from 5.4 (fifth grade reading level) to 16.9 (undergraduate senior reading level). There was variability among the readability scales, with New Fog Count scoring the overall lowest mean readability (mean 9.3, SD 2.7) and Fry scoring the overall highest mean readability (mean 14, SD 2.9). FRE scored an overall average of 47.2 (SD 11.4), corresponding to difficult and representative of college-level reading. Based on the averages of the nine readability scales for each article, only 3 (1.6%) articles were written at the recommended sixth grade levels [9], with 44 (23.5%) written beyond a high school level (Figure 6).

Figure 6. Number of articles (n=187) stratified by educational stages; 187 COVID-19–related articles' readability score breakdown: middle school, <9th grade; some high school, 9th to 11th grade; high school degree, 12th grade; some college, 13th to 15th grade; college degree or higher, ≥16th grade.

A



B

Reading level (Corresponding readability score)	Number of articles
elementary school level (<6)	1
middle school level (6-8.9)	16
high school level (9-12.9)	126
undergraduate level (13-16.9)	44
postgraduate level (>17)	0

Discussion

Analysis of Results

This study systematically assessed and determined that articles resulting from the most frequently Google-searched health-related COVID-19 terms in the United States were of higher quality and readability than hypothesized. QUEST proved versatile in synthesizing aggregate data regarding different aspects of literature including authorship, attribution, conflict of interest, currency, complementarity, and tone. Despite the high prevalence of misinformation on the internet, analysis of our data set revealed that 89% of articles scored in the upper quartiles, suggesting that online information in the United States regarding COVID-19 was of a higher quality than anticipated.

With the uptick in news dissemination by national media after the onset of the pandemic, national news outlets were unsurprisingly the largest source of our sample set followed by private health-focused entities. Interestingly, there was less output from global organizations such as the World Health Organization and United Nations; these organizations only had 1% of total article output, contradicting their organizational goals of far-reaching public health campaigns and initiatives [24]. This discrepancy in expected and observed output could be due to organizational choices to frequently update a centralized information page as opposed to generating new articles that would show up as separate listings. Additionally, different categorical sources allocate varying levels of resources and personnel to public-facing operations that would result in searchable online information [25].

The lack of significant trends associated with an increasing search order product value implied a qualitatively homogenous sampling of articles by exposure in our analysis, validating the decision to analyze only the first three pages of each term's Google search (Figure 5). Using the product of search metrics instead of a direct numerical order of results allowed us to place increased weight on articles listed toward the top of later result pages. Among the sample set, the average quality score of the first articles on the first page (mean 18.7, SD 0.0) was not significantly different from the average quality score of the 10th article on the third page (mean 17.1, SD 1.88) of Google search results.

Stratification of the QUEST scores by article categorization revealed source-based qualitative differences, in part due to different data-gathering and publishing processes. Output from peer-reviewed journals had the highest average score (mean 22.7, SD 3.18), likely due to their rigorous scientific vetting process prior to publication [26]. Increased variation in this category (σ^2=10.1) may be a reflection of the varying requirements set forth by journals and a smaller sample size included in our analysis. Peer-reviewed journals had a significantly increased average QUEST score as compared to national news outlets (mean 18.3, SD 2.20; *P*=.002), regional news outlets (mean 17.6, SD 2.24; *P*=.003), national government sites (mean 17.5, SD 1.81; *P*=.001), state government sites (mean 17.4, SD 1.65; *P*=.002), and global health organizations (mean 17.2, SD 1.18; *P*=.046). Unfortunately, the latter categories were marked by the lowest average information quality, though still mostly in the second (7.5-15) and third (14-22.5) score quartiles. Additionally, entertainment and cultural outlets (mean 20.6, SD 1.95) were characterized by significantly higher quality information compared to regional news outlets and state government sites, perhaps reflecting their tendency to target wider audiences [27].

The majority of articles (170/187, 91%) were written well above American Medical Association–recommended reading levels [21] (Figure 6). Results from the readability analysis substantiated our hypothesis that most COVID-19 articles would be too difficult for the average American to read, in line with the results of prior smaller studies [28]. With the overall FRE score representative of college-level reading and categorically falling under "Difficult," the ability of these most-searched articles to convey accurate information is automatically diminished. Though not directly contributing to misinformation, mismatched comprehension levels lead to knowledge gaps; this may push the public to turn toward other less reliable modalities to stay informed [29]. Ahmed Siddiqui et al [29] specifically discussed the pervasiveness of nonevidence-based medical advice on social media as a "hidden epidemic" considering its ability to transcend geographic and cultural boundaries. Less readable online sources may indirectly facilitate the spread of misinformation regardless of high article quality. This has been affirmed by studies using different criteria from our own including the DISCERN scoring system [6], JAMA benchmarks [30], and even the HONcode system [31]. Additionally, with massive public health awareness efforts underway to encourage COVID-19 vaccination and safe social practices, accurate online media has become increasingly important as a direct source of information for all demographics [32].

Moving forward, publishing sources may benefit from using resources to optimize communication of health information. The Agency for Healthcare Research and Quality updated the second edition of their Health Literacy Universal Precautions Toolkit (HLUPT) recently in September 2020. The document outlines strategies to enhance overall patient understanding, and even contains a section focused on written communication [33]. In a 2015 study, Brega et al [34] determined that applying the sections of the HLUPT pertaining to written materials led to better readability of revised documents. Subjecting patient-facing articles to rigorous quality and literacy guidelines will aide in improving both publishing standards and consumer understanding, both of which are required to best communicate vital information.

Conclusion

The COVID-19 pandemic in the United States was accompanied by an influx of online health information. To investigate the need for quality control of this information, we assessed articles resulting from the most-searched health-related terms in the United States using the QUEST rubric and readability software. Despite the high prevalence and transmission of misinformation during the COVID-19 pandemic, the most frequently searched Google articles had good information quality. Still, the majority of these articles were written above the recommended reading level for the public, diminishing their ability to counteract the spread of misinformation.

Limitations

The limitations of this study include the small sample size, use of only three raters, and lack of individual comparative analysis when determining search keywords. Although GT was used to identify popular keywords, the search volumes of multi-word keywords were not compared against their results without use of quotations. This likely resulted in some bias of listing order because it excluded results only found if only one of the multi-word terms were searched. Furthermore, Google's newer quality rating guidelines adopted in 2019 have resulted in increased automatic filtering criteria, likely resulting in higher quality and personalized results than would have otherwise been listed [35].

Only the QUEST scale was used to measure article validity, and there are a range of other evaluation tools in the literature that may provide differing or complementary data. Additionally, the Fleiss's kappa value of our study was 0.0095, indicating a slight agreement between raters when it came to absolute scores. This may have been due to the subjective nature of the QUEST rubric especially in areas such as attribution, tone, and conflict of interest. Even still, author relative rankings of articles correlated to a greater extent than the absolute score values, indicating a shifted but similarly trending rating among raters. Additionally, facets of QUEST, such as authorship and currency, allocate points for characteristics that do not directly correlate with information accuracy, explaining the lower scores of government and global health organizations due to inherent formatting preferences (ie, omitting authors). The QUEST scale does not address every aspect of misinformation, although it does focus on some aspects such as attribution that is seen in Table 1. Further studies on the spread of misinformation, especially against the backdrop of the COVID-19 pandemic, would benefit from examining media outside articles such as radio, social media, and television.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. PDP is currently supported by the University of Arkansas for Medical Sciences Translational Research Institute grants KL2TR003108 and UL1TR003107 through the National Center for Advancing Translational Sciences of the National Institutes of Health and Arkansas Breast Cancer Research Program.

A supplementary data file (Multimedia Appendix 1) has been provided with the raw data online.

Authors' Contributions

All authors were involved in the design and conception of this manuscript. HM performed the literature search. VSB, AEH, and DMB collected the data. VSB and TC analyzed the articles. VSB, HM, TC, PDP, and AVP wrote the primary manuscript. All authors critically revised the manuscript. AVP supervised the study. All authors have approved the manuscript as it is written.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary data file.
[XLSX File (Microsoft Excel File), 75 KB-Multimedia Appendix 1]

References

1. Hydroxychloroquine has about 90 percent chance of helping COVID-19 patients. Association of American Physicians and Surgeons. URL: https://aapsonline.org/hcq-90-percent-chance/ [accessed 2020-09-06]
2. Boulware DR, Pullen MF, Bangdiwala AS, Pastick KA, Lofgren SM, Okafor EC, et al. A randomized trial of hydroxychloroquine as postexposure prophylaxis for Covid-19. N Engl J Med 2020 Aug 06;383(6):517-525 [FREE Full text] [doi: 10.1056/NEJMoa2016638] [Medline: 32492293]
3. Skipper CP, Pastick KA, Engen NW, Bangdiwala AS, Abassi M, Lofgren SM, et al. Hydroxychloroquine in nonhospitalized adults with early COVID-19: a randomized trial. Ann Intern Med 2020 Oct 20;173(8):623-631 [FREE Full text] [doi: 10.7326/M20-4207] [Medline: 32673060]
4. Ruiz JB, Bell RA. Understanding vaccination resistance: vaccine search term selection bias and the valence of retrieved information. Vaccine 2014 Oct 07;32(44):5776-5780. [doi: 10.1016/j.vaccine.2014.08.042] [Medline: 25176640]
5. Farooq A, Laato S, Islam AKMN. Impact of online information on self-isolation intention during the COVID-19 pandemic: cross-sectional study. J Med Internet Res 2020 May 06;22(5):e19128 [FREE Full text] [doi: 10.2196/19128] [Medline: 32330115]
6. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the internet: infodemiology study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: 10.2196/18444] [Medline: 32250960]

7. Robillard JM, Jun JH, Lai J, Feng TL. The QUEST for quality online health information: validation of a short quantitative tool. BMC Med Inform Decis Mak 2018 Oct 19;18(1):87 [FREE Full text] [doi: 10.1186/s12911-018-0668-9] [Medline: 30340488]
8. Evaluating internet health information: a tutorial from the National Library of Medicine. MedlinePlus. URL: https://medlineplus.gov/webeval/webeval.html [accessed 2020-09-06]
9. Weiss BD, Schwartzberg JG, Davis TC, Parker RM, Williams MV, Wang CC. Health Literacy: A Manual for Clinicians. Chicago, IL: American Medical Association Foundation and American Medical Association; 2003.
10. Daraz L, Morrow AS, Ponce OJ, Farah W, Katabi A, Majzoub A, et al. Readability of online health information: a meta-narrative systematic review. Am J Med Qual 2018;33(5):487-492. [doi: 10.1177/1062860617751639] [Medline: 29345143]
11. Ortiz-Martínez Y, Garcia-Robledo JE, Vásquez-Castañeda DL, Bonilla-Aldana DK, Rodriguez-Morales AJ. Can Google® trends predict COVID-19 incidence and help preparedness? The situation in Colombia. Travel Med Infect Dis 2020;37:101703 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101703] [Medline: 32360323]
12. Hu D, Lou X, Xu Z, Meng N, Xie Q, Zhang M, et al. More effective strategies are required to strengthen public awareness of COVID-19: evidence from Google Trends. J Glob Health 2020 Jul;10(1):011003 [FREE Full text] [doi: 10.7189/jogh.10.011003] [Medline: 32373339]
13. CDC, Washington State report first COVID-19 death. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/media/releases/2020/s0229-COVID-19-first-death.html [accessed 2020-09-06]
14. Beitzel SM, Jensen EC, Chowdhury A, Frieder O, Grossman D. Temporal analysis of a very large topically categorized Web query log. J Am Soc Inf Sci Technol 2006 Jan 15;58(2):166-178. [doi: 10.1002/asi.20464]
15. Coleman M, Liau T. A computer readability formula designed for machine scoring. J Appl Psychol 1975 Jul;60(2):283-284. [doi: 10.1037/h0076540]
16. Kincaid JP, Fishburne Jr R, Rogers R, Chissom B, Fishburne JR, Rogers R. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Defense Technical Information Center 1975:1.
17. Caylor J, Sticht T, Fox L, Ford J. Methodologies for determining reading requirements military occupational specialties. Hum Resources Res Organ 1973;7:81.
18. Fry E. A readability formula that saves time. J Reading 1968;11:575-578.
19. Gunning R. The Technique of Clear Writing. New York, NY: McGraw-Hill; 1952.
20. Chall J, Dale E. Readability Revisited: The New Dale-Chall Readability Formula. Brookline, MA: Brookline Books; 1995.
21. Raygor A. The Raygor readability estimate: a quick and easy way to determine difficulty. Read Theory Res Pract 1977:259-263.
22. McLaughlin H. SMOG Grading: a new readability formula. J Reading 1969;12:639-646.
23. Flesch R. A new readability yardstick. J Appl Psychol 1948 Jun;32(3):221-233. [doi: 10.1037/h0057532] [Medline: 18867058]
24. WHO global health days. World Health Organization. URL: https://www.who.int/campaigns [accessed 2020-09-18]
25. Das T. "Basically a house of experts: the production of World Health Organization information". Afr Health Sci 2010 Dec;10(4):390-394 [FREE Full text] [Medline: 21416042]
26. Swartz K. Peer-reviewed journals and quality. Inquiry 1999;36(2):119-121 [FREE Full text]
27. Understanding Media and Culture: An Introduction to Mass Communication. Minnesota: University of Minnesota Libraries; 2016.
28. Basch CH, Mohlman J, Hillyer GC, Garcia P. Public health communication in time of crisis: readability of on-line COVID-19 information. Disaster Med Public Health Prep 2020 Oct;14(5):635-637 [FREE Full text] [doi: 10.1017/dmp.2020.151] [Medline: 32389144]
29. Ahmed Siddiqui MY, Mushtaq K, Mohamed M, Al Soub H, Hussein Mohamedali MG, Yousaf Z. "Social Media Misinformation"-an epidemic within the COVID-19 pandemic. Am J Trop Med Hyg 2020 Aug;103(2):920-921 [FREE Full text] [doi: 10.4269/ajtmh.20-0592] [Medline: 32534600]
30. Fan KS, Ghani SA, Machairas N, Lenti L, Fan KH, Richardson D, et al. COVID-19 prevention and treatment information on the internet: a systematic analysis and quality assessment. BMJ Open 2020 Sep 10;10(9):e040487 [FREE Full text] [doi: 10.1136/bmjopen-2020-040487] [Medline: 32912996]
31. Valizadeh-Haghi S, Khazaal Y, Rahmatizadeh S. Health websites on COVID-19: are they readable and credible enough to help public self-care? J Med Libr Assoc 2021 Jan 01;109(1):75-83 [FREE Full text] [doi: 10.5195/jmla.2021.1020] [Medline: 33424467]
32. Naveed MA, Shaukat R. Health literacy predicts Covid-19 awareness and protective behaviours of university students. Health Info Libr J 2021 Oct 01:10.1111/hir.12404 [FREE Full text] [doi: 10.1111/hir.12404] [Medline: 34595814]
33. Health Literacy Universal Precautions Toolkit, 2nd Edition. Agency for Healthcare Research and Quality. URL: https://www.ahrq.gov/health-literacy/improve/precautions/toolkit.html [accessed 2021-03-13]

34. Brega AG, Freedman MAG, LeBlanc WG, Barnard J, Mabachi NM, Cifuentes M, et al. Using the health literacy universal precautions toolkit to improve the quality of patient materials. J Health Commun 2015;20 Suppl 2:69-76 [FREE Full text] [doi: 10.1080/10810730.2015.1081997] [Medline: 26513033]
35. Google. General Guidelines Overview. Search Quality Rating Program. 2021 Oct 19. URL: https://static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf [accessed 2021-12-01]

Abbreviations

FRE: Flesch Reading Ease
GT: Google Trends
HLUPT: Health Literacy Universal Precautions Toolkit
QUEST: Quality Evaluation Scoring Tool
SMOG: Simple Measure of Gobbledygook

Edited by T Leung; submitted 28.07.21; peer-reviewed by A Mavragani, P Ghezzi; comments to author 10.11.21; revised version received 31.12.21; accepted 06.01.22; published 11.02.22

Please cite as:
Bachu VS, Mahjoub H, Holler AE, Crihalmeanu T, Bachu DM, Ayyaswami V, Parker PD, Prabhu AV
Assessing COVID-19 Health Information on Google Using the Quality Evaluation Scoring Tool (QUEST): Cross-sectional and Readability Analysis
JMIR Form Res 2022;6(2):e32443
URL: https://formative.jmir.org/2022/2/e32443
doi: 10.2196/32443
PMID: 34995206