



# Evaluation of COVID-19 Information Provided by Digital Voice Assistants

ALYSEE SHIN YING GOH

LI LIAN WONG

KEVIN YI-LWERN YAP

\*Author affiliations can be found in the back matter of this article

## ABSTRACT

**Background:** Digital voice assistants are widely used for health information seeking activities during the COVID-19 pandemic. Due to the rapidly changing nature of COVID-19 information, there is a need to evaluate COVID-related information provided by voice assistants, to ensure consumers' needs are met and prevent misinformation. The objective of this study is to evaluate COVID-related information provided by the voice assistants in terms of relevance, accuracy, comprehensiveness, user-friendliness and reliability.

**Materials and Methods:** The voice assistants evaluated were Amazon Alexa, Google Home, Google Assistant, Samsung Bixby, Apple Siri and Microsoft Cortana. Two evaluators posed COVID-19 questions to the voice assistants and evaluated responses based on relevance, accuracy, comprehensiveness, user-friendliness and reliability. Questions were obtained from the World Health Organization, governmental websites, forums and search trends. Data was analyzed using Pearson's correlation, independent samples t-tests and Wilcoxon rank-sum tests.

**Results:** Google Assistant and Siri performed the best across all evaluation parameters with mean scores of 84.0% and 80.6% respectively. Bixby performed the worst among the smartphone-based voice assistants (65.8%). On the other hand, Google Home performed the best among the non-smartphone voice assistants (60.7%), followed by Alexa (43.1%) and Cortana (13.3%). Smartphone-based voice assistants had higher mean scores than voice assistants on other platforms (76.8% versus 39.1%,  $p = 0.064$ ). Google Assistant consistently scored better than Google Home for all the evaluation parameters. A decreasing score trend from Google Assistant, Siri, Bixby, Google Home, Alexa and Cortana was observed for majority of the evaluation criteria, except for accuracy, comprehensiveness and credibility.

**Conclusion:** Google Assistant and Apple Siri were able to provide users with relevant, accurate, comprehensive, user-friendly, and reliable information regarding COVID-19. With the rapidly evolving information on this pandemic, users need to be discerning when obtaining COVID-19 information from voice assistants.



**IJS Press**

Part of the IJS Publishing Group

## CORRESPONDING AUTHOR:

**Kevin Yi-Lwern Yap, PhD**

Senior Lecturer in Public Health (Digital Health), Department of Public Health, School of Psychology and Public Health, La Trobe University, Melbourne (Bundoora), VIC 3086, Australia

[kevinyap.ehealth@gmail.com](mailto:kevinyap.ehealth@gmail.com);  
[k.yap@latrobe.edu.au](mailto:k.yap@latrobe.edu.au)

## KEYWORDS:

Voice assistants; COVID-19 information; COVID-19 pandemic; Health information; Infodemic; Quality evaluation

## TO CITE THIS ARTICLE:

Goh ASY, Wong LL, Yap KY-L. Evaluation of COVID-19 Information Provided by Digital Voice Assistants. *International Journal of Digital Health*. 2021; 1(1): 3, 1–11. DOI: <https://doi.org/10.29337/ijdh.25>



QUALITY  
IMPROVEMENT  
STUDY

## 1. INTRODUCTION

Digital voice assistants are becoming widely used in today's world. In 2020, there were 4.2 billion voice assistants used in various digital platforms worldwide [1], such as smartphones, laptops and smart speakers. Commonly used smartphone voice assistants included Apple Siri (44% in consumer usage), Google Assistant (30%) and Samsung Bixby (4%) [2]. Other home-based speakers like Amazon Alexa (64.6% in consumer usage) and Google Home (19.6%) [2], and the laptop's Microsoft Cortana (11.4%) were also commonly used [2]. In a recent survey, it was shown that 51.9% of US consumers would consider a voice assistant for healthcare-related issues [3].

There are many instances whereby voice assistants have been used for healthcare-related issues. For example, the Cedars-Sinai Medical Center had used an Alexa-powered platform for patients to verbally request for their nurses, which would be sent to the nurses' mobile phones [4]. Another study by Boyd and Wilson found that Google internet searches and Google Assistant fared better than Siri for smoking cessation information, but there was room for improvement for all three voice assistants in sourcing expert content [5]. Alagha and Helbing found that Google Assistant and Siri understood consumer queries about vaccine safety and use better and provided more reliable sources than Alexa [6]. In contrast, Miner *et al.* reported that Siri, Cortana, Google Now and S Voice were inconsistent and incomplete in their responses to queries regarding mental health, interpersonal violence and physical health [7]. Similarly, in the study by Kocaballi *et al.* [8], the authors suggested that Alexa, Siri, Google Assistant, Google Home, Cortana and Bixby were limited in their ability to deal with prompts about mental and physical health, violence and lifestyle. These studies have shown inconsistency in the responses of voice assistants. Furthermore, from our knowledge, there have been no studies that have evaluated voice assistants on sudden disease outbreaks and pandemics.

The usage of voice assistants by consumers to access news and information about the current coronavirus disease (COVID-19) has been increasing [9]. The rapidly evolving information about this pandemic has led to an infodemic, and there are many sources with poor quality information being generated on the Internet that voice assistants may access and provide to consumers [10]. Voice assistants can relieve the burden of healthcare professionals by informing consumers about COVID-19 symptoms and help them recognize their symptoms [11]. Voice assistants also offer anonymity, which can benefit consumers who fear disclosing their worries or symptoms to a healthcare professional [12]. Given the benefits that voice assistants offer in such situations, developers need to quickly update their voice assistants with the necessary abilities in order to prevent

misinformation during the pandemic [13]. COVID-19 is an infectious disease that is transmissible via fomites [14], thus another advantage of using voice assistants is its hands-free accessibility, since consumers do not have to touch their devices to communicate, hence reducing possible transmission of the virus.

Major companies, such as Apple and Amazon, have equipped their voice assistants, Siri and Alexa, with the functionality to screen users for COVID-19 based on their symptoms and to provide advice accordingly [15, 16]. However, research has not been done on voice assistants' ability to provide consumers with relevant, accurate, comprehensive, user-friendly and reliable health information regarding a pandemic, such as COVID-19. Relevant, comprehensive and user-friendly information is important to ensure consumers' needs are fully met, while accurate and reliable information will ensure consumers are not misinformed. Hence, this study aims to evaluate the COVID-19-related information provided by voice assistants in terms of relevance, accuracy, comprehensiveness, user-friendliness and reliability.

## 2. METHODOLOGY

### 2.1. VOICE ASSISTANTS EVALUATED

The voice assistants that were evaluated were: Amazon Alexa, Google Assistant, Google Home, Apple Siri, Microsoft Cortana and Samsung Bixby. Alexa was accessed via Echo Dot. Google Assistant and Siri were accessed on an iPhone 11. Cortana was accessed via a Windows laptop and Bixby via a Samsung Galaxy S8.

### 2.2. QUESTIONS ON COVID-19

A series of commonly asked COVID-19 questions was compiled along with their respective answers from the websites of the World Health Organization (WHO) [17], United States Centers for Disease Control and Prevention (US CDC) [18], United Kingdom National Health Service (UK NHS) [19], European Centre for Disease Prevention and Control [20], Public Health Agency of Canada [21], Australian Government's Department of Health [22], Government of India's Ministry of Health and Family Welfare [23], Ministry of Health Singapore (MOH) [24] and National Centre for Infectious Diseases Singapore (NCID) [25].

A total of 56 questions were collated and organized into 6 categories: general information, prevention, transmission, screening, diagnosis, and treatment ([Appendix A](#)). The questions were checked against frequently asked questions found on public forums such as AskDr [26], [Patient.info](#) [27] and MedHelp [28]. Questions not in the original list by WHO and the government websites, but had appeared multiple times across these forums were compared with search trend data from Google Search and AnswerThePublic [29] to confirm that they were frequently asked questions. Some questions were rephrased to add context and questions

that incorporated more than one topic were split into their respective categories.

### 2.3 EVALUATION RUBRIC

The rubric used was adapted from 3 studies on voice assistants in healthcare [5, 6, 8] and the DISCERN [30] and HONcode [31, 32] quality evaluation tools (Figure 1). The

point system was adapted from Alagha and Helbing [6]. The rubric evaluated 5 parameters: relevance, reliability, accuracy, comprehensiveness, and user-friendliness of information provided. Relevance was evaluated based on how well the voice assistant’s response understood (comprehension ability) and addressed the question (applicability of information). Comprehension ability

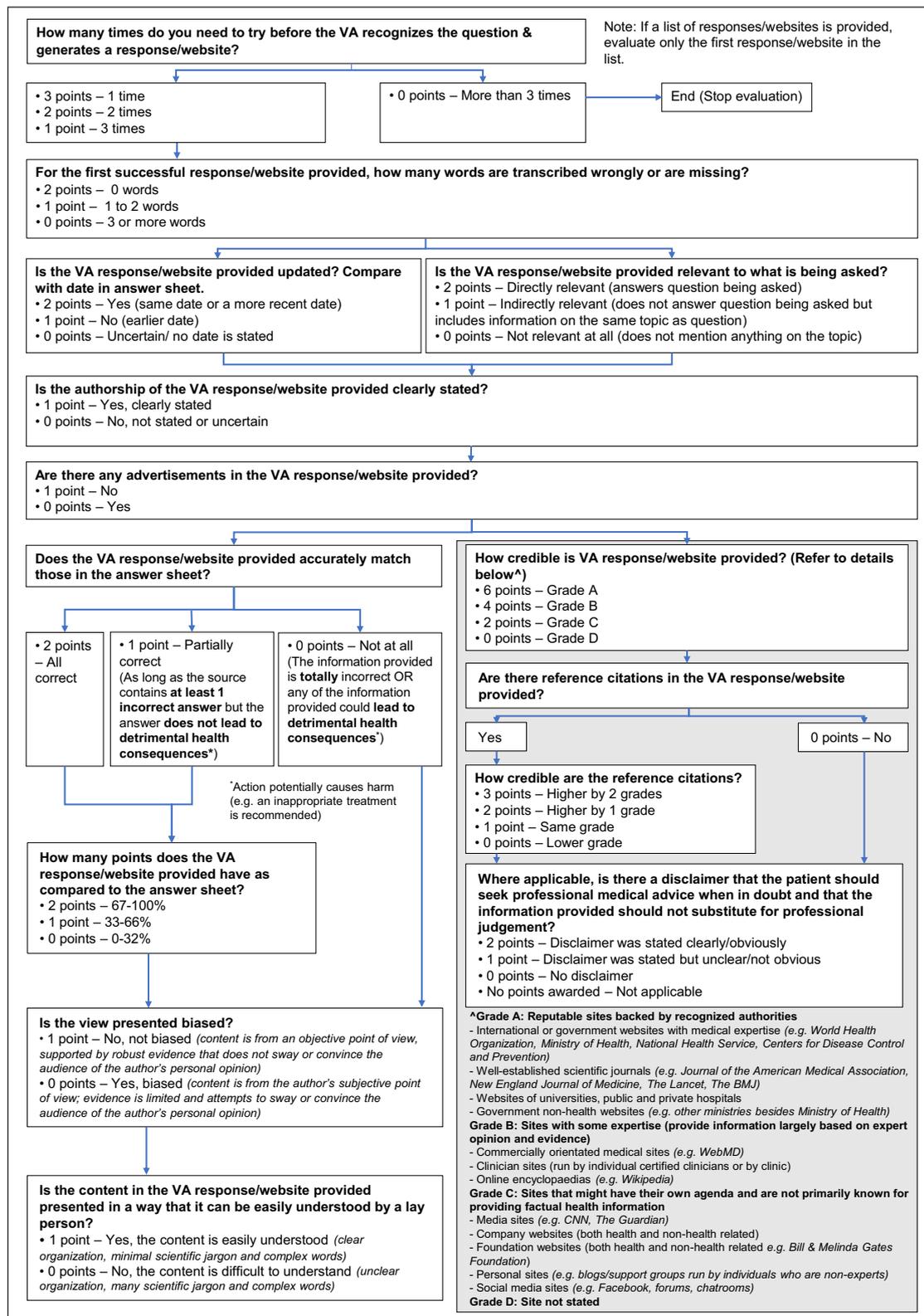


Figure 1 Evaluation rubric for assessing the voice assistants (VAs) used in this study.

was evaluated through the voice assistants' ability to recognize the question posed and provide a response. If the voice assistant was unable to provide a response after 3 attempts, the evaluation would end with zero points awarded. A successful response was further evaluated through the number of wrongly transcribed or missing words. Applicability of information was evaluated based on how updated and relevant the response was to the question.

Reliability was evaluated based on 3 criteria: transparency, presence of bias and credibility. Transparency was assessed based on whether the authorship of the response was clearly stated, and whether there were any advertisements. Biasness was defined as information provided from the author's subjective point of view, having limited evidence and attempting to sway or convince the audience of the author's personal opinion. Credibility was assessed according to 4 grading categories on the voice assistants' responses and the reference citations provided. Grade A was defined as reputable sites/references backed by recognized authorities, such as WHO, governmental websites and scientific journals. Grade B was defined as sites/references that provided information largely based on expert opinion, such as commercially orientated medical sites, clinician sites and online encyclopaedias. Grade C was defined as sites/references that might have their own agenda and were not primarily known for providing factual health information, such as social media and company websites. Grade D was used if the site/reference was not stated. In addition, the presence of a disclaimer stating that the information provided should not substitute a healthcare professional's advice/professional judgement would be evaluated for questions relating to consumer health advice, treatment, and special populations.

Accuracy was assessed through comparing the voice assistants' responses with our list of compiled answers ([Appendix A](#)). Answers that were totally incorrect or would lead to detrimental health consequences were awarded zero points, while partially or fully correct answers were awarded 1 and 2 points respectively. Comprehensiveness was determined based on the proportion of information provided by the voice assistant matched against our list of compiled answers. User-friendliness was assessed based on the understandability of the response by a layperson, with a clear organization of content and minimal scientific jargon and complex words.

The rubric was reviewed by 3 individuals (WLL, KY and QX). One of them (QX) pilot-tested the rubric using Google Assistant with 2 questions from each category in our compiled question list ([Appendix A](#)). The feedback obtained was used to refine the rubric for the final evaluation.

## 2.4 EVALUATION

Two independent evaluators (AG, female and JB, male) assessed the voice assistants using the same devices with the search history reset before and after each evaluator's use. All devices' languages were set as English (US) and the location function was switched off. For each question, the evaluator would score the voice assistant's response based on the evaluation rubric. If more than one weblink was provided by the voice assistant, the first weblink was evaluated. For each evaluator, after all responses were scored, each question's score was converted to a percentage and the mean percentage across all the questions was taken as that evaluator's score for the voice assistant. This was repeated for all voice assistants.

## 2.5 ANALYSIS

Descriptive statistics were used to report the proportion of successful responses and the cited sources by the voice assistants. These proportions were reported separately for each evaluator. Evaluation scores for the voice assistants were reported as a mean of both evaluators' scores. Normality (Shapiro-Wilk) tests were performed, and the data was analyzed at a significance level of 0.05 on the Statistical Package for Social Sciences (SPSS) software (version 25). Independent samples t-tests were used for comparing smartphone-based voice assistants and voice assistants on other platforms, and voice assistants accessing Bing versus those accessing Google search engines. Wilcoxon rank-sum tests were used to compare the comprehension abilities across genders for each voice assistant. Pearson's correlation coefficient was used to determine correlation between the percentage of successful responses and the comprehension abilities of the voice assistants.

## 3. RESULTS

The number of successful responses for the 56 COVID-19 questions differed across the voice assistants ([Table 1](#)). Google Assistant achieved the highest proportion of successful responses (97.3%), while Siri and Bixby were the other two voice assistants that achieved more than 90% of successful responses. Cortana had the lowest proportion of successful responses (22.4%).

Google Assistant had the highest score (84.0%), followed by Siri (80.6%). Bixby performed the worst of all the smartphone-based voice assistants (65.8%). On the other hand, Google Home performed the best out of the non-smartphone voice assistants (60.7%) when compared to Alexa (43.1%) and Cortana (13.3%). Smartphone-based voice assistants (Google Assistant, Siri and Bixby) had higher mean scores than the voice assistants on other platforms (Alexa, Google Home

and Cortana) (76.8% versus 39.1%,  $p = 0.064$ ). Google Assistant often responded verbally in short paragraphs (34/54, 63.0% for Evaluator 1; 24/55, 43.6% for Evaluator 2), but Siri would often only provide short verbal responses accompanied by weblinks, such as “I found this on the web” or “Here’s what I found” (53/55, 96.4% for Evaluator 1; 47/50, 94.0% for Evaluator 2).

Google Home and Google Assistant often provided similar responses/websites to the questions posed (20/56, 35.7% for Evaluator 1; 19/56, 33.9% for Evaluator 2). However, Google Home often responded with “Sorry, I don’t have any information about that. But I found something related.” and would then offer another question related to what the user had asked (17/56, 30.4% for both evaluators). The scores for Google

Assistant were consistently higher than Google Home for each of the evaluation criteria (Figure 2).

Google Assistant consistently scored the best in all the evaluation criteria (Figure 3). In terms of relevance, Google Assistant scored the highest for its comprehension ability (92.0%) and applicability of information (87.3%), followed by Siri (comprehension ability 88.8%, applicability of information 86.6%) (Figure 3a). A statistically significant positive correlation was observed between the proportion of successful responses provided by the voice assistants and their comprehension ability ( $r = 0.981$ ,  $p = 0.001$ ). There was a decreasing score trend for both relevance and reliability (transparency and presence of bias) from Google Assistant, Siri, Bixby, Google Home, Alexa and Cortana (Figure 3a and 3b). Cortana was the only voice

NUMBER OF SUCCESSFUL RESPONSES (%), N = 56						
	GOOGLE ASSISTANT	APPLE SIRI	SAMSUNG BIXBY	GOOGLE HOME	AMAZON ALEXA	MICROSOFT CORTANA
Evaluator 1	54 (96.4%)	55 (98.2%)	50 (89.3%)	41 (73.2%)	35 (62.5%)	15 (26.8%)
Evaluator 2	55 (98.2%)	50 (89.3%)	51 (91.1%)	41 (73.2%)	30 (53.6%)	10 (17.9%)
Mean (%)	54.5 (97.3%)	52.5 (93.8%)	50.5 (90.2%)	41.0 (73.2%)	32.5 (58.1%)	12.5 (22.4%)

Table 1 Number of successful responses (%) provided by each voice assistant, out of 56 possible responses.

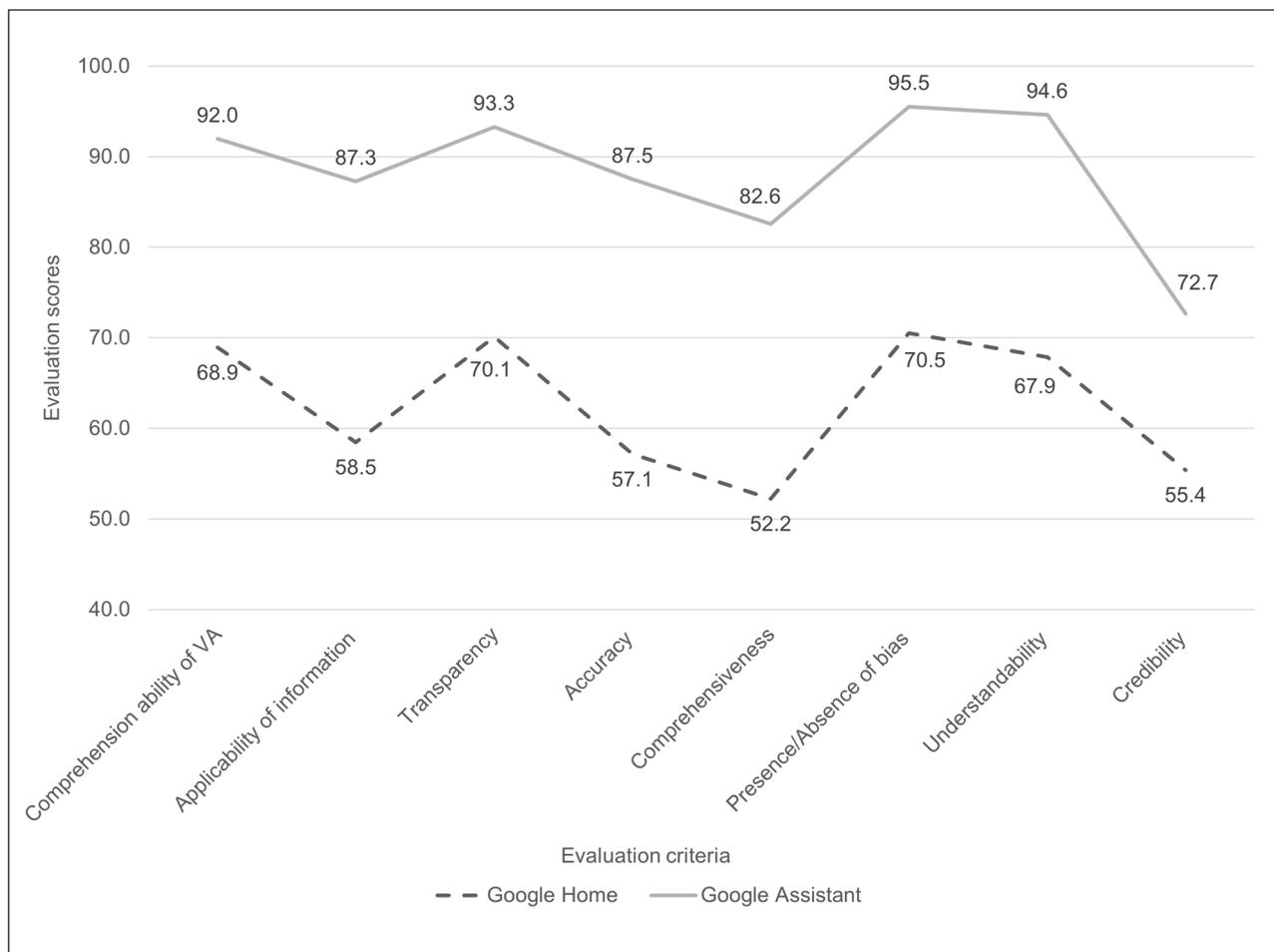
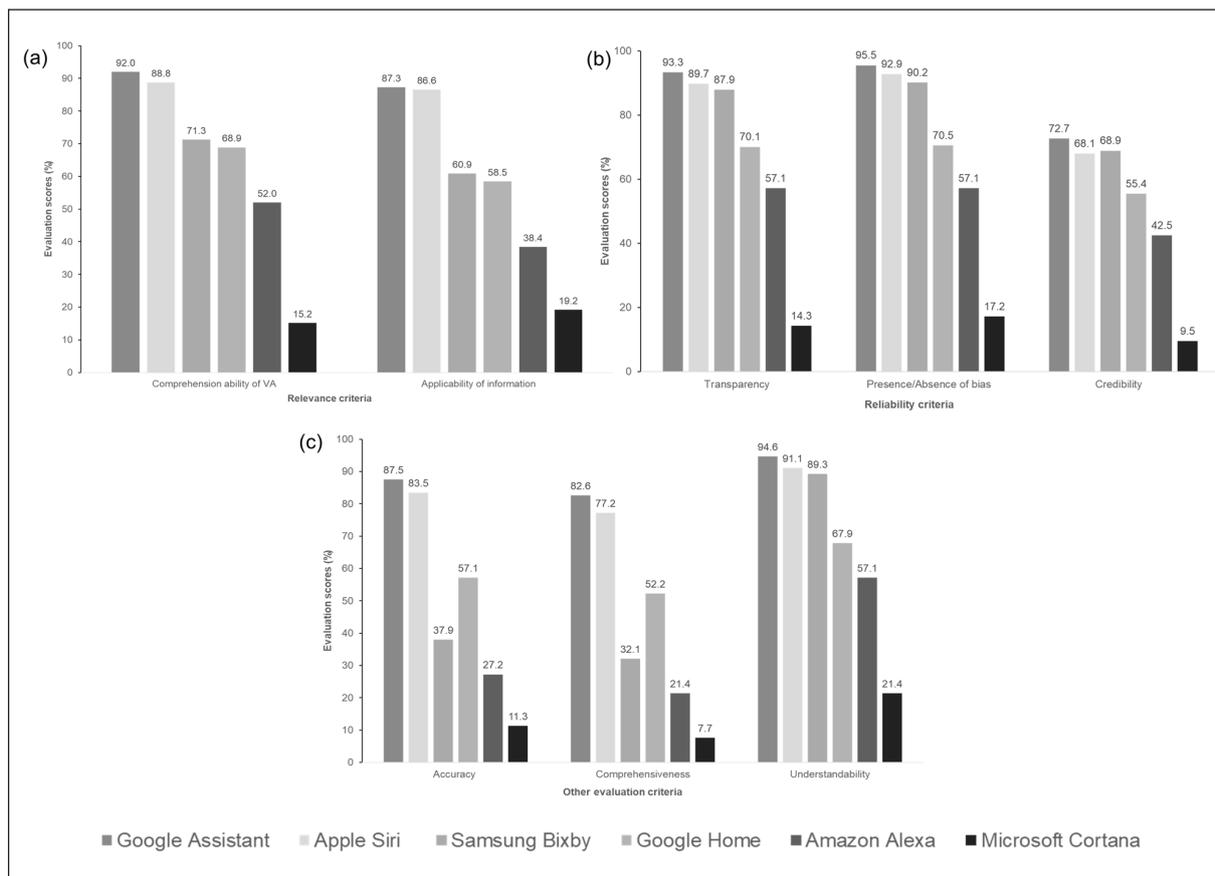


Figure 2 Evaluation scores of Google Assistant and Google Home for each criterion.



**Figure 3** Evaluation scores of the voice assistants (VAs) for each criterion.

assistant that consistently scored below 50% for all the evaluation criteria.

Bixby scored higher than Siri in terms of credibility (68.9% versus 68.1%), but lower for accuracy (37.9% versus 57.1%) and comprehensiveness (32.1% versus 52.2%) compared to Google Home (Figure 3c). Mean credibility scores were significantly lower between voice assistants that used Bing as a search engine (Alexa and Cortana, 26.0%) and the other voice assistants that used Google for searches (66.3%,  $p = 0.025$ ).

Majority of the responses by Google Assistant (45/54, 83.3% for Evaluator 1; 46/55, 83.6% for Evaluator 2) and Siri (40/55, 72.7% for Evaluator 1; 37/50, 74.0% for Evaluator 2) were from Grade A sources, such as WHO and CDC websites. Similarly, majority of responses by Bixby were also from Grade A sources (44/50, 88.0% for Evaluator 1; 45/51, 88.2% for Evaluator 2). The most cited source by Bixby was CDC (40/50, 80.0% for Evaluator 1; 40/51, 78.4% for Evaluator 2). However, Bixby also cited Grade C sources, such as news sites like the American Broadcasting Company (ABC) News (3/50, 6.0% for both evaluators). In contrast, only half of Alexa's responses were cited from CDC (18/34, 52.9% for Evaluator 1; 15/30, 50% for Evaluator 2), followed by a Grade B source – First Databank (5/34, 14.7% for both Evaluators 1 and 2). A large proportion of weblinks provided by Cortana were Grade C sources, such as

media and company sites (11/15, 73.3% for Evaluator 1; 8/10, 80% for Evaluator 2).

Comprehension abilities of the voice assistants differed between genders, but they were not significantly different (mean scores: females 67.5% versus males 61.8%,  $p = 0.738$ ). The largest difference in comprehension ability between genders occurred for Siri, whereby the median score for females (100.0%, interquartile range 100.0–100.0%) was higher than males (100.0%, interquartile range 80.0–100.0%,  $p = 0.012$ ).

## 4. DISCUSSION

Both Google Assistant and Siri performed well in the evaluation criteria, suggesting that COVID-19 information provided by these voice assistants was relevant, reliable, accurate, comprehensive, and user-friendly. The comparatively high scores that Google Assistant, Siri and Bixby achieved for transparency and credibility might have been due to their installation on smartphones, which allowed their responses to be displayed on screen. Unlike voice assistants on smart speakers, which could only provide verbal responses to the questions, the smartphone-based voice assistants enabled the authorship and reference citations to be more clearly identified.

Majority of the COVID-19 questions that were posed to the voice assistants were found on the frequently asked questions section of the government websites. However, there were two questions ([Appendix A](#), under General Information, questions 9 and 10) that were based on Google search trends. Although question 9 was not a frequently asked question on government websites, it was a common question asked by consumers in Google searches. The WHO had referred to the answers in their publication [33], hence this was included as a question to be evaluated in our study. On the other hand, question 10 was rephrased from a question from CDC as the original question was not reflective of consumers' actual search queries on Google. According to the company, Google Trends is able to categorize, aggregate and anonymize actual search requests made to Google, so that interests in particular topics can be displayed [34]. Thus, the question was adapted from Google Trends instead, since it would be more representative of how consumers would ask their questions to the voice assistants.

Google Assistant had the best comprehension ability among all the voice assistants. It also provided longer verbal responses than Siri. Our findings were similar to another study comparing the abilities of Google Assistant, Siri and Alexa in comprehending medication names [35]. The authors reported that Google Assistant had the best comprehension accuracy, while Alexa was the worst. Our study showed that there was a correlation between the comprehension ability of the voice assistants and the proportion of successful responses, thus Google Assistant might be the best voice assistant to answer COVID-19 questions posed by the general public.

Bixby performed worse than Google Assistant in terms of all the evaluation criteria. Our results were contrary to a study by Kocaballi and colleagues who reported that Bixby was second to Siri when responding appropriately to health and lifestyle prompts, and it outperformed Google Assistant and other voice assistants [8]. The difference was that in their study, Kocaballi and colleagues only evaluated the applicability of information, but did not assess the other evaluation criteria in our study, such as accuracy, comprehensiveness, user-friendliness and reliability. Our results showed that Bixby did not score as well in terms of accuracy and comprehensiveness of its responses, but also suffered in terms of providing relevant responses. During our evaluations, Bixby repeatedly produced the same generic responses when asked a variety of questions on COVID-19 ([Appendix B](#)). In this regard, Bixby's adaptability to the types of questions posed by the general public regarding the COVID-19 pandemic can be improved.

Google Assistant had consistently scored higher than Google Home in all the evaluation parameters, even though they used the same search engine. Our

findings were similar to the Kocaballi study, in which the smartphone-based voice assistants outperformed their counterparts on other platforms [8]. A possible explanation could be due to the different search algorithms and prioritization of the search results due to the different capabilities of the devices [36]. Unlike the smartphone-based Google Assistant which could provide a list of resources on screen, Google Home could only vocalize their responses. Thus, instead of answering the question directly, sometimes Google Home would pose another related question back to the user, which might not have captured the essence of the user's initial question. For example, in response to "Am I protected against COVID-19 if I had the influenza vaccine this year?", Google Home posed back the question "Do you want to know what is the mortality rate of the coronavirus disease versus influenza?" When rejected, Google Home was unable to perform any further searches, hence it scored poorly for most of the evaluation parameters. Similar to Bixby, Google Home's adaptability to the types of questions posed by users can be improved.

Alexa had provided long verbal responses (61.2 words on average per response) to the COVID-19 questions, which was similar to another study that reported that Alexa had the greatest number of spoken words in the responses compared to Siri and Google Assistant [6]. Furthermore, Alexa provided clear disclaimers in its verbal responses, thus bringing to the user's attention regarding any precautions that needed to be taken when accessing the information provided. The long verbal responses by Alexa could be an advantage to special populations who could not read small fonts on smartphones [37], such as the elderly and those with poor eyesight. More importantly, this could be beneficial to users who choose not to touch easily avoidable surfaces in the current COVID-19 pandemic [38]. However, during our evaluations, Alexa seemed to perform poorly with regards to applicability, credibility, accuracy and comprehensiveness of information on COVID-19, despite it being used in various healthcare settings [4, 39, 40]. Alexa's poor scores could be due to the differences between the Bing and Google search engines, which had different search engine optimization factors that affected the search results [41, 42]. For example, Google focuses on the quality rather than the quantity of backlinks, unlike Bing which treats both quality and quantity similarly. Furthermore, Bing favors backlinks with official domains, such as .edu, .org and .gov sites. In addition, the Google algorithm works on the context of search queries, unlike Bing, which uses targeted keywords and metadata as ranking parameters. Last, but not least, in contrast to Google searches, social media signals are used as a ranking factor in Bing searches. Since Alexa's source of answers were from Bing, the difference between its scores and those of the other voice assistants that utilized Google as a search engine

(i.e. Google Assistant, Google Home, Siri and Bixby) was expected. Nonetheless, Alexa had an algorithm embedded to identify the user’s risk for COVID-19. When prompted with questions on concerns over exposure to or having COVID-19, Alexa would start the algorithm with a prompt of “If you’re concerned about COVID-19, I can ask you a few questions based on CDC’s guidelines to help you understand your risk and make a decision about seeking medical care. Do you have a few minutes for this?” Evaluation of this algorithm found it to be thorough in identifying related symptoms along with risk factors such as age, health conditions, and close contact with infected people. However, if the user answered “no” to the prompt, Alexa would just end the process. The usefulness of this algorithm, combined with efforts from healthcare organizations such as the Mayo Clinic to further enhance Alexa’s skills in responding to COVID-19 questions [43], can potentially improve its credibility as a one-stop resource on the pandemic in time to come.

Cortana performed the worst among all the voice assistants. Besides a lack in comprehension ability, there was also a lack of reliable sources in its responses. Three-quarters of the sources provided by Cortana were Grade C sources, such as media and company sites like the ABC News, which might contain health information that lacked in completeness and accuracy [44]. Moreover, the media had been shown to present health issues in a perspective that disproportionately

emphasized risk, which might result in unnecessary heightened fear among consumers [45]. The higher selection of media sites by Cortana compared to other voice assistants could potentially also be linked to the search optimization factors of its Bing search engine instead of Google. As the search optimization factors for Bing continue to evolve [46], hopefully future pandemic-related information provided by voice assistants using Bing as a search engine would improve in terms of credibility and relevance.

Among all the parameters evaluated for voice assistants in this study, our author consensus was that even though the accuracy, credibility and comprehensiveness of pandemic-related information would have the greatest public health impact, these parameters would require a substantial amount of effort to develop, maintain and keep up-to-date, especially in relation to the rapid spread of the infodemic (Figure 4). On the other hand, understandability, comprehension ability and applicability of information could be “quick wins” if these parameters could be tailored towards a pandemic-related situation, so as to increase public awareness regarding the pandemic, as well as enhance the user-friendliness of the voice assistants. In contrast, while little effort is needed to improve the transparency and biasness of the voice assistants, these would only be useful if the other evaluation parameters were enhanced. As such, developers are encouraged to prioritize the features of

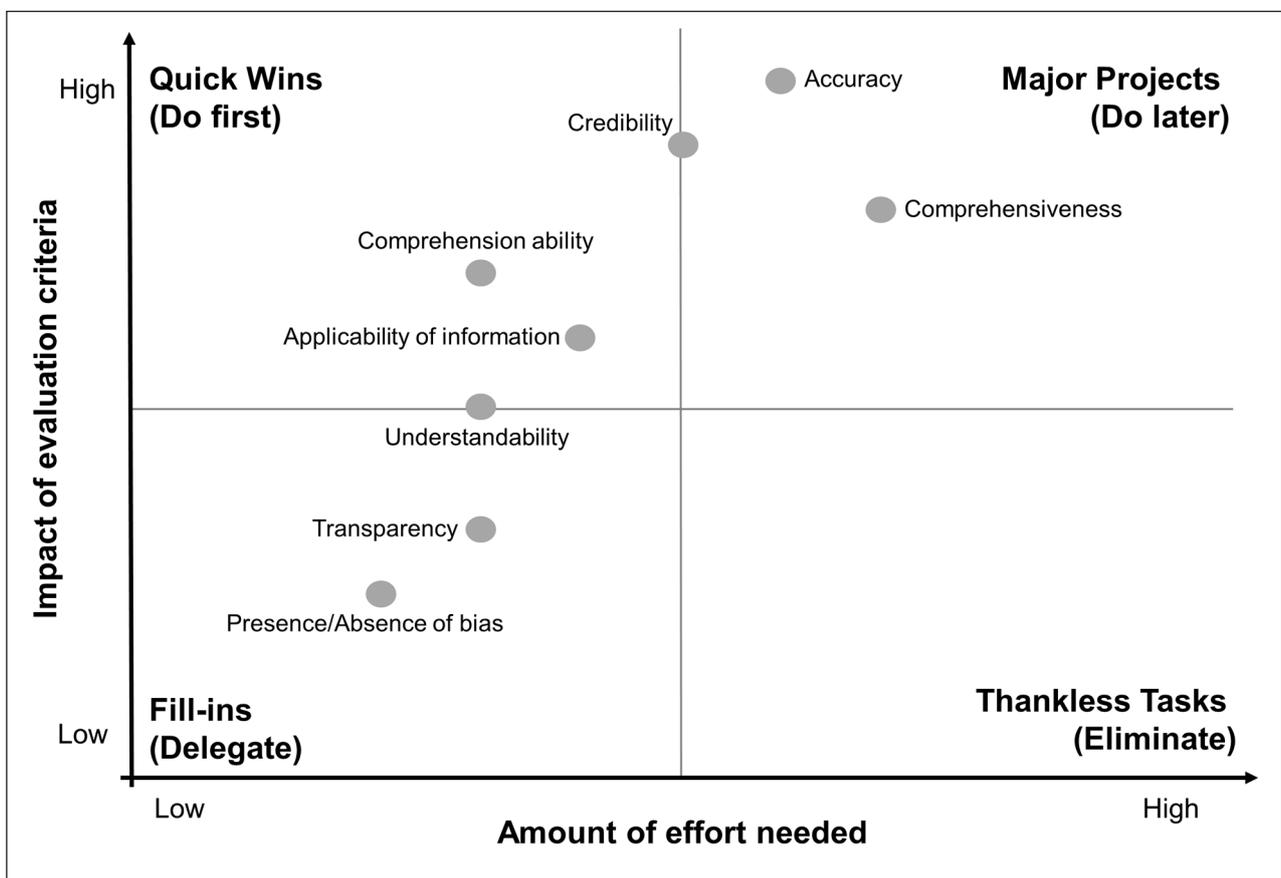


Figure 4 Action priority matrix for each evaluation criterion.

voice assistants according to their societal impact and amount of effort needed to develop these features in pandemic-related situations, such as COVID-19.

## 5. LIMITATIONS AND FUTURE WORK

As this study was conceived due to the rapidly evolving nature of the COVID-19 infodemic, the evaluation framework has not been validated. The information on COVID-19 is continually changing with new and updated information, thus we were not able to evaluate the quality of information longitudinally as it would also change over time. Our author consensus was that it would be timely to create public awareness regarding the quality of voice assistants during this crucial time in order to combat the infodemic on COVID-19. As such, we intend to validate this framework for pandemic-related information as part of future research. Another limitation was that the evaluation process might not have accurately mimicked the questioning process of an average consumer's usage of a voice assistant. If the voice assistant did not understand the question on the first attempt, a total of three attempts would be made by the evaluator and any successful response provided out of the three attempts would be evaluated. In reality, consumers might have given up on their first attempt and the voice assistant would have failed to provide the appropriate information required. Although the location feature was switched off, the responses provided by the voice assistants could still have been adapted to suit Singapore's local context where the evaluation was conducted, as the Internet Protocol address of the devices might have been used to provide the results [47, 48]. Hence, caution is advised when extrapolating the results of this study to other countries where the devices might provide different responses. Lastly, Chinese voice assistants were excluded. Given that the COVID-19 virus was first reported in China [49] and that Chinese voice assistants occupy a large part of the voice assistant market [50], future studies should also consider evaluating these voice assistants for pandemic-related information.

## 6. CONCLUSION

This study identified that Google Assistant and Siri were the best voice assistants in providing consumers with pandemic-related information about COVID-19. Consumers need to be discerning when obtaining health-related information from voice assistants, including examining the sources of information cited by the voice assistants. On the other hand, developers should also continue to enhance the skills of voice assistants in order to ensure that the information provided to consumers is reliable, accurate, comprehensive, user-friendly and relevant.

## ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Appendix A.** COVID-19 questions posed to voice assistants. DOI: <https://doi.org/10.29337/ijdh.25.s1>

## ACKNOWLEDGEMENTS

The authors would like to thank Mr Qihuang Xie for reviewing the evaluation rubric and pilot testing the rubric with a subset of COVID-19 questions; Mr Jerome Yan Heng Boon for assisting with the evaluation of the voice assistants; Mr Jiayi Loh, Ms Gwyneth Ang and Ms Clariis Yi Ning Woon for loaning the devices for evaluation of the voice assistants (Google Home, Echo Dot, Samsung Galaxy S8 respectively).

## COMPETING INTERESTS

The authors have no competing interests to declare.

## AUTHOR CONTRIBUTION

KY and LLW conceived and designed the study. AG conducted the study and analyzed the results. KY, LLW and AG wrote and revised the manuscript. All authors agreed to the publication of the manuscript.

## AUTHOR AFFILIATIONS

**Alysee Shin Ying Goh**  [orcid.org/0000-0002-2357-3122](https://orcid.org/0000-0002-2357-3122)

Department of Pharmacy, Faculty of Science, National University of Singapore, Block S4A, Level 2, 18 Science Drive 4, Singapore 117543

**Li Lian Wong**  [orcid.org/0000-0003-0649-2010](https://orcid.org/0000-0003-0649-2010)

Department of Pharmacy, Faculty of Science, National University of Singapore, Block S4A, Level 2, 18 Science Drive 4, Singapore 117543

**Kevin Yi-Lwern Yap**  [orcid.org/0000-0001-7322-4396](https://orcid.org/0000-0001-7322-4396)

Department of Public Health, School of Psychology and Public Health, La Trobe University, Melbourne (Bundoora), Victoria 3086, Australia

## REFERENCES

1. **Tankovska H.** Number of digital voice assistants in use worldwide 2019–2024 (in billions). <https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/>, 2020 (accessed 2 Jan 2021).
2. **Kinsella B, Mutchler A.** Voice assistant consumer adoption report – November 2018. <https://voicebot.ai/wp-content/uploads/2018/11/voice-assistant-consumer-adoption-report-2018-voicebot.pdf>, 2018 (accessed 2 Jan 2021).

3. **Orbita.** Voice assistant consumer adoption report for healthcare 2019. <https://voicebot.ai/voice-assistant-consumer-adoption-report-for-healthcare-2019/>, 2019 (accessed 2 Jan 2021).
4. **Leibler S.** Cedars-Sinai taps Alexa for smart hospital room pilot. <https://www.cedars-sinai.org/newsroom/cedars-sinai-taps-alexa-for-smart-hospital-room-pilot/>, 2019 (accessed 2 Jan 2021).
5. **Boyd M, Wilson N.** Just ask Siri? A pilot study comparing smartphone digital assistants and laptop Google searches for smoking cessation advice. *PLoS One.* 2018; 13: e0194811. DOI: <https://doi.org/10.1371/journal.pone.0194811>
6. **Alagha EC, Helbing RR.** Evaluating the quality of voice assistants' responses to consumer health questions about vaccines: An exploratory comparison of Alexa, Google Assistant and Siri. *BMJ Health Care Inform.* 2019; 26. DOI: <https://doi.org/10.1136/bmjhci-2019-100075>
7. **Miner AS, Milstein A, Schueller S, Hegde R, Mangurian C, Linos E.** Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Intern. Med.* 2016; 176: 619–625. DOI: <https://doi.org/10.1001/jamainternmed.2016.0400>
8. **Kocaballi AB, Quiroz JC, Rezaadegan D, Berkovsky S, Magrabi F, Coiera E,** et al. Responses of conversational agents to health and lifestyle prompts: Investigation of appropriateness and presentation structures. *J. Med. Internet Res.* 2020; 22: e15823. DOI: <https://doi.org/10.2196/15823>
9. **National Public Radio Inc.** Edison Research, The smart audio report. <https://www.nationalpublicmedia.com/insights/reports/smart-audio-report/>, 2020 (accessed 2 Jan 2021).
10. **Cuan-Baltazar JY, Munoz-Perez MJ, Robledo-Vega C, Perez-Zepeda MF, Soto-Vega E.** Misinformation of COVID-19 on the Internet: Infodemiology study. *JMIR Public Health Surveill.* 2020; 6: e18444. DOI: <https://doi.org/10.2196/18444>
11. **Ting DSW, Carin L, Dzau V, Wong TY.** Digital technology and COVID-19. *Nat. Med.* 2020; 26: 459–461. DOI: <https://doi.org/10.1038/s41591-020-0824-5>
12. **Miner AS, Laranjo L, Kocaballi AB.** Chatbots in the fight against the COVID-19 pandemic. *NPJ Digit Med.* 2020; 3: 65. DOI: <https://doi.org/10.1038/s41746-020-0280-0>
13. **Sezgin E, Huang Y, Ramtekkar U, Lin S.** Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. *NPJ Digit Med.* 2020; 3: 122. DOI: <https://doi.org/10.1038/s41746-020-00332-0>
14. **Chia PY, Coleman KK, Tan YK, Ong SWX, Gum M, Lau SK,** et al., Detection of air and surface contamination by SARS-CoV-2 in hospital rooms of infected patients. *Nat Commun.* 2020; 11: 2800. DOI: <https://doi.org/10.1038/s41467-020-16670-2>
15. **Porter J.** Apple's Siri voice assistant now provides coronavirus advice. <https://www.theverge.com/2020/3/23/21190600/apple-siri-coronavirus-covid-19-symptoms-healthcare-advice-information, 2020> (accessed 2 Jan 2021).
16. **Amazon Inc.** Helpful things Alexa can do during COVID-19. <https://www.amazon.com/b?ie=UTF8&node=21155035011, 2020> (accessed 2 Jan 2021).
17. **World Health Organization.** Q&As on COVID-19 and related health topics. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub, 2020> (accessed 23 Sep 2020).
18. **US Centers for Disease Control and Prevention.** Frequently Asked Questions. <https://www.cdc.gov/coronavirus/2019-ncov/faq.html, 2020> (accessed 23 Sep 2020).
19. **UK National Health Service.** Coronavirus (COVID-19). <https://www.nhs.uk/conditions/coronavirus-covid-19/, 2020> (accessed 23 Sep 2020).
20. **European Centre for Disease Prevention and Control.** Questions and answers on COVID-19. <https://www.ecdc.europa.eu/en/covid-19/questions-answers, 2020> (accessed 23 Sep 2020).
21. **Government of Canada.** Coronavirus disease (COVID-19). <https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19.html, 2020> (accessed 23 Sep 2020).
22. **Australian Government Department of Health.** What you need to know about coronavirus (COVID-19). <https://www.health.gov.au/news/health-alerts/novel-coronavirus-2019-ncov-health-alert/what-you-need-to-know-about-coronavirus-covid-19, 2020> (accessed 23 Sep 2020).
23. **Government of Karnataka.** Detail question and answers on COVID-19 for public. <https://www.mohfw.gov.in/pdf/FAQ.pdf, 2020> (accessed 2 Jan 2021).
24. **Ministry of Health Singapore.** FAQs on the COVID-19 situation. <https://www.moh.gov.sg/covid-19/faqs, 2020> (accessed 23 Sep 2020).
25. **National Centre for Infectious Diseases Singapore.** Viral pneumonia due to COVID-19 – Frequently Asked Questions (FAQ). <https://www.ncid.sg/Health-Professionals/Diseases-and-Conditions/Documents/FAQ%20on%20Viral%20Pneumonia%20due%20to%20Novel%20Coronavirus.pdf, 2020> (accessed 23 Sep 2020).
26. **AskDr.** Coronavirus COVID-19. <https://www.askdr.co/spaces/coronavirus-covid-19, 2020> (accessed 28 Aug 2020).
27. **Patient Platform Limited.** Coronavirus (COVID-19). <https://patient.info/forums/discuss/browse/coronavirus-covid-19--4541, 2020> (accessed 28 Aug 2020).
28. **MedHelp.** Coronavirus Community. <https://www.medhelp.org/forums/Coronavirus/show/2203, 2020> (accessed 28 Aug 2020).
29. **Answer The Public.** Covid-19 en-au - Results for COVID-19. <https://answerthepublic.com/reports/e3d1bd1b-be1b-4643-95f5-a5d5b2a00fe6, 2020> (accessed 20 Sep 2020).
30. **Charnock D.** The DISCERN Handbook: Quality criteria for consumer health information on treatment choices. Abingdon, Oxon, UK: Radcliffe Medical Press, 1998. <http://www.discern.org.uk/discern.pdf> (accessed 2 Jan 2021).

31. **Health On the Net.** HONcode guidelines: Find the guidelines for the certification of health website, the HONcode. <https://www.hon.ch/en/guidelines-honcode.html#guidelinesdetail>, 2020 (accessed 2 Jan 2021).
32. **Laversin S, Baujard V, Gaudinat A, Simonet MA, Boyer C.** Improving the transparency of health information found on the internet through the HONcode: A comparative study. *Stud. Health Technol. Inform.* 2011; 169: 654–658.
33. **World Health Organization.** WHO Director-General's opening remarks at the media briefing on COVID-19 – 24 February 2020. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---24-february-2020>, 2020 (accessed 23 Sep 2020).
34. **Google.** FAQ about Google Trends data. [https://support.google.com/trends/answer/4365533?hl=en&ref\\_topic=6248052](https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052) (accessed 11 Feb 2021).
35. **Palanica A, Thommandram A, Lee A, Li M, Fossat Y.** Do you understand the words that are comin outta my mouth? Voice assistant comprehension of medication names. *NPJ Digit Med.* 2019; 2: 55. DOI: <https://doi.org/10.1038/s41746-019-0133-x>
36. **Google.** How Google Assistant helps you get things done. <https://developers.google.com/assistant/howassistantworks/responses> (accessed 2 Jan 2021).
37. **Hoy MB.** Alexa, Siri, Cortana, and more: An introduction to voice assistants. *Med. Ref. Serv. Q.* 2018; 37: 81–88. DOI: <https://doi.org/10.1080/02763869.2018.1404391>
38. **Ozdemir S, Ng S, Chaudhry I, Finkelstein EA.** Adoption of preventive behaviour strategies and public perceptions about COVID-19 in Singapore. *Int J Health Policy Manag.* 2020; Online ahead of print. DOI: <https://doi.org/10.34172/ijhpm.2020.199>
39. **Omron Healthcare Inc.** Ask Alexa about your blood pressure readings. <https://omronhealthcare.com/alexa/>, 2020 (accessed 28 Aug 2020).
40. **Mayo Clinic.** Skills from Mayo Clinic. <https://www.mayoclinic.org/voice/apps>, 2020 (accessed 28 Aug 2020).
41. **Theuring J.** Bing vs Google: Search engine comparison. <https://www.impression.co.uk/blog/bing-differ-google/>, 2020 (accessed 28 Aug 2020).
42. **Ford D.** SEO differences between Google and other search engines. <https://curatti.com/seo-differences-various-search-engines/>, 2019 (accessed 9 Feb 2021).
43. **Eddy N.** Mayo Clinic adds COVID-19 skills to Amazon Alexa. <https://www.healthcareitnews.com/news/mayo-clinic-adds-covid-19-skills-amazon-alexa>, 2020 (accessed 4 Jan 2021).
44. **Wilson A, Bonevski B, Jones A, Henry D.** Media reporting of health interventions: Signs of improvement, but major problems persist. *PLoS One.* 2009; 4: e4831. DOI: <https://doi.org/10.1371/journal.pone.0004831>
45. **Berry TR, Wharf-Higgins J, Naylor PJ.** SARS wars: An examination of the quantity and construction of health information in the news media. *Health Commun.* 2007; 21: 35–44. DOI: <https://doi.org/10.1080/10410230701283322>
46. **Lincoln JE.** Bing SEO vs. Google SEO (Updated 2020). <https://ignitevisibility.com/how-is-bing-seo-different-than-google-seo/>, 2020 (accessed 4 Jan 2020).
47. **Google.** How IP addresses work on Google. <https://support.google.com/websearch/answer/1696588> (accessed 28 Aug 2020).
48. **Google.** Manage your Android device's location settings. [https://support.google.com/accounts/answer/3467281?hl=en&ref\\_topic=7189122](https://support.google.com/accounts/answer/3467281?hl=en&ref_topic=7189122) (accessed 28 Aug 2020).
49. **World Health Organization.** Disease outbreak news: Pneumonia of unknown cause – China. <https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/>, 2020 (accessed 28 Aug 2020).
50. **Tankovska H.** Global smart speaker market share 2018 and 2019, by platform. <https://www.statista.com/statistics/1005558/worldwide-smart-speaker-market-share/#:~:text=In%202019%2C%20Amazon's%20Alexa%20is,with%20a%2031.4%20percent%20share>, 2020 (accessed 2 Jan 2021).

---

#### TO CITE THIS ARTICLE:

Goh ASY, Wong LL, Yap KY-L. Evaluation of COVID-19 Information Provided by Digital Voice Assistants. *International Journal of Digital Health.* 2021; 1(1): 3, 1–11. DOI: <https://doi.org/10.29337/ijdh.25>

Submitted: 04 January 2021

Accepted: 14 February 2021

Published: 08 March 2021

#### COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

*International Journal of Digital Health* is a peer-reviewed open access journal published by IJS Publishing Group.