ORIGINAL RESEARCH

Laryngoscope
Investigative Otolaryngology

TrachGPT: Appraisal of tracheostomy care recommendations from an artificial intelligent Chatbot

Oluwatobiloba Ayo-Ajibola BS[1] | **Ryan J. Davis BS**[1] | **Matthew E. Lin MD**[2] | **Neelaysh Vukkadala MD**[2] | **Karla O'Dell MD**[3] | **Mark S. Swanson MD**[3] | **Michael M. Johns III MD**[3] | **Elizabeth A. Shuman MD**[3]

[1]Keck School of Medicine of the University of Southern California, Los Angeles, California, USA

[2]Department of Head and Neck Surgery, University of California Los Angeles, Los Angeles, California, USA

[3]Department of Otolaryngology-Head and Neck Surgery, University of Southern California Caruso, Los Angeles, California, USA

Correspondence
Elizabeth A. Shuman, USC Department of Otolaryngology-Head and Neck Surgery, Keck School of Medicine of USC, 1537 Norfolk Street, Suite 5800, Los Angeles, CA 90033, USA.
Email: elizabeth.shuman@med.usc.edu

Abstract

Objective: Safe home tracheostomy care requires engagement and troubleshooting by patients, who may turn to online, AI-generated information sources. This study assessed the quality of ChatGPT responses to such queries.

Methods: In this cross-sectional study, ChatGPT was prompted with 10 hypothetical tracheostomy care questions in three domains (complication management, self-care advice, and lifestyle adjustment). Responses were graded by four otolaryngologists for appropriateness, accuracy, and overall score. The readability of responses was evaluated using the Flesch Reading Ease (FRE) and Flesch–Kincaid Reading Grade Level (FKRGL). Descriptive statistics and ANOVA testing were performed with statistical significance set to $p < .05$.

Results: On a scale of 1–5, with 5 representing the greatest appropriateness or overall score and a 4-point scale with 4 representing the highest accuracy, the responses exhibited moderately high appropriateness (mean = 4.10, SD = 0.90), high accuracy (mean = 3.55, SD = 0.50), and moderately high overall scores (mean = 4.02, SD = 0.86). Scoring between response categories (self-care recommendations, complication recommendations, lifestyle adjustments, and special device considerations) revealed no significant scoring differences. Suboptimal responses lacked nuance and contained incorrect information and recommendations. Readability indicated college and advanced levels for FRE (Mean = 39.5, SD = 7.17) and FKRGL (Mean = 13.1, SD = 1.47), higher than the sixth-grade level recommended for patient-targeted resources by the NIH.

Conclusion: While ChatGPT-generated tracheostomy care responses may exhibit acceptable appropriateness, incomplete or misleading information may have dire

clinical consequences. Further, inappropriately high reading levels may limit patient comprehension and accessibility. At this point in its technological infancy, AI-generated information should not be solely relied upon as a direct patient care resource.

KEYWORDS

artificial intelligence, education, head and neck cancer, patient knowledge, tracheostomy

1 | INTRODUCTION

Home tracheostomy management requires a unique collaboration between healthcare providers and patients. Patients with tracheostomy dependence encounter a multitude of challenges functionally, physically, and psychosocially. From cannulation to discharge and home care, patients may experience deficits in their ability to breathe, communicate, and obtain nutrition, which can negatively affect their well-being and quality of life.[1–6]

Moreover, complications occur in 11%–40% of tracheostomies, with these including but not limited to bleeding, blockage or displacement of the cannula, tracheoesophageal fistula, and excessive scarring.[7–11] Other common issues include mucus plugging, difficulty talking or swallowing, and accidental decannulation, often due to coughing. These complications comprise a critical element of post-discharge tracheostomy care, and as patients navigate these challenges, they are likely to seek alternative sources of information beyond appointments with medical professionals.

Over the past two decades, there has been a significant rise in patients turning to the Internet as an initial source of health information, often prioritizing it over direct contact with healthcare professionals.[12] In the United States alone, over half of US adults polled in 2019 reported utilizing search engines and social media for health-related purposes, underscoring the immense power of the internet in guiding patient health inquiries.[13] Moreover, those in poorer health are at an increased likelihood of online health information seeking.[14,15]

ChatGPT, an internet-based natural language processor (NLP) developed by OpenAI, was released in December 2022.[16] Since then, its user base has grown to over 1 billion users as many are attracted to its ability to synthesize complex information and provide personalized responses in their native language.[17] Emerging as a competitor to popular internet search engines, ChatGPT's exponential growth signifies its potential to revolutionize the way individuals seek information, particularly in reshaping how patients access medical knowledge.[18] Thus, it is likely that tracheostomy patients may also turn to this tool as a source of tracheostomy care information.

Recognizing the widespread adoption of ChatGPT and expressing concern with the danger of potentially incorrect recommendations, the medical community has actively delved into investigating the capabilities of ChatGPT in addressing patient medical information needs, with various studies showing a high degree of accuracy within responses.[19,20] Notably, studies have emerged within otolaryngology examining its efficacy in generating post-operative instructions for otolaryngology procedures and various applications in clinical otolaryngology.[21–23] Moreover, as more investigations proceed in this area of research, some critiques include the large potential for lack of specificity and comprehensiveness within responses.[24]

Despite the growing popularity of ChatGPT, there has been a notable lack of investigations examining its efficacy in generating accurate tracheostomy care recommendations in response to common patient inquiries. Thus, in this study, we aimed to characterize the clinical appropriateness, factual accuracy, and overall quality of ChatGPT-generated responses to likely tracheostomy care questions. In addition, readability scores provided an approximation of the understandability of responses. Our central hypothesis was that ChatGPT responses, while potentially of moderate to high accuracy, may lack the specificity and nuance required in the potentially critical nature of advising tracheostomy home care, thus impacting the response appropriateness and overall quality.

2 | METHODS

This cross-sectional cohort study was deemed IRB-exempt by the Institutional Review Board of the University of Southern California. Ten inquiries regarding tracheostomy care were created based on the clinical experiences of a board-certified otolaryngologist as well as the "frequently asked questions" section of the Johns Hopkins Tracheostomy website, a popular source of information from a high-volume head and neck cancer center.[25]

Groups of three questions were semantically categorized as "complication symptom," which modeled a patient asking for advice secondary to the rise of new symptoms, and "self-care advice," which modeled questions surrounding suction, humidification, and obstruction management. Finally, "lifestyle change" questions focused on the effect of tracheotomy on patient activities, such as activities to avoid and any possible side effects to expect soon after discharge. Additionally, a separate inquiry related explicitly to Passy Muir valve care was categorized as a "special circumstance."

After review by a board-certified otolaryngologist, a medical student researcher (OAA) inputted each inquiry into a new ChatGPT inquiry session. This approach prevented the development of progressive learning abilities through knowledge from past inquiries and allowed for closer modeling of initial patient contact with ChatGPT. The same researcher recorded only the first generated outputs.

TABLE 1 Laryngologist assessment and readability of ChatGPT tracheostomy care responses.

	Question	Physician assessment			Readability (FRE)	
		Mean appropriateness (Mean, SD)	Mean accuracy (mean, SD)	Mean overall score (Mean, SD)	Ease	Level
Complications	Hemoptysis	4.00 (0.82)	3.75 (0.50)	4.00 (0.82)	38.6	13.8
	Mucus plug	3.25 (0.96)	3 (0.00)	3.00 (1.15)	42	11.9
	Fluid issues	4.75 (0.50)	3.75 (0.50)	4.75 (0.50)	37.4	14.9
	Category total	4.00 (0.95)	3.50 (0.52)	3.92 (1.08)	39.33 (2.39)	13.53 (1.52)
Self-care	Humidification	4.50 (0.58)	3.75 (0.50)	4.25 (0.5)	31.8	14.1
	Suction	4.50 (0.58)	3.75 (0.50)	4.25 (0.96)	37.1	12
	Dislodgement	4.00 (0.92)	3.25 (0.50)	4.00 (0.82)	50.6	10.7
	Category total	4.33 (0.65)	3.58 (0.51)	4.17 (0.72)	39.83 (9.69)	12.27 (1.72)
Lifestyle	Precautions	3.75 (1.26)	3.25 (0.50)	2.75 (0.96)	32.3	13.5
	Side effects	4.25 (0.50)	3.5 (0.58)	4.00 (0.00)	37.7	13.3
	Smoking	3.75 (1.50)	3.75 (0.50)	4.00 (1.54)	34.5	14
	Category total	3.67 (0.79)	3.33 (0.50)	3.67 (0.71)	34.83 (2.72)	13.6 (0.36)
Special circumstances	Passy-Muir	4.25 (0.96)	3.75	4.25	53.0	10.7
	Total (Mean, SD)	4.10 (0.90)	3.55 (0.50)	4.02 (0.86)	39.5 (7.17)	13.1 (1.47)
		Scored 1–5 (5 as most appropriate)	Scored 1–4 (4 as highest accuracy)	Scored 1–5 (5 as highest score)		

ChatGPT Model 3.5 was queried on March 28, 2023 to create these responses (Appendix S1).

Four board-certified otolaryngologists from two tertiary academic centers assessed the responses for appropriateness, factual accuracy, and overall quality. Clinical appropriateness, defined as the suitability of the response for the request at hand, was graded on a 5-point Likert scale from 5 (very appropriate) to 1 (very inappropriate). Factual accuracy measures the number of factual errors in the response, from 4 (no factual errors) to 1 (totally inaccurate). The overall quality of the response was graded on an A (5) to F (1) scale. For appropriateness and quality, scores were categorized as high (4.21–5), moderately high (3.41–4.20), moderate (2.61–3.40), moderately low (1.81–2.60), and low (1–1.80). Factual accuracy scores were categorized as low (1, 2), moderate (2.01–3), and high (3.01–4).

For sub-A overall quality, respondents were given a choice of six multiple-choice options: lack of nuance, incorrect information, incorrect recommendation, irrelevant information, not specific to the request, or other. The selection of the options above also exposed free-response questions in which question-specific reasons for deficits were elicited. Readability was evaluated using the Flesch Reading Ease (FRE) and Flesch–Kincaid Reading Grade Level (FKRGL) algorithms. While both use average sentence length and syllabus per word for calculation, the FKRGL is correlated to an equivalent grade level. In contrast, FRE scores are graded along a scale in which lower scores indicate a higher education level. Although ChatGPT is capable of providing answers of a desired readability level based on a prompt, no such prompts were utilized in this study to mimic patient interactions as readability specification is not likely in these scenarios.

Data analysis was performed using Microsoft Excel (version 16.71, Microsoft Software, Richmond, Washington) and IBM SPSS Statistics for Macintosh, version 28.0.0.0 (IBM Corp., Armonk, N.Y., USA). Descriptive statistics were used to characterize the data cohort as means and standard deviations were used for continuous variables. ANOVA testing was performed to compare scores and readability between categories of inquiries. If ANOVA was significant, post-hoc testing was further analyzed for intergroup variability. Significance was set at $p < .05$.

3 | RESULTS

Overall, the ChatGPT-generated responses to tracheostomy care inquiries displayed high accuracy (mean = 3.55, SD = 0.50), and moderately high appropriateness (mean = 4.10, SD = 0.90) and overall scores (mean = 4.02, SD = 0.86) (Table 1).

With 1 representing the greatest appropriateness and overall, self-care advice inquiries exhibited the greatest appropriateness (mean = 4.33, SD = 0.65) and overall score (mean = 4.17, SD = 0.72) (Figure 1). These inquiries also scored the highest accuracy on a 4-point Likert scale, in which 4 was the highest accuracy (mean = 3.58, SD = 0.51). Lifestyle adjustment inquiries displayed the lowest appropriateness (mean = 3.67, SD = 0.79), accuracy (mean = 3.33, SD = 0.50) and overall score (mean = 3.67, SD = 0.71). Among these three categories, ANOVA testing yielded no significant differences in rating across categories and scoring dimensions (appropriateness: $p = .579$, accuracy: $p = .940$; overall:
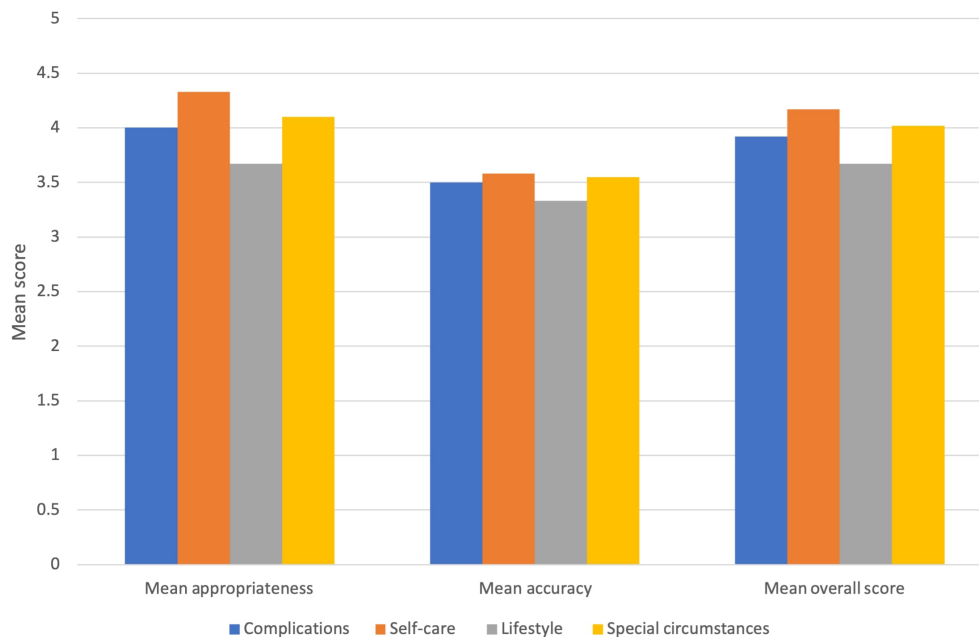
$p = .801$). The inquiry regarding Passy Muir valve care exhibited higher than average appropriateness (4.25), accuracy (3.75), and overall score (4.25).

Average overall FRE and FKRGL scores were 39.50 (SD = 7.17) and 13.10 (SD = 1.47) respectively. The category with the highest readability ease, denoted by higher scores corresponding with lower education level, was the self-care inquiry category (mean = 39.83, SD = 9.69), whereas the lowest ease was the lifestyle adjustment category (mean = 34.83, SD = 2.72). Other than the special circumstance inquiry, readability levels for all inquiry categories were of 12th grade or above.

Physician graders often cited a lack of nuance and incorrect information or recommendations as justification for subpar grading (Table 2).

4 | DISCUSSION

This study sought to evaluate the appropriateness, accuracy, overall quality, and readability of AI-generated responses to patient inquiries regarding tracheostomy care. Our hypothesis that, due to ChatGPT not being trained explicitly by and for medical practice, the chatbot would produce responses of poor appropriateness, accuracy, and overall quality was not supported by the results of this study. High factual accuracy, and moderately high appropriateness and overall quality of responses were displayed for complication, self-care, lifestyle, and special device consideration inquiries as evaluated by four otolaryngologists. However, categorical and free-response critiques of AI-formulated answers from the otolaryngologist-graders revealed a risk of patients receiving incomplete or unnuanced tracheostomy care information from ChatGPT. Moreover, readability levels for all responses displayed a 10th-grade or higher level, much greater than

the 5th-grade reading level of the average adult and the 6th-grade reading level recommended for patient-targeted material.[26–28] As patients with tracheostomies are prone to post-discharge complications,[1–3] and sizeable negative quality of life deficits, they are likely to seek information from alternative medical sources to address concerns and cope with their treatment.[29] This leaves them particularly vulnerable to receiving incorrect medical recommendations from AI-powered NLPs like ChatGPT.

To our knowledge, this study is among the first to evaluate the use of AI chatbot technology to address tracheostomy care inquiries. Overall quality and appropriateness were moderate, while factual accuracy was high, as evaluated by physician graders. Johnson et al., which evaluated ChatGPT for 247 medical queries across 17 specialties, found that ChatGPT provided mostly correct information to the array of inquiries as judged by physician specialists.[19] A similar study conducted by Zhu et al.[24] for prostate cancer treatment inquiries displayed the capability for ChatGPT also to respond with moderate to high accuracy, comparativeness, and overall quality, albeit with concerns surrounding specificity and comprehensiveness.

While the present study supports literature in the exhibition of ChatGPT as a source of accurate patient recommendations, it remains essential to refine AI natural language processors like ChatGPT further to increase the clinical appropriateness of recommendations for the specific inquiry. Appropriate recommendations are especially vital as complications such as infection, bleeding, or blockage of the tracheostomy may have grave outcomes if not managed properly.

While not significantly different, responses addressing self-care advice (humidification, suction, and dislodgement) were of the highest appropriateness, accuracy, and overall quality score. High-quality responses to self-care inquiries are critical as these patients are discharged with laborious, demanding, and disruptive home care tasks

TABLE 2 Examples of physician feedback for subpar responses.

Feedback category	Example of physician response	Response frequency
Lack of Nuance	"The response fails to take into account the amount of blood and the fact that the trach was recently placed which can mean that some amount of blood is normal. Depending on the volume, seeking attention immediately may be unnecessary. It would have also been reasonable to have some discussion of whether the blood came from around the stoma, through the tube or from the mouth. It could have also discussed suction trauma."	13
Incorrect information	"The response hedges when it says 'generally not recommended' which may imply in some situations that it might be ok. A more definitive answer recommending against smoking should have been made."	9
Incorrect recommendation	"I'm grading this response more harshly because it concerns a very dangerous issue and misses a key recommendation namely it should have recommended removing and changing the inner cannula (if present). This is a critical step that is not mentioned at all."	6
Irrelevant Info	"I'm not sure what passive vs. active humidification means."	2
Not specific to this issue	"The recommendations are not really specific for blood from the trach, but more hemoptysis in general so while the recommendations are not incorrect, it's just not specific."	1
Other	"Redundant info."; "Omitted Important Information."	4

and may require support as they navigate the novelty of their condition.[30,31]

Readability was analyzed as a proxy for the understandability of the responses garnered from the chatbot. While ChatGPT is capable of dispelling information at a desired readability level and ease, we aimed to model likely patient interactions by not including this language within prompts. Fleisch Kincaid Readability Grade level scores were consistently higher than the 6th-grade level recommended for patient material, a designation that is already higher than the average 5th-grade reading level of the average American adult.[26,27,32] Improving the readability of these responses is especially critical for head and neck cancer patients with tracheostomies in particular, as they exhibit lower levels of education than other cancer patients.[33]

However, the aforementioned result supports current literature that has found that general patient material, particularly those focused on tracheostomy care, far eclipses the recommended reading level.[34,35] However, the highly manipulatable nature of ChatGPT and other online, artificially intelligent chatbots allows users to "prime" the chatbot to rewrite and even formulate original answers at a desired reading level.[36] Therefore, more research is required to discern whether readability and answer quality improve when primed to achieve a set reading level.

When allowed to provide categorized and free response feedback for AI-generated responses that did not receive high-quality gradings, physician graders consistently reported a lack of nuance and incorrect information or recommendations. For example, in response to an inquiry on acute hemoptysis after a recent cannulation, ChatGPT's response gave recommendations for hemoptysis in general, as well as failing to take into account the nuance of the recency of placement and the volume of hemoptysis. Similarly, in regards to a mucus plug emergency, clinicians perceived the response to contain incorrect information and recommendations as it failed to discuss that, in an emergency, removal of the plug is recommended by anyone, not just a clinician, as well as the need to remove the inner cannula if applicable to the patient. Interestingly, this trend and these remarks contradict the high factual accuracy gradings previously mentioned; however, it is likely that despite the accuracy of the dispelled information, there was a possibility for incomplete or nonspecific recommendations. These themes are concerning as patients often use information-seeking as a source of coping and control over their diagnosis; incorrect recommendations may falsify this sense of control and result in noncompliance with physician-guided care.[37,38]

ChatGPT has the potential to disrupt the space of online medical information due to its ability to serve as a highly accessible source of personalized patient support and education.[39] As similar NLPs have improved their efficacy via training on content specific to various domains in medicine, this model is capable of rapidly improving its ability to disseminate quality medical information to patients.[19,40] However, in its present form, the model is not yet perfect at conveying clinically appropriate, accurate, and high-quality responses for every inquiry. It is important to note that the perfect online patient resource for comprehensive, specific, and nuanced tracheostomy care information may not exist and, moreover, may not be realistic for most patient-focused online resources, let alone a natural language processor like ChatGPT, which is very much in its infancy and not explicitly trained on and for medical information.

Nonetheless, otolaryngologists and other professionals involved with the care of tracheostomy patients must counsel patients on the risks of adhering solely to recommendations from ChatGPT, as mismanagement of the tracheostomy apparatus and its complications may carry great consequences.

The present study had several limitations. These responses were queried on ChatGPT, which is merely one of many other AI-powered NLPs at various stages of development. The version of ChatGPT may also significantly impact the results as the most recent spring 2023 update to version 4, albeit less readily accessible through a paid subscription, has been proclaimed more capable than version 3.5 utilized in this study. Furthermore, FRE and FKRGL measures, which measure

the word length and syllable content of sentences, may undervalue the complexity of medical terminology that is short in length but still unfamiliar to patients. Readability alone does not ensure understanding, as factors such as layout, inclusion of graphics, and overall design play a role in the effectiveness of text-based patient content.[41] Future studies may choose to utilize the Patient Education Materials Assessment Tool (PEMAT) or similar instruments to more adequately measure the understandability, while also considering results from multiple natural language processors. The use of a larger sample of inquiries and graders may also increase the power of future investigations.

5 | CONCLUSION

While ChatGPT-generated tracheostomy care responses may exhibit moderate to high levels of appropriateness, accuracy, and overall quality, there is a risk of incomplete or unnuanced information and readability levels far exceeding recommended standards. Healthcare professionals must advise patients on the potential risks of relying solely on ChatGPT recommendations for tracheostomy care.

CONFLICT OF INTEREST STATEMENT
None.

ORCID
Oluwatobiloba Ayo-Ajibola https://orcid.org/0009-0002-4824-7160
Ryan J. Davis https://orcid.org/0009-0002-0408-8380
Matthew E. Lin https://orcid.org/0000-0001-9914-9670
Elizabeth A. Shuman https://orcid.org/0009-0009-0872-869X

REFERENCES

1. Foster A. More than nothing: the lived experience of tracheostomy while acutely ill. *Intens Crit Care Nurs*. 2010;26(1):33-43. doi:10.1016/j.iccn.2009.09.004
2. Bowers B, Scase C. Tracheostomy: facilitating successful discharge from hospital to home. *Br J Nurs*. 2007;16(8):476-479. doi:10.12968/bjon.2007.16.8.23418
3. Kumar V, Malhotra V, Sinha V. Evaluation of individual Quality of Life (QOL) among patients with tracheostomy using WHO-QOL BREF Questionnaire. *Indian J Otolaryngol Head Neck Surg*. 2022;74(S3):5207-5216. doi:10.1007/s12070-020-02052-z
4. Nakarada-Kordic I, Patterson N, Wrapson J, Reay SD. A systematic review of patient and caregiver experiences with a tracheostomy. *Patient*. 2018;11(2):175-191. doi:10.1007/s40271-017-0277-1
5. Phookan J, Talukdar R. A study on quality of life in post-tracheostomised patients. *Indian J Otolaryngol Head Neck Surg*. 2023;75(2):848-856. doi:10.1007/s12070-023-03511-z
6. Terrell JE, Ronis DL, Fowler KE, et al. Clinical predictors of quality of life in patients with head and neck cancer. *Arch Otolaryngol Head Neck Surg*. 2004;130(4):401. doi:10.1001/archotol.130.4.401
7. Castling B, Telfer M, Avery BS. Complications of tracheostomy in major head and neck cancer surgery; a retrospective study of 60 consecutive cases. *Br J Oral Maxillofac Surg*. 1994;32(1):3-5. doi:10.1016/0266-4356(94)90162-7
8. Halfpenny W, McGurk M. Analysis of tracheostomy-associated morbidity after operations for head and neck cancer. *Br J Oral Maxillofac Surg*. 2000;38(5):509-512. doi:10.1054/bjom.2000.0310
9. Taicher S, Givol N, Peleg M, Ardekian L. Changing indications for tracheostomy in maxillofacial trauma. *J Oral Maxillofac Surg*. 1996;54(3):292-295. doi:10.1016/S0278-2391(96)90744-2
10. Goldenberg D, Ari EG, Golz A, Danino J, Netzer A, Joachims HZ. Tracheotomy complications: a retrospective study of 1130 cases. *Otolaryngol Head Neck Surg*. 2000;123(4):495-500. doi:10.1067/mhn.2000.105714
11. Ruohoalho J, Xin G, Bäck L, Aro K, Tapiovaara L. Tracheostomy complications in otorhinolaryngology are rare despite the critical airway. *Eur Arch Otorhinolaryngol*. 2021;278(11):4519-4523. doi:10.1007/s00405-021-06707-7
12. Swoboda CM, Van Hulle JM, McAlearney AS, Huerta TR. Odds of talking to healthcare providers as the initial source of healthcare information: updated cross-sectional results from the Health Information National Trends Survey (HINTS). *BMC Fam Pract*. 2018;19(1):146. doi:10.1186/s12875-018-0805-7
13. Goldberg R. More than half of American households used the Internet for health-related activities in 2019, NTIA Data Show|National Telecommunications and Information Administration. Accessed October 3, 2023. https://www.ntia.gov/blog/2020/more-half-american-households-used-internet-health-related-activities-2019-ntia-data-show
14. Bansil P, Keenan NL, Zlot AI, Gilliland JC. Health-related information on the web: results from the HealthStyles Survey, 2002-2003. *Prev Chronic Dis*. 2006;3(2):A36.
15. Bundorf MK, Wagner TH, Singer SJ, Baker LC. Who searches the Internet for health information? *Health Serv Res*. 2006;41(3p1):819-836. doi:10.1111/j.1475-6773.2006.00510.x
16. ChatGPT: optimizing language models for dialogue. Accessed April 15, 2023. https://openai.com/blog/chatgpt
17. Ruby D. 59+ ChatGPT statistics—users, trends & more (July 2023). *DemandSage*. 2023. Accessed July 6, 2023. https://www.demandsage.com/chatgpt-statistics/
18. Marr B. Is Google's reign over? ChatGPT emerges as a serious competitor. *Forbes*. Accessed July 6, 2023. https://www.forbes.com/sites/bernardmarr/2023/02/20/is-googles-reign-over-chatgpt-emerges-as-a-serious-competitor/
19. Johnson D, Goodman R, Patrinely J, et al. Assessing the accuracy and reliability of AI-generated medical responses: an evaluation of the chat-GPT model. In review; 2023. doi:10.21203/rs.3.rs-2566942/v1
20. Van Bulck L, Moons P. What if your patient switches from Dr. Google to Dr. ChatGPT? A vignette-based survey of the trustworthiness, value, and danger of ChatGPT-generated responses to health questions. *Eur J Cardiovasc Nurs*. 2023;23:zvad038. doi:10.1093/eurjcn/zvad038
21. Temsah O, Khan SA, Chaiah Y, et al. Overview of early ChatGPT's presence in medical literature: insights from a hybrid literature review by ChatGPT and human experts. *Cureus*. 2023;15:e37281. doi:10.7759/cureus.37281
22. Park I, Joshi AS, Javan R. Potential role of ChatGPT in clinical otolaryngology explained by ChatGPT. *Am J Otolaryngol*. 2023;44(4):103873. doi:10.1016/j.amjoto.2023.103873
23. Ayoub NF, Lee YJ, Grimm D, Balakrishnan K. Comparison between ChatGPT and Google search as sources of postoperative patient instructions. *JAMA Otolaryngol Head Neck Surg*. 2023;149(6):556-558. doi:10.1001/jamaoto.2023.0704
24. Zhu L, Mou W, Chen R. Can the ChatGPT and other large language models with internet-connected database solve the questions and concerns of patient with prostate cancer and help democratize medical knowledge? *J Transl Med*. 2023;21(1):269. doi:10.1186/s12967-023-04123-5

25. Tracheostomy. Johns Hopkins Trach. Accessed July 6, 2023. https://www.hopkinsmedicine.org/tracheostomy

26. Edmunds MR, Barry RJ, Denniston AK. Readability assessment of online ophthalmic patient information. *JAMA Ophthalmol*. 2013; 131(12):1610. doi:10.1001/jamaophthalmol.2013.5521

27. Weiss BD, Blanchard JS, McGee DL, et al. Illiteracy among Medicaid recipients and its relationship to health care costs. *J Health Care Poor Underserved*. 1994;5(2):99-111. doi:10.1353/hpu.2010.0272

28. Eltorai AEM, Ghanian S, Adams CA, Born CT, Daniels AH. Readability of patient education materials on the american association for surgery of trauma website. *Arch Trauma Res*. 2014;3(2):e18161. doi:10.5812/atr.18161

29. Mills ME, Sullivan K. The importance of information giving for patients newly diagnosed with cancer: a review of the literature. *J Clin Nurs*. 1999;8(6):631-642. doi:10.1046/j.1365-2702.1999.00296.x

30. Balfe M, Keohane K, O'Brien K, Sharp L. Social networks, social support and social negativity: a qualitative study of head and neck cancer caregivers' experiences. *Eur J Cancer Care (Engl)*. 2017;26(6):e12619. doi:10.1111/ecc.12619

31. Balfe M, Keohane K, O'Brien K, et al. In a bad place: carers of patients with head and neck cancer experiences of travelling for cancer treatment. *Eur J Oncol Nurs*. 2017;30:29-34. doi:10.1016/j.ejon.2017.07.001

32. Weiss B. Health literacy and patient safety: help patients understand. *Manual for Clinicians*. 2nd ed.; 2010. Accessed July 10, 2023. https://psnet.ahrq.gov/issue/health-literacy-and-patient-safety-help-patients-understand-manual-clinicians-2nd-ed

33. Kim C, Dillon R, Nica L, Berthelet E, Keyes M, Hamilton SN. Information needs of patients diagnosed with head and neck cancer undergoing radiation therapy: a survey of patient satisfaction. *J Canc Educ*. 2021;36(5):941-949. doi:10.1007/s13187-020-01719-z

34. Sabharwal S, Badarudeen S, Unes Kunju S. Readability of online patient education materials from the AAOS web site. *Clin Orthop Relat Res*. 2008;466(5):1245-1250. doi:10.1007/s11999-008-0193-8

35. Kong KA, Hu A. Readability assessment of online tracheostomy care resources. *Otolaryngol Head Neck Surg*. 2015;152(2):272-278. doi:10.1177/0194599814560338

36. Haver HL, Lin CT, Sirajuddin A, Yi PH, Jeudy J. Use of ChatGPT, GPT-4, and bard to improve readability of ChatGPT's answers to common questions about lung cancer and lung cancer screening. *AJR Am J Roentgenol*. 2023;221(5):701-704. doi:10.2214/AJR.23.29622

37. Ziebland S, Chapple A, Dumelow C, Evans J, Prinjha S, Rozmovits L. How the internet affects patients' experience of cancer: a qualitative study. *BMJ*. 2004;328(7439):564. doi:10.1136/bmj.328.7439.564

38. Akakpo MG. The role of care-seeking behavior and patient communication pattern in online health information-seeking behavior – a cross-sectional survey. *Pan Afr Med J*. 2022;42:124. doi:10.11604/pamj.2022.42.124.33623

39. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT – reshaping medical education and clinical management. *Pak J Med Sci*. 2023;39(2):605-607. doi:10.12669/pjms.39.2.7653

40. BioMedLM: a domain-specific large language model for biomedical text. Accessed July 10, 2023. https://www.mosaicml.com/blog/introducing-pubmed-gpt

41. Friedman DB, Hoffman-Goetz L. A systematic review of readability and comprehension instruments used for print and web-based cancer information. *Health Educ Behav*. 2006;33(3):352-373. doi:10.1177/1090198105277329

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.