

ChatGPT Responses to Frequently Asked Questions on Ménière's Disease: A Comparison to Clinical Practice Guideline Answers

Rebecca A. Ho, BA , Ariana L. Shaari, BA ,
 Paul T. Cowan, DO, MS , and Kenneth Yan, MD, PhD 

OTO Open
 2024, Vol. 8(3):e163
 © 2024 The Author(s). OTO Open
 published by Wiley Periodicals LLC
 on behalf of American Academy of
 Otolaryngology–Head and Neck
 Surgery Foundation.
 DOI: 10.1002/oto2.163
<http://oto-open.org>
 WILEY

Abstract

Objective. Evaluate the quality of responses from Chat Generative Pre-Trained Transformer (ChatGPT) models compared to the answers for “Frequently Asked Questions” (FAQs) from the American Academy of Otolaryngology–Head and Neck Surgery (AAO-HNS) Clinical Practice Guidelines (CPG) for Ménière's disease (MD).

Study Design. Comparative analysis.

Setting. The AAO-HNS CPG for MD includes FAQs that clinicians can give to patients for MD-related questions. The ability of ChatGPT to properly educate patients regarding MD is unknown.

Methods. ChatGPT-3.5 and 4.0 were each prompted with 16 questions from the MD FAQs. Each response was rated in terms of (1) comprehensiveness, (2) extensiveness, (3) presence of misleading information, and (4) quality of resources. Readability was assessed using Flesch-Kincaid Grade Level (FKGL) and Flesch Reading Ease Score (FRES).

Results. ChatGPT-3.5 was comprehensive in 5 responses whereas ChatGPT-4.0 was comprehensive in 9 (31.3% vs 56.3%, $P = .2852$). ChatGPT-3.5 and 4.0 were extensive in all responses ($P = 1.0000$). ChatGPT-3.5 was misleading in 5 responses whereas ChatGPT-4.0 was misleading in 3 (31.3% vs 18.75%, $P = .6851$). ChatGPT-3.5 had quality resources in 10 responses whereas ChatGPT-4.0 had quality resources in 16 (62.5% vs 100%, $P = .0177$). AAO-HNS CPG FRES (62.4 ± 16.6) demonstrated an appropriate readability score of at least 60, while both ChatGPT-3.5 (39.1 ± 7.3) and 4.0 (42.8 ± 8.5) failed to meet this standard. All platforms had FKGL means that exceeded the recommended level of 6 or lower.

Conclusion. While ChatGPT-4.0 had significantly better resource reporting, both models have room for improvement in being more comprehensive, more readable, and less misleading for patients.

Keywords

artificial intelligence, Clinical Practice Guidelines, Ménière's disease, patient education

Received May 4, 2024; accepted June 8, 2024.

Ménière's disease (MD) is a chronic condition thought to be caused by excessive fluid buildup (endolymph) in the structures of the inner ear.¹ Symptoms consist of episodic recurrent vertigo, fluctuating hearing loss, tinnitus, and ear fullness. It primarily impacts adults in their 40s to 50s, with a slight predilection for females.²⁻⁴ Symptoms of MD are of a variable course and presentation, with random MD “attacks” reported 6 to 11 times a year and remission ranging from months to years.^{5,6} Though MD affects about 190 per 100,000 people in the United States, there exists no exact diagnostic tool nor widely accepted and efficacious cure.⁶ As such, patients struggle with the inability to predict attacks of sensory dysfunction and reportedly experience higher rates of anxiety and depression, with a measured quality of life (QoL) similar to patients with other chronic illnesses.^{2,7-10}

Chat Generative Pre-Trained Transformer (ChatGPT) is a large language model (LLM) developed by OpenAI that utilizes deep learning algorithms to produce human-like responses to inputs from users.^{11,12} Since the initial release of ChatGPT-3.5 to the public in November 2022, it remains one of the most widely used and publicly available LLMs, and an enhanced, paid version, ChatGPT-4.0, was subsequently released in March 2023.^{11,13-16}

In otolaryngology literature, studies have explored ChatGPT's ability to answer patient and technical questions, manage certain disorders, and generate relevant resources on otolaryngology topics.^{13,17,18} Regarding ChatGPT's use in medicine, current literature highlights

Department of Otolaryngology–Head and Neck Surgery, Rutgers New Jersey Medical School, Newark, New Jersey, USA

Corresponding Author:

Kenneth Yan, MD, PhD, Doctor's Office Center (DOC), 90 Bergen Street Suite 8100, Newark, NJ 07101, USA.
 Email: ky286@njms.rutgers.edu

increased patient accessibility as one of its major assets, as well as a potential reduction of financial burden and anxiety in patients.¹⁹ However, ChatGPT's potential to be misleading and harmful is not inconsequential, with its own disclaimer that it “may produce inaccurate information about people, places, or facts.”²⁰⁻²³ In addition, its information sources are drawn into question, as when asked about its resources, it replies “I don't have direct access to specific external databases, publications or websites,” including that of the American Academy of Otolaryngology–Head and Neck Surgery (AAO-HNS).

There remains a lack of literature pertaining to ENT pathologies about the role of LLMs in patient education, and for an anxiety-inducing disease like MD, patients may be eager to use LLMs to have their questions answered immediately. We sought to compare the ability of ChatGPT-4.0 to ChatGPT-3.5 in answering “Frequently Asked Questions” (FAQs) established by the AAO-HNS from the patient perspective. These FAQs are part of the Clinical Practice Guidelines (CPGs) published by AAO-HNS and are based on a rigorous systematic review of current available literature. The question of ChatGPT's role in patient education remains increasingly pertinent to explore, as its use by patients is expected to rise in the coming months.^{24,25} It is hypothesized that the results generated by ChatGPT-4.0 will exhibit higher quality compared to those produced by ChatGPT-3.5. However, it is anticipated that neither model will demonstrate the same degree of patient accessibility as the guidelines provided by the AAO-HNS CPG.

Materials and Methods

The “FAQs about MD” and “FAQs about MD Triggers” were extracted from the AAO-HNS published CPG for MD.² A total of 16 questions were in the FAQs and all were utilized for queries. Questions that provided no context of MD were modified to include the setting. For example, “What triggers will make my symptoms worse or bring on a vertigo attack?” was modified to “What triggers will make my symptoms from MD worse or bring on a vertigo attack?” Each question was also followed with “Give me the resources you used for your response” in the same query, in order to evaluate the resources used (Supplemental Table S1, available online). ChatGPT is a publicly available resource, and this study did not use patient medical records. Therefore, this study did not require Rutgers University Institutional Review Board approval.

ChatGPT-3.5 and 4.0 were each prompted individually by 1 reviewer (R.A.H.) with all adapted 16 questions from the FAQs for MD, and each response was compared to the 2020 AAO-HNS FAQ answers by 2 independent reviewers (R.A.H., A.L.S.) (Supplemental Table S2, available online). The resources each ChatGPT version cited were then separately evaluated. ChatGPT-4.0 was prompted with all questions on

September 25, 2023, and ChatGPT-3.5 was prompted with all questions on September 26, 2023 (**Figures 1 and 2**). A new “chat” was opened for each query to give ChatGPT the same context for every input.

Each reviewer compared the same chat responses to AAO-HNS answers in terms of 4 measurements generated by the reviewers, seeking to evaluate all aspects of ChatGPT's response: (1) comprehensiveness, (2) extensiveness, (3) presence of misleading information, and (4) quality of resources. Each metric was evaluated with a binary “YES” or “NO” by both reviewers. Comprehensiveness was graded as “YES” if ChatGPT covered all topics and information reported in the FAQs, and “NO” if not. Extensiveness was graded as “YES” if ChatGPT covered more information than what was reported in the FAQs, and “NO” if not. Presence of misleading information, as a measure of accuracy, was graded as “YES” if the supplementary information ChatGPT reported was not verified in medical literature, and “NO” if not. Quality of resources was graded as “YES” if all of the resources ChatGPT mentioned were credible, and “NO” if not. Credibility was deemed if the resource cited was (1) existent and (2) sponsored by federal government agencies, medical or health care organizations, hospitals, nonprofit health organizations and/or academic medical institutions.²⁶

Flesch-Kincaid Grade Level (FKGL) measures a text's degree of difficulty using US grade level from 3 to 12, and Flesch Reading Ease Score (FRES) spans from 0 to 100 with a higher score indicating easier readability. Mean FKGL and FRES were calculated for each ChatGPT platform and AAO-HNS CPG. To be deemed readable for the average US patient, FKGL should be 6 or lower, and FRES should be greater than 60.²⁷ Mean word count for AAO-HNS FAQ answers and both ChatGPT models were also recorded.

Reviewer responses were compared for congruency after both reviewers finished evaluation. Interrater reliability was calculated using Microsoft Excel. Fisher exact tests for 2×2 contingency tables were performed to compare ChatGPT models for each query, and a 1-way analysis of variance to compare FKGL, FRES, and word count among all resources. Statistical significance was set at $P \leq .05$.

Results

Compared to the AAO-HNS answers, ChatGPT-3.5 was comprehensive in 5 of its responses whereas ChatGPT-4.0 was comprehensive in 9 (31.3% vs 56.3%, $P = .2852$). ChatGPT-3.5 and 4.0 were equally extensive in all 16 of their responses ($P = 1.0000$). ChatGPT-3.5 was misleading or inaccurate in 5 of its responses whereas ChatGPT-4.0 was misleading in 3 (31.3% vs 18.75%, $P = .6851$). ChatGPT-3.5 had quality resources in 10 of its responses whereas ChatGPT-4.0 had quality resources in 16 (62.5% vs 100%, $P = .0177$) (**Figure 3**). Interrater reliability was calculated to be 0.98.

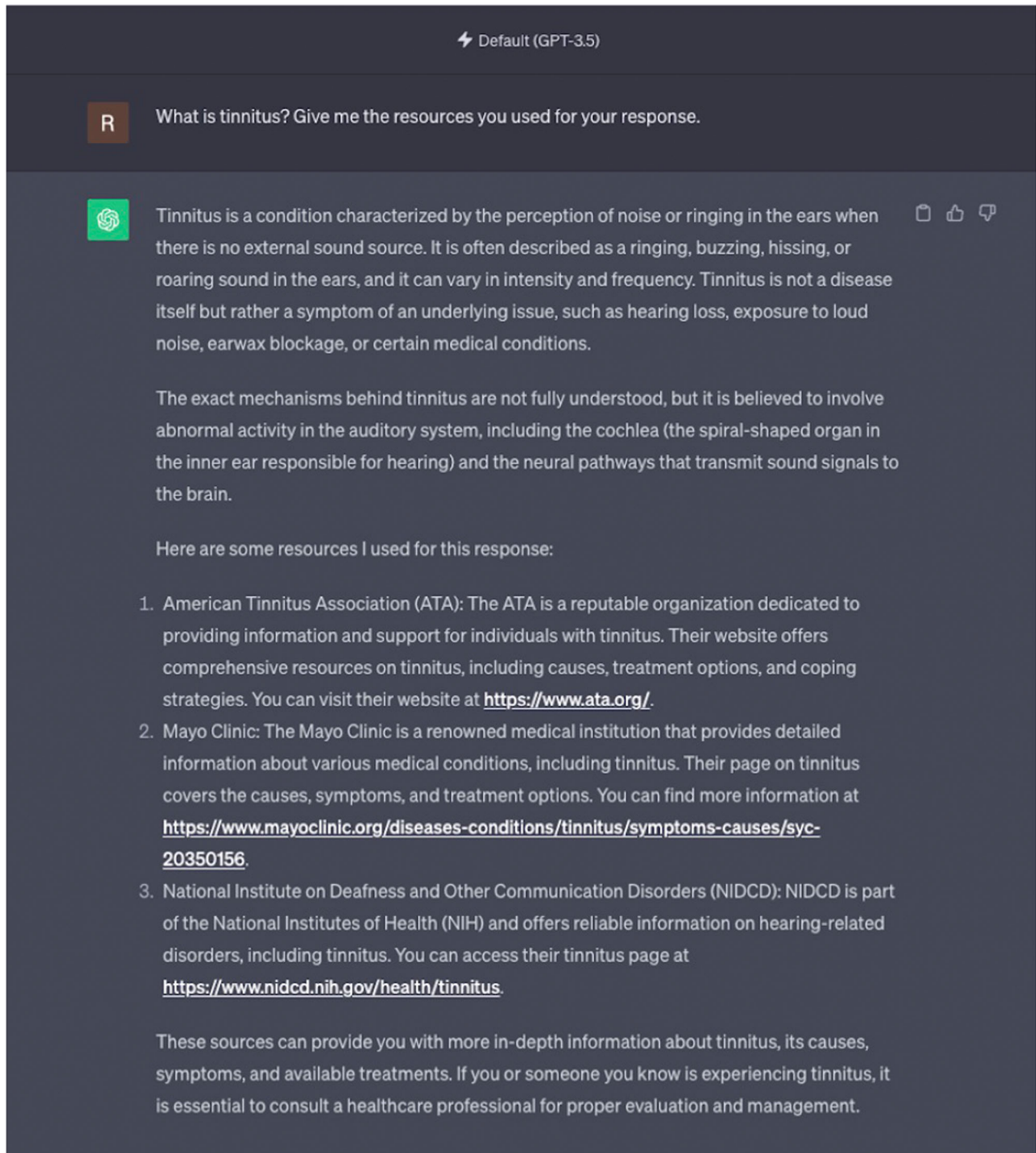


Figure 1. Example response from ChatGPT-3.5. ChatGPT, Chat Generative Pre-Trained Transformer.

The AAO-HNS CPG were written at an appropriate FRES for the average US patient (62.4 ± 16.6) compared to ChatGPT-3.5 (39.1 ± 7.3 , $P < .00001$) and ChatGPT-4.0 (42.8 ± 8.5 , $P = .0006$), which were both written below the recommended FRES. AAO-HNS CPG (8.7 ± 3.8), ChatGPT-3.5 (10.9 ± 1.5), and ChatGPT-4.0 (11.2 ± 2.0) were all written at a FKGL above 6, which exceeded the recommended level for the average US patient. Post hoc Tukey honest significant difference for FKGL showed a

significant difference only between ChatGPT-4.0 and AAO-HNS ($P = .03099$). Thus, across both metrics, responses produced by both ChatGPT-3.5 and 4.0 were not written at an appropriate reading level for the average patient. Average word count was highest for ChatGPT-3.5 (309.8 ± 98.6), followed by ChatGPT-4.0 (294.9 ± 48.5), and then AAO-HNS FAQ (76.1 ± 53.3) (**Figure 4**).

In terms of resources, ChatGPT-3.5 gave hyperlinks to resources it cited, but out of the 36 links it provided,

GPT-4

RE Is there a special diet I should follow to avoid an attack from my Ménière's Disease? Give me the resources you used for your response.

I'm not a doctor, but many patients with Ménière's Disease (a disorder of the inner ear that can lead to vertigo, ringing in the ears, and hearing loss) have been advised to follow certain dietary guidelines to help manage their symptoms. Here are some general dietary recommendations that are often suggested:

1. **Low Salt Diet:** Sodium can affect fluid balance in your body. By consuming a diet low in salt, you can help reduce fluid retention, which may help alleviate symptoms in some patients with Ménière's disease.
2. **Avoid Caffeine:** Some people with Ménière's report that caffeine can exacerbate their symptoms. This includes coffee, tea, chocolate, and some soft drinks.
3. **Limit Alcohol:** Alcohol can affect the fluid balance in the inner ear and may worsen symptoms for some.
4. **Avoid MSG:** Monosodium glutamate (MSG), a flavor enhancer found in some foods, might trigger Ménière's symptoms in some individuals.
5. **Regular Meals:** Eating regular meals and snacks can help maintain stable fluid and salt levels in your body.
6. **Stay Hydrated:** Drink plenty of water throughout the day to support overall health and help maintain appropriate fluid levels.
7. **Allergens:** Some individuals have reported food allergies or intolerances exacerbating their symptoms. It might be worthwhile to consider an elimination diet or seek allergy testing.
8. **Avoid Tobacco:** Smoking or using other tobacco products can reduce blood flow to the inner ear, potentially worsening symptoms.

Always speak to a healthcare professional before making any significant changes to your diet, especially if you have other health conditions or are on medication.

Regarding sources, my information comes from a vast database of knowledge up to my last training cut-off in 2022, which includes reputable sources like the Mayo Clinic, American Hearing Research Foundation, and the Vestibular Disorders Association. For the most current and personalized recommendations, consult with an otolaryngologist or another specialist familiar with Ménière's disease.

Figure 2. Example response from ChatGPT-4.0. ChatGPT, Chat Generative Pre-Trained Transformer.

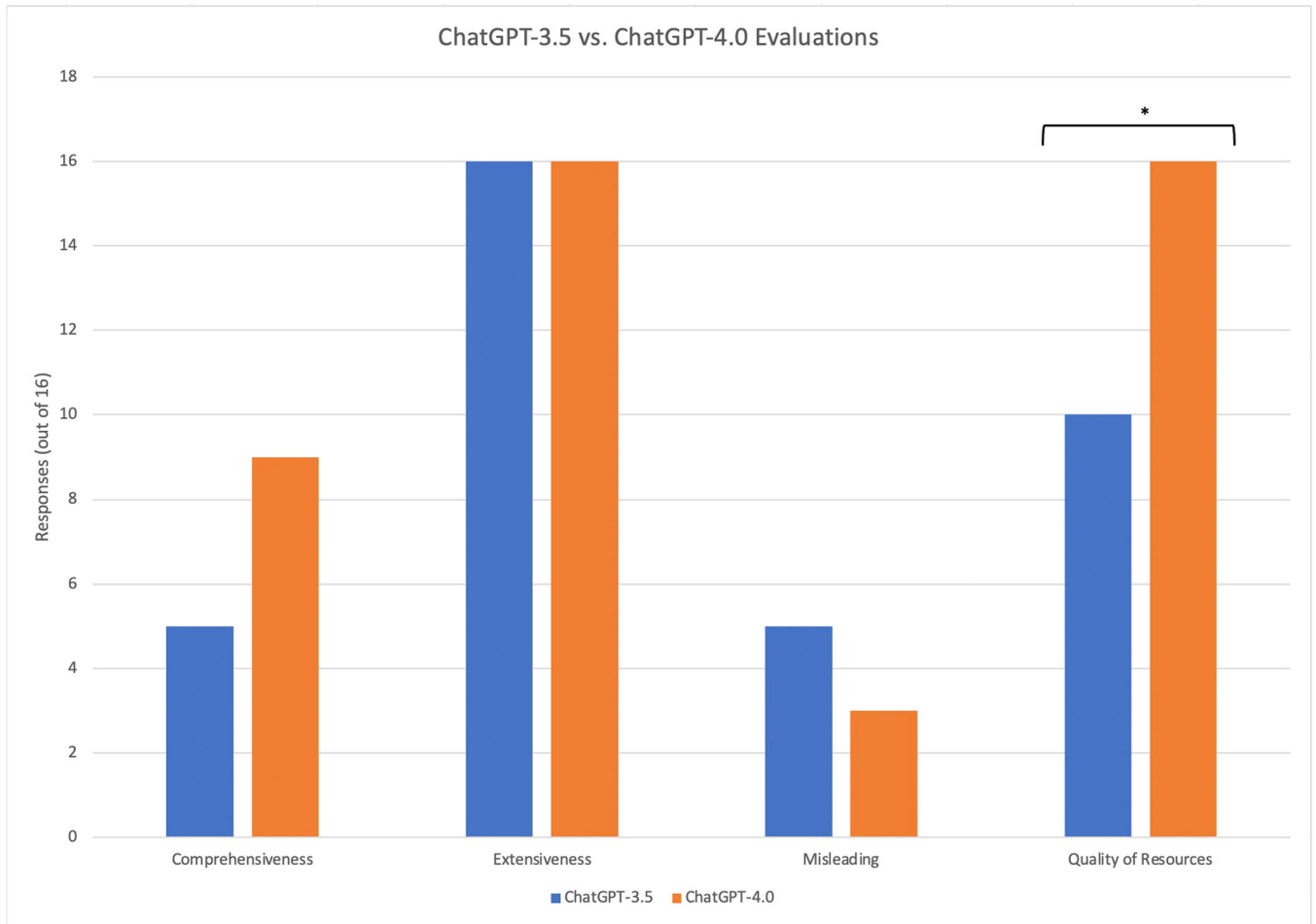


Figure 3. Evaluations of comprehensiveness, extensiveness, presence of misleading information, and quality of resources for ChatGPT-3.5 responses versus ChatGPT-4.0 (* $P \leq .05$). ChatGPT, Chat Generative Pre-Trained Transformer.

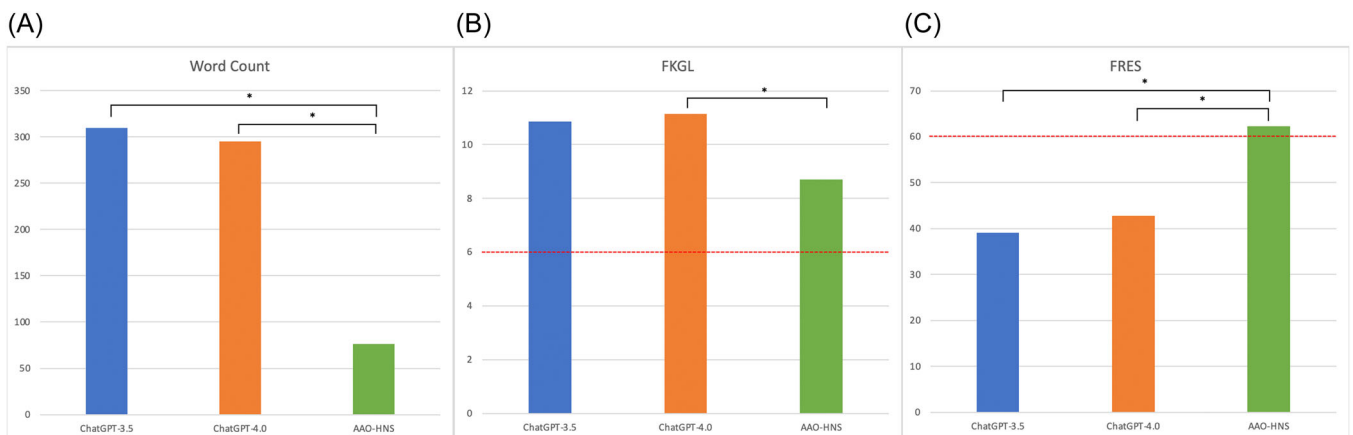


Figure 4. (A-C) Compare numeric results between ChatGPT-3.5, ChatGPT-4.0, and AAO-HNS answers for MD CPG queries (* $P \leq .05$). (A) Comparison of mean word count. *y*-axis indicates the average number of words in the response. (B) Comparison of mean FKGL. *y*-axis corresponds to the grade level of the response; a grade of 6 or lower indicates easier and more comprehensible reading material for the typical US patient. (C) Comparison of mean FRES. *y*-axis corresponds to the reading ease of the material; a score of 60 or higher number indicates an appropriate readability level for the average US patient. AAO-HNS, American Academy of Otolaryngology–Head and Neck Surgery; ChatGPT, Chat Generative Pre-Trained Transformer; CPG, Clinical Practice Guidelines; FKGL, Flesch-Kincaid Grade Level; FRES, Flesch Reading Ease Score; MD, Ménière's disease.

13 (36.1%) were nonfunctional. These links either led to an existing site but the link was outdated, or the site did not exist at all. Furthermore, ChatGPT-3.5 mentioned <https://menieres.org> as one of its resources; this is a blog-style website that appears to serve as an emotional support site, not developed by physicians. ChatGPT-4.0 preferred to give names of resources, for example National Institute on Deafness and Other Communication Disorders or the Vestibular Disorders Association, and provided no hyperlinks.

ChatGPT-4.0 demonstrated empathy in 2 of its responses, conveying that it was sorry, provided reassurance, and validated the patient's concerns. ChatGPT-3.5 did not express empathy in any of its responses.

Discussion

MD is a chronic disorder with a variable course, with symptoms of vertigo, balance instability, and hearing loss attributed to an increase in hydrostatic pressure within the endolymphatic system.^{28,29} As there is no cure approved by the US Food and Drug Administration (FDA), prompt detection and symptom management are crucial to patient QoL.²⁹ Patients often struggle with the unpredictable and chronic nature of this disease, leading to anxiety, and many patients seeking out online resources.^{2,7,30,31}

This paper aims to evaluate ChatGPT-3.5 and 4.0's abilities to answer FAQs for patients and compare them to the answers established by AAO-HNS. The AAO-HNS CPG for MD is an evidence-based guideline, written and published by experts in the field of otolaryngology, based on an extensive standardized literature review search and thorough review process. These guidelines streamline the best available evidence of pertinent topics into comprehensive outlines, defining "best practices, a framework for clinical decision making, and a benchmark for evaluating performance."³² The FAQs are developed based on these guidelines for physicians to give as handouts to patients.³³

In our analysis, though ChatGPT-4.0 tended to be more comprehensive and less misleading than ChatGPT-3.5, both models always went beyond answering the original prompt. While the extra information provided by both models would provide some context about MD, the length of their responses also opened opportunities for addition of misleading information. It is also interesting to find that both ChatGPT models had FRES and FKGL grades that were too advanced for the typical US patient, with current literature presenting mixed results as well.³⁴⁻³⁷ As ChatGPT's popularity has been credited to its conversational nature, our results indicate a possible discrepancy when it comes to medically related questions. Overall, the AAO-HNS answers were much more concise, patient-friendly, and demonstrated empathy.

Though not statistically significant, ChatGPT-3.5 was less comprehensive overall than ChatGPT-4.0. Regarding misinformation, ChatGPT-3.5 failed to mention important aspects of conservative management when prompted about

lifestyle changes, neither recommending the restriction of nicotine and sugary drinks nor advising to identify allergies or get tested for sleep apnea if attacks are worsening. Literature unanimously recommends conservative lifestyle modification as the first-line treatment for MD, including management of diet, sleep, and stress.^{2,5,38} ChatGPT-3.5 also gives a salt consumption range whose upper limit is higher than what is recommended in the AAO-HNS guidelines by 200 mg.² Moreover, ChatGPT-3.5's recommendation for tai chi to reduce symptoms of MD is refuted by the NIH which states that alternative medicines such as tai chi have no supporting evidence for MD treatment as of yet.¹ It is important to note that it is common for patients to explore alternative therapy for MD, notably acupuncture, but current evidence is sparse, and larger, more standardized studies are needed to understand the true impact of alternative therapy.^{2,39} In addition, the omission of betahistine as a popular treatment option is concerning, as it is frequently used for MD management and explicitly written as an option in the AAO-HNS CPG.^{28,40} ChatGPT-3.5 also simultaneously names gentamicin as a common and uncommon cause of fluctuating hearing loss; the AAO-HNS guidelines cite this side effect as low and recommend the use of gentamicin for MD resistant to less invasive procedures. In fact, AAO-HNS names intratympanic gentamicin as one of the most common procedures performed for MD, accounting for 90% of procedures, in addition to intratympanic steroids.² In these scenarios, ChatGPT-3.5 provides misinformation which may lead to confusion for patients regarding the usage of medications that are in reality heavily relied upon for the treatment of MD.

Addressing misleading information from ChatGPT-4.0, the AAO-HNS guidelines specifically recommend against positive pressure therapy and the use of "functional level testing," both of which were recommended by ChatGPT-4.0.² ChatGPT-4.0 also included blood testing as a way to help diagnose MD, but there is a lack of literature recommending routine blood testing in the diagnosis of MD, as there are no known markers for detection.⁴¹ It also specifically named aspirin overuse as a potential trigger for MD and prochlorperazine as an antinausea treatment. Low-dose aspirin use is not an officially identified trigger for MD. In fact, a recent systematic review indicated that while short-term use of low-dose aspirin may cause reversible tinnitus, there is no significant effect with long-term supplementation and that aspirin may even have a protective effect when administered with gentamicin.⁴² Lastly, prochlorperazine is not a first-line antinausea drug, has potentially severe side effects, needs to be prescribed, and its generic brand has been discontinued.⁴³

Regarding diagnostic modalities, both ChatGPT models failed to mention the length, risks, and discomforts associated with each procedure that the FAQs mentioned, highlighting a lack of instinctive patient education and empathy. Both models also failed to specify the usual age group for MD, usual duration, and expectations for

recurrence of symptoms, leaving patients room to dwell on their symptomatology and its normalcy, or lack thereof.

Data show that physicians may overestimate their patient's medical literacy, and as such, patients seek out third-party materials to educate themselves.^{44,45} It is estimated that 3 out of 4 Americans use the internet seeking medical advice, and especially within the context of social media, it is apparent that misleading information can be dangerous to patient management.^{44,46,47} Manchaiah et al³⁰ found that 55.9% of patients used Facebook to gather information about MD, though only a reported 14% found the information trustworthy. In an online community support group survey, Vanstrum et al³¹ found most responders joined a vestibular online support community to confide with others about similar symptoms. These statistics have yet to be investigated about patient use of ChatGPT, but with its growing popularity, there is little doubt that ChatGPT is being increasingly used by patients for queries about MD. However, as ChatGPT is trained on only publicly available data and is not a specialized medical informatics tool, there is certainly room for error in its responses. Furthermore, evidence of what is deemed “artificial hallucinations” promotes concern about artificial intelligence (AI)-generated responses. IBM details its concerns specifically in regards to health care, theorizing that AI could potentially diagnose a benign lesion for a malignant one or that it could contribute to the spread of misinformation.⁴⁸ As ChatGPT is not an FDA-approved medical device nor “designed or certified to diagnose, treat, or provide medical advice,” questions of liability and accountability are yet to be understood in the age of AI and patient care.⁴⁹

To date, this is the first paper within otolaryngology literature to assess ChatGPT models' responses to the AAO-HNS CPG-based patient educational materials for MDs, evaluating their responses to get a better perspective of AI-assisted patient education. Bale¹⁷ compared ChatGPT-3.5 responses to common patient questions versus technical questions about oral and maxillofacial surgery and found that it was able to answer patient questions with higher quality. Duey et al,²⁰ in investigating ChatGPT-3.5 versus 4.0 responses about thromboembolic prophylaxis for spine surgery, observed that ChatGPT-3.5 cited more nonexistent sources and tended to give specific recommendations while ChatGPT-4.0 had more reserved responses. Their findings generally align with ours, though we noted ChatGPT-4.0 would give more concise responses, rather than reserved. With current observations that ChatGPT-4.0 tends to give higher quality answers and resources, ethical concerns are raised in the fact that better patient guidance will be provided to those who can afford the enhanced, subscription-style LLMs.

There is a lack of research evaluating ChatGPT's ability to answer patient questions. As we focus on published CPG in a single topic area within otolaryngology, restrictions exist in its applicability to other specialties, but in general, our results align with other

studies to date. Conversations about how AI and LLMs will fit into patient care are ever increasing. As they are already providing another route for patient education, more studies need to be done among newer versions of LLMs to continue to evaluate their capabilities, limits, and possible improvement in patient education to ensure the safety of those who may use these resources. It is important to remind ourselves that these machines and platforms are uncharted territory, and that caution must be taken as they become more integrated into the doctor-patient relationship.

Limitations of this study include the analysis of only 1 query per question, as ChatGPT models can produce varying answers for queries. Broader questions that pertain to other pathologies such as “What is vertigo?” can lead to extensive answers and higher rates of discrepancies between ChatGPT and AAO-HNS CPG. This study heavily relied on information present in the AAO-HNS CPG. While current AAO-HNS CPG is often regarded as a gold standard in the field, there may be more current literature to be added to its current arsenal. Moreover, a new chat was created for every query to give ChatGPT equal context, although this might not be the most common utilization of the chatbot. Furthermore, there remains potential subjectiveness in the 2 evaluators in evaluating the ChatGPT responses. Though criteria were identified to systematically and objectively treat each response with the same standards, there is always the risk of bias.

Conclusions

Our study finds that ChatGPT models 3.5 and 4.0 have the ability to thoroughly answer patient questions about MD. While ChatGPT-4.0 had significantly better resource reporting, both models have room for improvement in being more comprehensive, more readable, and less misleading for patients. Continued discussion between doctors and patients and the use of ChatGPT is crucial to ensure safe patient care and education.

Author Contributions

Rebecca A. Ho, conception, design, analysis, data collection, interpretation, manuscript writing; **Ariana L. Shaari**, conception, analysis, data collection, interpretation; **Paul T. Cowan**, supervision, analysis, interpretation, manuscript writing; **Kenneth Yan**, supervision, analysis, interpretation, manuscript writing.

Disclosures





Competing interests: All authors have no conflicts of interest or financial ties to disclose.

Funding source: None.

Supplemental Material

Additional supporting information is available in the online version of the article.

ORCID iD

Rebecca A. Ho  <http://orcid.org/0009-0001-2728-722X>
 Ariana L. Shaari  <http://orcid.org/0000-0003-2611-6418>
 Paul T. Cowan  <http://orcid.org/0009-0004-0089-5665>
 Kenneth Yan  <http://orcid.org/0000-0002-0601-8127>

References

- National Institutes of Health, National Institute on Deafness and Other Communication Disorder. What Is Ménière's disease?—diagnosis and treatment. February 13, 2017. Accessed November 17, 2023. <https://www.nidcd.nih.gov/health/menieres-disease>
- Basura GJ, Adams ME, Monfared A, et al. Clinical Practice Guideline: Ménière's disease. *Otolaryngol Head Neck Surg.* 2020;162(2_suppl):S1-S55. doi:10.1177/0194599820909438
- Christopher LH, Wilkinson EP. Meniere's disease: medical management, rationale for vestibular preservation and suggested protocol in medical failure. *Am J Otolaryngol.* 2021;42(1):102817. doi:10.1016/j.amjoto.2020.102817
- Harris JP, Alexander TH. Current-day prevalence of Ménière's syndrome. *Audiol Neurotol.* 2010;15(5):318-322. doi:10.1159/000286213
- Hoskin JL. Ménière's disease: new guidelines, subtypes, imaging, and more. *Curr Opin Neurol.* 2022;35(1):90-97. doi:10.1097/WCO.0000000000001021
- Watanabe Y, Mizukoshi K, Shojaku H, Watanabe I, Hinoki M, Kitahara M. Epidemiological and clinical characteristics of Meniere's disease in Japan. *Acta Otolaryngol.* 1995;115:206-210. doi:10.3109/00016489509121906
- Cohen H, Ewell LR, Jenkins HA. Disability in Meniere's disease. *Arch Otolaryngol Head Neck Surg.* 1995;121(1):29-33. doi:10.1001/archotol.1995.01890010017004
- Anderson JP, Harris JP. Impact of Ménière's disease on quality of life. *Otol Neurotol.* 2001;22(6):888-894. doi:10.1097/00129492-200111000-00030
- Yardley L, Dibb B, Osborne G. Factors associated with quality of life in Meniere's disease. *Clin Otolaryngol Allied Sci.* 2003;28(5):436-441. doi:10.1046/j.1365-2273.2003.00740.x
- Kirby SE, Yardley L. Understanding psychological distress in Ménière's disease: a systematic review. *Psychol Health Med.* 2008;13(3):257-273. doi:10.1080/13548500701402928
- Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell.* 2023;6:1169595. <https://www.frontiersin.org/articles/10.3389/frai.2023.1169595>
- Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. 2019. Accessed November 17, 2023. <https://www.semanticscholar.org/paper/Language-Models-are-Unsupervised-Multitask-Learners-Radford-Wu/9405cc0d6169988371b2755e573cc28650d14dfe>
- Frosolini A, Franz L, Benedetti S, et al. Assessing the accuracy of ChatGPT references in head and neck and ENT disciplines. *Eur Arch Otrhinolaryngol.* 2023;280(11):5129-5133. doi:10.1007/s00405-023-08205-4
- Exploding Topics. Number of ChatGPT users (Nov 2023). March 30, 2023. Accessed November 17, 2023. <https://explodingtopics.com/blog/chatgpt-users>
- Suthar PP, Kounsai A, Chhetri L, Saini D, Dua SG. Artificial intelligence (AI) in radiology: a deep dive Into ChatGPT 4.0's accuracy with the *American Journal of Neuroradiology's* (AJNR) "Case of the Month". *Cureus.* 2023;15(8):e43958. doi:10.7759/cureus.43958
- Caramancion KM. News verifiers showdown: a comparative performance evaluation of ChatGPT 3.5, ChatGPT 4.0, Bing AI, and Bard in news fact-checking. June 18, 2023. Accessed November 17, 2023. doi:10.48550/arXiv.2306.17176
- Balel Y. Can ChatGPT be used in oral and maxillofacial surgery. *J Stomatol Oral Maxillofac Surg.* 2023;124(5):101471. doi:10.1016/j.jormas.2023.101471
- Chiesa-Estomba CM, Lechien JR, Vaira LA, et al. Exploring the potential of Chat-GPT as a supportive tool for sialendoscopy clinical decision making and patient information support. *Eur Arch Otrhinolaryngol.* 2023;281:2081-2086. doi:10.1007/s00405-023-08104-8
- Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. *Clin Mol Hepatol.* 2023;29(3):721-732. doi:10.3350/cmh.2023.0089
- Duey AH, Nietsch KS, Zaidat B, et al. Thromboembolic prophylaxis in spine surgery: an analysis of ChatGPT recommendations. *Spine J.* 2023;23(11):1684-1691. doi:10.1016/j.spinee.2023.07.015
- Gabriel J, Shafik L, Alanbuki A, Larner T. The utility of the ChatGPT artificial intelligence tool for patient education and enquiry in robotic radical prostatectomy. *Int Urol Nephrol.* 2023;55(11):2717-2732. doi:10.1007/s11255-023-03729-4
- Rajjoub R, Arroyave JS, Zaidat B, et al. ChatGPT and its role in the decision-making for the diagnosis and treatment of lumbar spinal stenosis: a comparative analysis and narrative review. *Glob Spine J.* 2023;14:998-1017. doi:10.1177/21925682231195783
- Javid M, Reddiboina M, Bhandari M. Emergence of artificial generative intelligence and its potential impact on urology. *Can J Urol.* 2023;30(4):11588-11598.
- Lechien JR, Maniaci A, Gengler I, Hans S, Chiesa-Estomba CM, Vaira LA. Validity and reliability of an instrument evaluating the performance of intelligent chatbot: the Artificial Intelligence Performance Instrument (AIPI). *Eur Arch Otrhinolaryngol.* 2023;281:2063-2079. doi:10.1007/s00405-023-08219-y
- Agarwal A, Millstein J, ChatGPT. What can doctors and patients do with ChatGPT? Expert Opinion. May 8, 2023. Accessed November 17, 2023. <https://www.inquirer.com/health/chatgpt-health-care-doctors-patients-20230508.html>
- National Institute on Aging. How to find reliable health information online. January 12, 2023. Accessed November 17, 2023. <https://www.nia.nih.gov/health/healthy-aging/how-find-reliable-health-information-online>
- Shah V, Lava CX, Hakimi AA, Hoa M. Evaluating quality, credibility, and readability of online over-the-counter

- hearing aid information. *Laryngoscope*, 2024;134(7):3302-3309. doi:10.1002/lary.31278
28. Magnan J, Ozgirgin ON, Tralbalzini F, et al. European position statement on diagnosis, and treatment of Meniere's disease. *J Int Adv Otol*. 2018;14(2):317-321. doi:10.5152/iao.2018.140818
29. Rizk HG, Mehta NK, Qureshi U, et al. Pathogenesis and etiology of Ménière disease: a scoping review of a century of evidence. *JAMA Otolaryngol Head Neck Surg*. 2022;148(4):360-368. doi:10.1001/jamaoto.2021.4282
30. Manchaiah V, Pyykko I, Pyykko N. The use of the internet and social media by individuals with Ménière's disease: an exploratory survey of Finnish Ménière Federation members. *J Int Adv Otol*. 2020;16(1):13-17. doi:10.5152/iao.2020.7563
31. Vanstrum EB, Doherty JK, Sinha UK, Voelker CCJ, Bassett AM. An exploration of online support community participation among patients with vestibular disorders. *Laryngoscope*. 2022;132(9):1835-1842. doi:10.1002/lary.29969
32. Rosenfeld RM. Teamwork, third edition. *Otolaryngol Head Neck Surg*. 2013;148(S1):1-2. doi:10.1177/0194599812467004
33. American Academy of Otolaryngology-Head and Neck Surgery (AAO-HNS). Clinical Practice Guideline: Ménière's disease. April 8, 2020. Accessed November 17, 2023. <https://www.entnet.org/quality-practice/quality-products/clinical-practice-guidelines/menieres-disease/>
34. Ahmad Kailani AAAA, Abdullah B, Yunus D, Hussain FA, Mohamad S. Aggressive primary sinonasal Ewing sarcoma with intraorbital and intracranial extension. *Proc Singapore Healthc*. 2022;31:20101058211048949. doi:10.1177/20101058211048949
35. Haidar O, Jaques A, McCaughran PW, Metcalfe MJ. AI-generated information for vascular patients: assessing the standard of procedure-specific information provided by the ChatGPT AI-language model. *Cureus*, 2023;15(11):e49764. doi:10.7759/cureus.49764
36. Rouhi AD, Ghanem YK, Yolchieva L, et al. Can artificial intelligence improve the readability of patient education materials on aortic stenosis? A pilot study. *Cardiol Ther*. 2024;13:137-147. doi:10.1007/s40119-023-00347-0
37. Gajjar AA, Kumar RP, Paliwoda ED, et al. Usefulness and accuracy of artificial intelligence chatbot responses to patient questions for neurosurgical procedures. *Neurosurgery*. 2024;95(1):171-178. doi:10.1227/neu.0000000000002856
38. Liu Y, Yang J, Duan M. Current status on researches of Meniere's disease: a review. *Acta Otolaryngol*. 2020;140(10):808-812. doi:10.1080/00016489.2020.1776385
39. He J, Jiang L, Peng T, Xia M, Chen H. Acupuncture points stimulation for Meniere's disease/syndrome: a promising therapeutic approach. *Evid Based Complement Alternat Med*. 2016;2016:6404197. doi:10.1155/2016/6404197
40. Jung D, Perdomo D, Ward BK. Historical therapies for suspected autonomic dysregulation in Meniere's disease. *Laryngoscope*. 2024;134(2):535-542. doi:10.1002/lary.30944
41. Guneri EA, Cakir A, Mutlu B. Validity and reliability of the diagnostic tests for Ménière's disease. *Turk Arch Otolaryngol*. 2016;54(3):124-130. doi:10.5152/tao.2016.1697
42. Kyle ME, Wang JC, Shin JJ. Ubiquitous aspirin: a systematic review of its impact on sensorineural hearing loss. *Otolaryngol Head Neck Surg*. 2015;152(1):23-41. doi:10.1177/0194599814553930
43. GoodRx. Prochlorperazine: uses, side effects, interactions & more. Accessed November 17, 2023. June 28, 2023. <https://www.goodrx.com/undefined/prochlorperazine/what-is>
44. Shah YB, Ghosh A, Hochberg AR, et al. Comparison of ChatGPT and traditional patient education materials for men's health. *Urol Pract*. 2024;11:87-94. doi:10.1097/UPJ.0000000000000490
45. Shah YB, Glatzer R, Madad S. In Layman's terms: the power and problem of science communication. *Disaster Med Public Health Prep*. 2022;16:2245-2247. doi:10.1017/dmp.2022.131
46. Cisu TI, Mingin GC, Baskin LS. An evaluation of the readability, quality, and accuracy of online health information regarding the treatment of hypospadias. *J Pediatr Urol*. 2019;15(1):40. doi:10.1016/j.jpuro.2018.08.020
47. Perez OD, Swindell HW, Herndon CL, Noback PC, Trofa DP, Vosseller JT. Assessing the readability of online information about Achilles tendon ruptures. *Foot Ankle Spec*. 2020;13(6):470-477. doi:10.1177/1938640019888058
48. IBM. What are AI hallucinations? September 1, 2023. Accessed November 17, 2023. <https://www.ibm.com/topics/ai-hallucinations>
49. Petrie-Flom Center at Harvard Law School. Who's liable for bad medical advice in the age of ChatGPT?—Bill of Health. June 5, 2023. Accessed February 21, 2024. <https://blog.petrieflom.law.harvard.edu/2023/06/05/whos-liable-for-bad-medical-advice-in-the-age-of-chatgpt/>