

# BPPV Information on Google Versus AI (ChatGPT)

Jeffrey R. Bellinger BS<sup>1</sup>, Julian S. De La Chapa MD<sup>2</sup>,  
 Minhie W. Kwak BA<sup>1</sup>, Gabriel A. Ramos BA<sup>1</sup>,  
 Daniel Morrison MD<sup>2</sup>, and Bradley W. Kesser MD<sup>2</sup>

<sup>1</sup> University of Virginia School of Medicine, Charlottesville, VA

<sup>2</sup> Department of Otolaryngology – Head and Neck Surgery, University of Virginia Hospital, VA



Scan the QR code to view the article published in *Otolaryngology – Head and Neck Surgery*

## Purpose

- Online information pertaining to benign paroxysmal positional vertigo (BPPV) that is high in quality and written at an appropriate level is critical for patients who may be replacing or supplementing provider information with information found on the internet.
- We sought to quantitatively compare online patient education materials found using traditional search engines (Google) versus conversational artificial intelligence (AI) models (ChatGPT) for BPPV.

## Background

- ChatGPT is a chatbot developed by the company OpenAI that was released to the public Nov. 30<sup>th</sup>, 2022.<sup>1</sup>
- ChatGPT uses the Generative Pre-Trained Transformer (GPT) which is an example of an artificial intelligence derived large language model (LLM).
- It uses a variety of algorithms to analyze the user’s input and develop a human-like response based on the patterns and structures it was trained on.<sup>2</sup>
- ChatGPT and other chatbots have the potential to revolutionize how people search for and interact with online health information.

## Methods

### Data Acquisition (Feb. 2023)

- The top 30 Google search results for “benign paroxysmal positional vertigo” were extracted.
  - 29/30 webpages were included in the analysis.
  - Of note, 4/30 (13.3%) webpages required membership/payment to read the full article in which only the available abstracts were reviewed.
- The definition, cause, symptoms, diagnosis, and treatment of BPPV were asked in the form of questions to ChatGPT.
  - This was repeated 3 times for each query for an average score.
  - The responses to each question were analyzed individually as well as in a combined response (to model questions being asked sequentially).
  - Example of individual response given in Table 1.

### Data Analysis

Two reviewers analyzed and scored the Google webpages and ChatGPT responses on the following domains:

#### Readability

- Assessed through Flesch-Kincaid Grade Level (FKGL) and Flesch Reading Ease (FRE) scores.
- Lower scores on FKGL, and higher scores on FRE, correspond to text that is easier to read (Table 2).

#### Quality

- The DISCERN instrument is a validated scoring metric assessing the quality of information on the publication’s reliability, quality of treatment choices, and overall quality of information.
  - 16 questions (3 parts) - scale of 0-80 with 80 representing the highest score.

#### Understandability and Actionability

- Assessed using the validated Patient Education Assessment Tool for Printed Materials (PEMAT-P).
  - 24 questions divided into sections for understandability, actionability, and a total score.
  - Scoring was recorded as a percentage out of a total possible score.

#### Accuracy and Currency

- 5-point Likert scale of the accuracy and currency of information on ChatGPT responses were determined by two neurotology experts.

#### References

1. OpenAI. <https://openai.com/blog/openai-api/>  
 2. Dheda G. Open AI Master. How does ChatGPT work? Published April 5, 2023. <https://openaimaster.com/how-does-chatgpt-work/>  
 3. Weis B. *Health Literacy: A Manual for Clinicians*. American Medical Association, American Medical Foundation; 2003  
 4. von Brevern M, Radtke A, Lezius F, et al. Epidemiology of benign paroxysmal positional vertigo: a population based study. *Journal of Neurology, Neurosurgery & Psychiatry*. 2007;78:710-715.

Question	ChatGPT Response
"what is benign paroxysmal positional vertigo?"	Benign paroxysmal positional vertigo (BPPV) is a common inner ear disorder that causes brief episodes of dizziness or vertigo, which can be accompanied by nausea and vomiting. It is caused by the displacement of small calcium carbonate crystals (otoconia) within the inner ear's semicircular canals, which are responsible for detecting changes in head position and movement. When these crystals become dislodged and float into one of the semicircular canals, they can cause abnormal fluid movement in the canal, leading to the sensation of spinning or vertigo.  BPPV typically presents with brief, intense episodes of vertigo that are triggered by changes in head position, such as rolling over in bed or tilting the head back to look up. The symptoms may last for a few seconds to several minutes, and then subside. Although BPPV can be distressing, it is not a serious condition, and it can be effectively treated with a series of simple head movements called the Epley maneuver, which can help reposition the displaced crystals and alleviate symptoms. In some cases, medication or surgery may be necessary if the vertigo persists or if there are underlying medical conditions contributing to the symptoms.

FRE Score	Descriptive Categories	Estimated Reading Grade
90-100	Very Easy	5th Grade
80-90	Easy to read	6th Grade
70-80	Fairly easy to read	7th Grade
60-70	Standard/Plain English	8th and 9th Grade
50-60	Fairly difficult to read	10th-12th Grade
30-50	Difficult to read	College Reading Level
0-30	Very difficult to read	College Graduate Reading Level

  

FKGL	Education Equivalence
5-6	5th-6th grade
7-8	7th-8th grade
9	9th grade
10	10th grade
11	11th grade
12	12th grade
13	University 1st year
14	University 2nd year
15-16	University 3rd year and 4th year
17-18	Master's and professional degree
19	Doctorate degree

Flesch Reading Ease Score (FRE) Equation:  $206.835 - 1.015(\text{total words}/\text{total sentences}) - 84.6(\text{total syllables}/\text{total words})$   
 Flesch-Kincaid Grade Level (FKGL) Equation:  $0.39(\text{total words}/\text{total sentences}) + 11.8(\text{total syllables}/\text{total words}) - 15.59$

## Results

### Readability

	Google	ChatGPT	P-value
FKGL, mean (SD)	10.7 (2.6)	13.9 (2.5)	<0.001
FRE, mean (SD)	46.5 (14.3)	34.9 (11.2)	0.005
Abbreviations: FKGL, Flesch-Kincaid Grade Level; FRE, Flesch Reading Ease; SD, standard deviation			

### Quality, Understandability, & Actionability

	Google	ChatGPT (Individual)	ChatGPT (Combined)	P value
DISCERN Total, mean (SD)	56.5 (9.5)	17.5 (3.9) ***	25.0 (0.9) ***	<0.001
DISCERN Part 1, mean (SD)	32.0 (4.2)	NA	NA	NA
DISCERN Part 2, mean (SD)	25.4 (7.5)	17.5 (3.9) **	25.0 (0.9)	0.001
PEMAT-P Total, mean (SD), %	71.5 (15.0)	59.8 (7.6) *	56.8 (1.4)	0.017
Understandability, mean (SD), %	82.3 (8.6)	72.3 (2.4) ***	56.0 (0.0) ***	<0.001
Actionability, mean (SD), %	45.2 (33.5)	34.2 (25.0)	60.8 (5.8)	
Abbreviations: SD, standard deviation; PEMAT-P, Patient Education Assessment Tool for Printed Materials. All P-values are derived from independent samples T-tests comparing either ChatGPT (individual) or ChatGPT (combined) to Google. * P = 0.017 ** P = 0.001 *** P < .001				

## Discussion

- Online information is only useful to patients if it is accurate and consists of high-quality information that is easy to read, understand, and apply.
- The American Medical Association (AMA) recommends patient information be between a 4<sup>th</sup>-6<sup>th</sup> grade reading level.<sup>3</sup>
  - Our results demonstrate information on Google is at about an 11<sup>th</sup> grade reading level, with ChatGPT results being even higher, at nearly a 2<sup>nd</sup> year university student level.
- Individual ChatGPT responses had lower quality relative to combined ChatGPT and Google responses.
  - People age >60 are most affected by BPPV and may not possess the online skills to manipulate conversational chatbot search algorithms to effectively serve their information needs.<sup>4</sup>
- While OpenAI concedes its models aren't to be used for diagnostic, triage, or treatment purposes, this blanket statement will likely not deter people, as they now have an interactive, conversational “Dr. Google”.
  - A combination of material that is difficult to understand and offering a higher degree of action could make this information more problematic.
- Our experts “agree” that ChatGPT information on BPPV is generally accurate and current, but did identify misleading and tangential information, as well as uncommon treatment paradigms.

### Accuracy and Currency of ChatGPT Responses

Section	Accuracy score, mean (SD)	Currency Score, mean (SD)	Comments
Definition	4.50 (0.55)	4.50 (0.55)	1. The crystals are located in the vestibule, on the surface of the utricle and saccule, not in the semicircular canals. They become dislodged and float into the semicircular canals, causing the characteristic symptoms. 2) Usually does not cause vomiting. 3) Medication or surgery is rarely needed to improve symptoms
Cause	4.33 (0.82)	4.50 (0.84)	1) Visual system really does not have anything to do with BPPV, other than the pathophysiology causes characteristic eye movements (nystagmus) 2) Typically does not cause unsteadiness or lightheadedness (but can cause lightheadedness during the episode) 3) Do not agree with headaches as a symptom
Symptoms	3.67 (0.52)	3.83 (0.41)	1) Canalth repositioning maneuvers are not a test for BPPV; they are a treatment! 2) This is more about vertigo than specifically about BPPV; we do not get an MRI or CT for BPPV. We also typically do not need to perform the head impulse test for BPPV. 3) ... Inaccurate description of the Dix Hallpike test
Diagnosis	3.67 (1.37)	4.00 (0.89)	1) The chat veers off into treatment of general vertigo or imbalance but is essentially accurate
Treatment	4.67 (0.52)	4.67 (0.52)	1) Surgery is essentially never indicated, not sure if I agree with risk factors they describe, Again, nystagmus is not a symptom. But most of the information is pretty good. 2) ... I would say that in MOST cases the cause of BPPV is unknown - it is much less likely to find an identifiable cause
Combined	4.33 (0.52)	4.33 (0.52)	
Total	4.19 (0.82)	4.31 (0.67)	

Scores represent the average of the two reviewer scores on a 5 point Likert Scale (1 - Strongly Disagree, 2 - Disagree, 3 - Neutral, 4 - Agree, 5 - Strongly Agree)

## Conclusions

- This analysis used BPPV to represent a common disease in the field of otolaryngology to assess and compare online information from two different modalities, a traditional search engine versus ChatGPT.
- Our results suggest information on ChatGPT is more difficult to read, of lower quality, and more difficult to understand compared to information on Google searches.
- Further research should explore the readability, quality, understandability, and accuracy on popular chatbots across a variety of diseases.
- Analyses such as this one should be intermittently revisited as developers release updates and new features.
- Over the last few decades, search engines revolutionized the access and knowledge patients had to health information, which was not without its own shortcomings and misinformation. ChatGPT and other chatbots may represent the next informational renaissance that patients will use, and providers should be aware of to offer the best care to their patients.

#### Disclosures

The authors have no potential conflicts of interest to disclose.

#### Contact Information

Please reach out with any questions to Jeffrey Bellinger – [jrb6dup@virginia.edu](mailto:jrb6dup@virginia.edu)