ORIGINAL RESEARCH

Laryngoscope
Investigative Otolaryngology

Search for medical information for chronic rhinosinusitis through an artificial intelligence ChatBot

Arsany Yassa BA[1] | **Olivia Ayad BS, MSc**[2,3,4] | **David Avery Cohen MD**[5] | **Aman M. Patel BA**[1] | **Ved A. Vengsarkar BS**[1] | **Michael S. Hegazin DO**[1] | **Andrey Filimonov MD, PharmD**[1] | **Wayne D. Hsueh MD**[1,6,7] | **Jean Anderson Eloy MD, FACS, FARS**[1,6,7,8,9]

[1]Department of Otolaryngology—Head and Neck Surgery, Rutgers New Jersey Medical School, Newark, New Jersey, USA

[2]Department of Architecture and Territory, Mediterranean University of Reggio Calabria, Calabria, Italy

[3]Department of Landscape Architecture, International Credit Hours Engineering Programs of Ain Shams University, Cairo, Egypt

[4]Arclivia, Bayonne, NJ, United States

[5]Department of Otolaryngology—Head and Neck Surgery, University of Florida College of Medicine, Gainesville, Florida, USA

[6]Department of Ophthalmology and Visual Science, Rutgers New Jersey Medical School, Newark, New Jersey, USA

[7]Center for Skull Base and Pituitary Surgery, Neurological Institute of New Jersey, Rutgers New Jersey Medical School, Newark, New Jersey, USA

[8]Department of Neurological Surgery, Rutgers New Jersey Medical School, Newark, New Jersey, USA

[9]Department of Otolaryngology and Facial Plastic Surgery, Cooperman Barnabas Medical Center—RWJBarnabas Health, Livingston, New Jersey, USA

Correspondence
Jean Anderson Eloy, Department of Otolaryngology—Head and Neck Surgery, Rutgers New Jersey Medical School, 90 Bergen St., Suite 8100, Newark, NJ 07103, USA.
Email: jean.anderson.eloy@gmail.com

Abstract

Objectives: Artificial intelligence is evolving and significantly impacting health care, promising to transform access to medical information. With the rise of medical misinformation and frequent internet searches for health-related advice, there is a growing demand for reliable patient information. This study assesses the effectiveness of ChatGPT in providing information and treatment options for chronic rhinosinusitis (CRS).

Methods: Six inputs were entered into ChatGPT regarding the definition, prevalence, causes, symptoms, treatment options, and postoperative complications of CRS. International Consensus Statement on Allergy and Rhinology guidelines for Rhinosinusitis was the gold standard for evaluating the answers. The inputs were categorized into three categories and Flesch–Kincaid readability, ANOVA and trend analysis tests were used to assess them.

Results: Although some discrepancies were found regarding CRS, ChatGPT's answers were largely in line with existing literature. Mean Flesch Reading Ease, Flesch–Kincaid Grade Level and passive voice percentage were (40.7%, 12.15%, 22.5%) for basic information and prevalence category, (47.5%, 11.2%, 11.1%) for causes and symptoms category, (33.05%, 13.05%, 22.25%) for treatment and complications, and

(40.42%, 12.13%, 18.62%) across all categories. ANOVA indicated no statistically significant differences in readability across the categories (p-values: Flesch Reading Ease = 0.385, Flesch–Kincaid Grade Level = 0.555, Passive Sentences = 0.601). Trend analysis revealed readability varied slightly, with a general increase in complexity.

Conclusion: ChatGPT is a developing tool potentially useful for patients and medical professionals to access medical information. However, caution is advised as its answers may not be fully accurate compared to clinical guidelines or suitable for patients with varying educational backgrounds.

Level of evidence: 4.

KEYWORDS
artificial intelligence, chronic rhinosinusitis, ICAR, medical information systems, readability

1 | INTRODUCTION

Artificial intelligence (AI) is a rapidly growing field that has the potential to revolutionize many industries, including health care.[1] One of the most flourishing subfields of AI is machine learning (ML), which has gained great recognition since the 1990's.[2] ML allows computers to learn by automatically recognizing significant patterns and relations within large amounts of data without the need for explicit programming.[3] In recent years, ML's algorithm has witnessed great improvements, allowing its applications to become beneficial across many fields.[4] Tasks that would normally require human intelligence such as understanding natural language, recognizing images, and making decisions are now being performed by AI.[5]

With the rise of technology, the use of the internet to search for health-related information (HRI) has become readily accessible for many people around the globe. However, medical information may be inappropriate or even harmful because of existing unverified content and a lack of strict online regulations.[6] Moreover, even if the information is accurate, some resources may use language above the lay level understanding of the public, rendering it effectively inaccessible.[6]

With the exponential evolution of online search demands comes the growing need for AI to revolutionize how we access dependable medical information. AI has already been successfully applied in the health care field in recent years. In the field of Neurology, an AI system was developed to restore the control of movement in patients with quadriplegia.[7] In the field of Dermatology, AI-based tools are being used to evaluate the severity of psoriasis[8] and to distinguish between onychomycosis and healthy nails.[9] In Otolaryngology, Powell et al. provided a proof of concept that human phonation can be decoded by AI to help in the diagnosis of voice disorders.[10] In the field of ophthalmology, researchers at Google developed and trained a deep convolutional neural network on thousands of retinal fundus images to classify diabetic retinopathy and macular edema in adults with diabetes.[11] In primary care fields, physicians can utilize AI to transcribe their notes, analyze patient discussions, and automatically input the necessary information into EHR systems.[12] Nevertheless, AI's ability to provide accurate and comprehensible medical information for various medical topics and disorders has yet to be proven via extensive demonstration.[13] Therefore, in this study, we sought to examine whether an AI ChatBot can provide accurate, comprehensive, and understandable information on chronic rhinosinusitis (CRS). CRS was chosen as the prototype for this research due to several compelling reasons. Firstly, CRS is one of the most prevalent conditions in Otolaryngology, affecting approximately 11% of the population and accounting for 15% of otolaryngologic outpatient consultations. This condition leads to significant morbidity, impacting the quality of life of millions of individuals.[14,15] In the US alone, there are over 30 million physician visits related to CRS annually, a figure that exceeds the number of medical visits for hypertension.[16] CRS's high health care burden and clinical complexity with its wide range of symptoms, causes, and treatments provides a robust test for AI-generated medical content. By evaluating ChatGPT's performance in providing information on a condition as widespread and multifaceted as CRS, we aim to assess its potential as a reliable tool for medical education and patient information.

2 | MATERIALS AND METHODS

2.1 | Generating medical information

The following data were generated on April 1, 2024. The website is accessible through OpenAI.

To examine ChatGPT's ability to respond with appropriate medical information, we provided the AI ChatBot with inputs in the form of questions about CRS and recorded the responses. These inputs included queries about CRS symptoms and questions about the disorder. A total of six unique ChatGPT outputs were examined, corresponding to the six questions posed. The requests were then categorized into several categories to fully evaluate the ChatBot's knowledge of CRS. ChatGPT's answers were then compared to ICAR

(International consensus statement on Allergy and rhinology guidelines for Rhinosinusitis) guidelines to evaluate their accuracy.

ChatGPT-3.5 uses an algorithm that is probabilistic in nature. In other words, it utilizes random sampling to generate a wide variety of responses, possibly including different answers to the same query. This investigation only included ChatGPT's initial answer to each query without regenerating the answers. Additional clarifications or explanations were not permitted. All queries were entered into a ChatGPT account owned by the author in a single day, guaranteeing accuracy in grammar and syntax. We placed each query into a new dialogue window to eliminate confounding factors and guarantee the accuracy and precision of the responses since ChatGPT-3.5 can adjust from the details of every interaction.

The study did not require institutional review board (IRB) approval from Rutgers New Jersey Medical School IRB since it does not utilize human participants and no patient identifying information was used.

2.2 | Data analysis

We conducted a thorough linguistic examination to assess the readability and complexity of the AI-generated responses. To accomplish this, we utilized the Flesch Reading Ease and Flesch–Kincaid Grade Level metrics, which are established measures that provide insights into the readability and the educational level required to comprehend the material. These metrics have been widely applied in numerous studies to evaluate the accessibility of online content for individuals facing various conditions, such as ACL injury, glaucoma, and dog bites.[17–19] In our study, we first determined the average (mean) and variability (standard deviation) for both the Flesch Reading Ease and Flesch–Kincaid Grade Level indices, as well as for the percentage of passive sentences, to evaluate ChatGPT's overall readability performance. The reason we also analyzed passive sentence percentage is that it can affect the readability and comprehension of the text. Passive constructions are generally harder to read and understand, particularly for individuals with lower literacy levels.[20] Next, we organized the six questions pertaining to CRS into three distinct categories: basic information and prevalence, causes and symptoms, and treatment and complications. For each category, we computed the mean and standard deviation for the Flesch Reading Ease, Flesch–Kincaid Grade Level, and the passive sentence percentage for ChatGPT's responses. These initial steps paved the way for more thorough statistical analyses, including ANOVA and trend analysis, enabling us to investigate whether ChatGPT's readability performance varied across different question categories. ANOVA is used to compare the means of three or more groups to determine if there are any statistically significant differences among them. In this study, ANOVA was used to compare the means of the readability metrics across the three different categories of questions. The assumptions for ANOVA were also verified prior to analysis. Trend analysis was conducted to observe patterns and shifts in the readability metrics across the different categories. The mean and standard deviation for each readability metric

were plotted to visualize trends in the data. This analysis helped to identify any systematic changes in readability and complexity of the responses based on the type of information provided. This comprehensive analytical approach allowed us to deeply understand the textual qualities of the AI-generated content and thoroughly evaluate its suitability for patient education.

3 | RESULTS

3.1 | Qualitative analysis

Figure 1 illustrates the response from ChatGPT when asked about the definition of CRS. Figure 2 displays the ChatGPT-provided prevalence data for CRS. Figures 3–8 detail the causes, symptoms, treatments, and postoperative complications of CRS, respectively, as described by ChatGPT.

We first want to report the accuracy of ChatGPT's answers to each of the questions.

For the definition question (Figure 1), ChatGPT defines CRS as it is commonly defined in the medical literature—with a symptom duration of at least 12 weeks for CRS diagnosis to be established.[21] However, when compared to ICAR guidelines, ChatGPT failed to mention that establishing a diagnosis of CRS requires two of four main symptoms (Pressure/pain in face, anosmia/hyposmia, nasal obstruction/blockage, and mucopurulent nasal drainage) to be present for at least 12 weeks in addition to objective evidence on physical exam (purulence from paranasal sinuses or osteomeatal complex, polyps, edema or evidence of inflammation on nasal endoscopy or CT), since symptoms alone have low specificity for CRS diagnosis.[21–23] ChatGPT also failed to mention that the presence of polyps further classifies CRS to CRSsNP (CRS without nasal polyps) or CRSwNP (CRS with nasal polyps), an important omission as treatment differs based on disease subtype according to ICAR.[23]

As for the prevalence question (Figure 2), ChatGPT reported a prevalence of 2%–5% for CRS in the US and Europe, and that the prevalence has been increasing over the past few years. However, ChatGPT's source for this percentage is unclear since many sources reports a rate of >10% in the US and Europe for the general population.[24,25] ICAR, on the other hand, reports a prevalence in the range of 2.1% to 13.8% in the US and 6.9% to 27.1% in Europe.[23] ICAR attributes the large difference in range between the lower and the upper limits to the fact that the diagnosis of CRS requires objective evidence, which makes it difficult to determine a true prevalence by ICAR.[23]

When asked about the causes of CRS (Figure 3), ChatGPT's answer aligned with the ICAR guidelines in that the exact etiology of CRS involves multiple factors.[23,26–30] The ChatBot then proceeded to mention some of the cited causes, including lifestyle and environmental factors (i.e., occupational hazards).[31] ChatGPT also explained how each factor can contribute to or increase the risk of developing CRS (i.e., structural abnormalities and nasal polyps can contribute to developing CRS by blocking the sinuses).[32–34] In the literature, some

Anonymous
"What is the definition of Chronic Rhinosinusitis?"

ChatGPT
Chronic rhinosinusitis (CRS) is a persistent inflammation of the nasal passages and sinuses. It typically lasts for at least 12 weeks and can cause symptoms such as nasal congestion, facial pain and pressure, and difficulty breathing through the nose. CRS can be caused by a variety of factors, including allergies, structural abnormalities in the nasal passages, and infections. It is a common condition and can have a significant impact on a person's quality of life.

FIGURE 1 ChatGPT's response to the question "What is the definition of chronic rhinosinusitis?"

Anonymous
Can you tell me the prevalence of Chronic Rhinosinusitis?

ChatGPT
The prevalence of chronic rhinosinusitis (CRS) varies depending on the population studied and the criteria used to diagnose the condition. However, it is estimated that CRS affects between 2-5% of the general population in the US and Europe. The condition is more common in certain groups of people, such as those with allergies, asthma, and certain genetic disorders. The prevalence of CRS tends to increase with age and it is more common in men than in women. Additionally, studies have shown that the prevalence of CRS has been increasing in recent years.

FIGURE 2 ChatGPT's response to the question "Can you tell me the prevalence of chronic rhinosinusitis?"

Anonymous
Can you tell me the causes of Chronic Rhinosinusitis?

ChatGPT
The exact cause of chronic rhinosinusitis (CRS) is not well understood, and it is likely that multiple factors contribute to the development of the condition. Some of the most commonly cited causes of CRS include:

- Allergies: Allergic reactions can cause inflammation in the nasal passages and sinuses, leading to CRS.
- Infections: Bacterial, viral, or fungal infections can cause inflammation in the sinuses and nasal passages, leading to CRS.
- Structural abnormalities: Certain structural abnormalities in the nasal passages, such as a deviated septum, can block the sinuses and increase the risk of CRS.
- Nasal polyps: Small growths in the nasal passages called nasal polyps can block the sinuses and increase the risk of CRS.
- Immune system dysfunction: Some experts believe that CRS may be caused by an abnormal immune response in the sinuses and nasal passages.
- Environmental factors: Exposure to pollutants, irritants, and certain occupational hazards may increase the risk of CRS.
- Lifestyle factors: such as smoking and alcohol consumption can also play a role in the development of CRS.

It's important to note that the cause of CRS can be multifactorial and, in some cases, the cause may not be clear.

FIGURE 3 ChatGPT's response to the question "Can you tell me the causes of chronic rhinosinusitis?"

lifestyle factors are proven in affecting the development of CRS such as exposure to hair-care products, dust, fumes, cleaning agents, allergens, and even cold, dry and low elevation areas.[35–37] Another study even found a 2.5 fold increased risk of CRS development with residential proximity (within 2 km radius) to commercial pesticides application.[38] However, according to ICAR guidelines, the link between environmental or lifestyle factors and CRS is very weakly supported.[23] The lifestyle factor that is most strongly associated with CRS is tobacco smoke exposure, according to ICAR.[23] Even though ChatGPT

mentioned that smoking is associated with CRS, it did not highlight the importance of this risk factor and only mentioned it as a potential factor. In other words, in this case, ChatGPT's answer is not entirely correct according to ICAR but correct according to other sources.

When asked about the symptoms of CRS (Figure 4), ChatGPT clearly provided the most common symptoms and even the less common ones according to ICAR.[22,23,39] At the end, it also correctly provided some caveats: (1) Symptoms may not always be persistent and can also be intermittent in nature. (2) CRS symptoms may fluctuate in

FIGURE 4 ChatGPT's response to the question "What are the symptoms of chronic rhinosinusitis?"
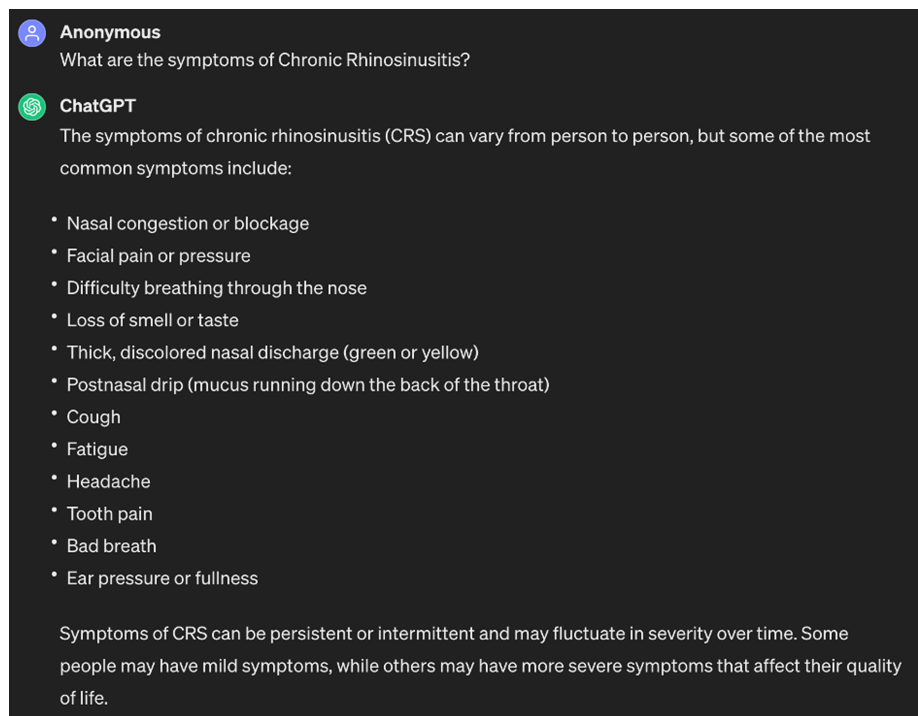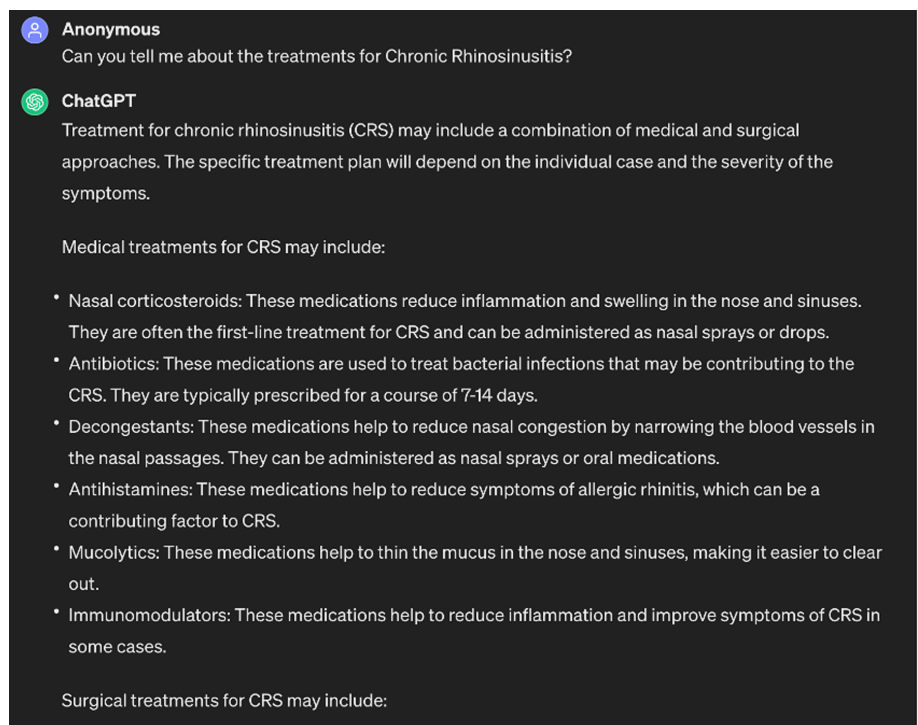
Anonymous
What are the symptoms of Chronic Rhinosinusitis?

ChatGPT
The symptoms of chronic rhinosinusitis (CRS) can vary from person to person, but some of the most common symptoms include:

- Nasal congestion or blockage
- Facial pain or pressure
- Difficulty breathing through the nose
- Loss of smell or taste
- Thick, discolored nasal discharge (green or yellow)
- Postnasal drip (mucus running down the back of the throat)
- Cough
- Fatigue
- Headache
- Tooth pain
- Bad breath
- Ear pressure or fullness

Symptoms of CRS can be persistent or intermittent and may fluctuate in severity over time. Some people may have mild symptoms, while others may have more severe symptoms that affect their quality of life.

FIGURE 5 ChatGPT's response to the question "Can you tell me about the treatments for chronic rhinosinusitis?"

Anonymous
Can you tell me about the treatments for Chronic Rhinosinusitis?

ChatGPT
Treatment for chronic rhinosinusitis (CRS) may include a combination of medical and surgical approaches. The specific treatment plan will depend on the individual case and the severity of the symptoms.

Medical treatments for CRS may include:

- Nasal corticosteroids: These medications reduce inflammation and swelling in the nose and sinuses. They are often the first-line treatment for CRS and can be administered as nasal sprays or drops.
- Antibiotics: These medications are used to treat bacterial infections that may be contributing to the CRS. They are typically prescribed for a course of 7-14 days.
- Decongestants: These medications help to reduce nasal congestion by narrowing the blood vessels in the nasal passages. They can be administered as nasal sprays or oral medications.
- Antihistamines: These medications help to reduce symptoms of allergic rhinitis, which can be a contributing factor to CRS.
- Mucolytics: These medications help to thin the mucus in the nose and sinuses, making it easier to clear out.
- Immunomodulators: These medications help to reduce inflammation and improve symptoms of CRS in some cases.

Surgical treatments for CRS may include:

severity. (3) Some patients might have mild symptoms, but others may suffer severe symptoms that affect their quality of life. All of the caveats are consistent with ICAR.[23,30,40]

Concerning the treatment options for CRS (Figures 5 and 6), ChatGPT correctly answered that treatment could be medical or surgical.[23,41] The answer also contained a list of medically oriented treatments with accurate corresponding rationales and another list of procedures and surgical techniques.[22,23,42–45] ChatGPT also delved into some accurate details as to how each procedure is performed.[23,44,45] Again, ChatGPT aptly mentioned that the treatment plan may vary depending on each individual's case. However, ChatGPT mentioned mucolytics as part of the medical management for CRS, but ICAR did not provide any recommendations regarding their use due to insufficient evidence.[23] ChatGPT also did not reveal

Surgical treatments for CRS may include:

- Functional endoscopic sinus surgery (FESS): This is a minimally invasive procedure that uses an endoscope (a thin, flexible tube with a light and camera on the end) to visualize and remove any blockages or abnormalities in the sinuses.
- Image-guided surgery: This is a surgical technique that uses imaging technology, such as computed tomography (CT) scans, to guide the surgeon during the procedure.
- Balloon sinuplasty: This is a minimally invasive procedure that uses a small balloon catheter to dilate blocked sinus ostia (the openings of the sinuses) and improve nasal airflow.

It's important to note that the treatment plan may vary depending on the individual case and the severity of the symptoms. A thorough evaluation and diagnostic workup is needed before deciding on the best course of treatment.

FIGURE 6 Rest of ChatGPT's response to the question "Can you tell me about the treatments for chronic rhinosinusitis?"

Anonymous
What are the possible post-operative complications for Chronic Rhinosinusitis operations?

ChatGPT
There are a number of potential complications that can occur after surgery for chronic rhinosinusitis (CRS). These complications can vary depending on the type of surgery that was performed and the individual patient's health status.

Here are some possible post-operative complications for CRS operations:

- Infection: As with any surgical procedure, there is a risk of infection after sinus surgery. This risk can be minimized by following proper post-operative care instructions and by taking antibiotics as prescribed.
- Bleeding: There may be some bleeding or oozing from the nose in the days following surgery. This is usually not a cause for concern, but if significant bleeding occurs, it is important to seek medical attention.
- Swelling and bruising: Some swelling and bruising around the eyes and cheeks is normal after sinus surgery. This will typically subside within a week or two.
- Scarring: Scarring can occur inside the nose and sinuses after surgery, which can lead to narrowing of the nasal passages.
- Persistent or recurrent symptoms: In some cases, the surgery may not fully resolve the patient's symptoms, and they may continue to experience nasal congestion, facial pain, or other symptoms.
- Anosmia: Loss of sense of smell may occur temporarily or permanently after sinus surgery.
- Epistaxis: Nosebleed may occur after sinus surgery.

FIGURE 7 ChatGPT's response to the question "What are the possible postoperative complications for chronic rhinosinusitis operations?"

- Anosmia: Loss of sense of smell may occur temporarily or permanently after sinus surgery.
- Epistaxis: Nosebleed may occur after sinus surgery.

It's important to note that these complications are relatively rare, and most patients experience a significant improvement in their symptoms after sinus surgery. However, it's important to be aware of the potential complications and to seek medical attention if any concerning symptoms occur.

It is always best to consult with your surgeon to understand the possible complications and risks of the surgery, as well as the expected recovery time.

FIGURE 8 Rest of ChatGPT's response to the question "What are the possible postoperative complications for chronic rhinosinusitis operations?"

that most of the people who have CRS report severe symptoms and a lack of satisfaction with the treatment options they currently undergo.[46]

ChatGPT also identified the most common postoperative complications and provided useful information on their management (Figures 7 and 8).[47] One of the most significant postoperative complications that ChatGPT identified was infection, which is a typical concern following any surgical operation[48] and aligns with ICAR.[23] This risk can be reduced, as ChatGPT correctly noted, by adhering to the proper postoperative care guidelines and taking prescribed antibiotics.[23,47–51] Prophylactic antibiotics have been shown in multiple studies, according to ICAR, to significantly

reduce the incidence of postoperative infections in individuals undergoing sinus surgery.[23,49] In addition, ChatGPT, consistent with ICAR, mentioned persistent or recurrent symptoms such as anosmia, epistaxis, swelling, bruising, and scarring, and other common postoperative complications that patients may experience after surgery.[23,50,51] In general, ChatGPT's response emphasizes the value of consulting a surgeon to learn about the risks and possible complications associated with the procedure as well as the anticipated length of recovery. To reduce the risk of complications, patients should also be provided with suitable postoperative care advice.[47,51]

3.2 | Quantitative analysis

3.2.1 | Statistical analysis

We meticulously assessed the Flesch Reading Ease, Flesch–Kincaid Grade Level, and the percentage of passive sentences across three distinct categories: basic information and prevalence, causes and symptoms, and treatment and complications.

3.3 | Overall readability metrics

Across all categories, our findings (Table 1) revealed an average Flesch Reading Ease score of 40.42 (SD = 9.43), signifying that the material is challenging for most readers. The Flesch–Kincaid Grade Level averaged at 12.13 (SD = 1.45) indicates that the content is suitable for an audience with at least a high school reading level. Passive sentence construction was employed in 18.62% (SD = 10.85%) of the sentences, suggesting a moderate use of passive voice.

3.4 | Category-specific readability metrics

Table 2 displays the readability metrics across the three questions' categories: basic information and prevalence, causes and symptoms, and treatment and complications. Mean Flesch Reading Ease, Flesch–Kincaid Grade Level and passive voice percentage were (40.7 ± 4.24, 12.15 ± 0.64, 22.5% ± 3.54%) for basic information and prevalence category, (47.5 ± 14.42, 11.2 ± 2.55, 11.1% ± 15.7%) for causes and symptoms category, (33.05 ± 3.04, 13.05 ± 0.49, 22.25% ± 12.66%) for treatment and complications. This shows variations in Flesch Reading Ease and Flesch–Kincaid Grade Level scores, indicating the complexity of each category. The treatment and complications section, for example, has the lowest Reading ease score and highest Flesch–Kincaid Grade level. Additionally, the Passive Sentences percentage highlights variations in writing style.

3.5 | Statistical significance

Applying ANOVA tests to these metrics, we ascertained the p-values: Flesch Reading Ease ($p = .385$), Flesch–Kincaid Grade Level ($p = .555$), and Passive Sentences ($p = .601$), all suggesting no statistically significant differences in the readability across the different categories.

3.6 | Trend analysis

A trend analysis with standard deviations was conducted to visualize the readability shifts between categories (Figure 9). The Flesch Reading Ease scores displayed a nominal increase from basic information to causes and symptoms but dropped in the treatment and
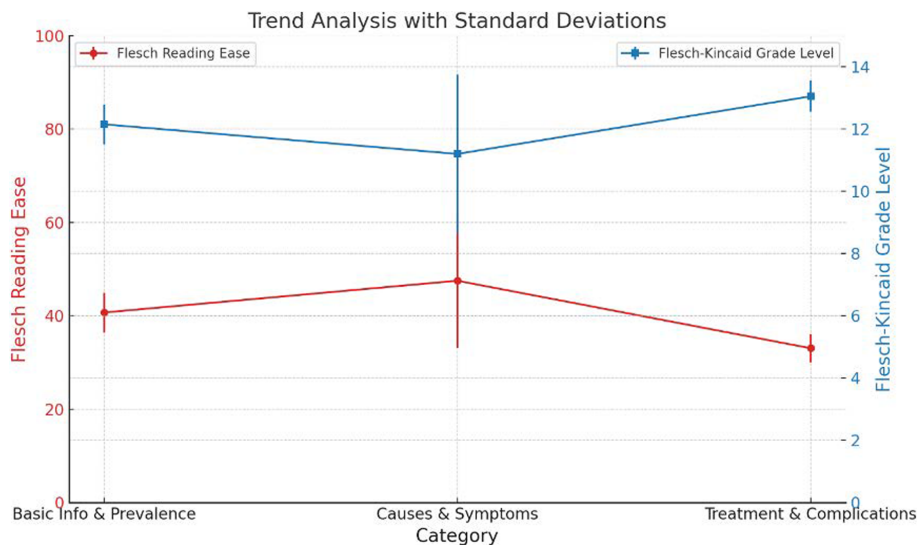
TABLE 1 Overall and question-specific readability metrics for chronic rhinosinusitis.

Category	Flesch reading ease score	Flesch–Kincaid grade level	Passive sentences
Q1: Definition of chronic rhinosinusitis	37.7	12.6	25%
Q2: Prevalence of chronic rhinosinusitis	43.7	11.7	20%
Q3: Causes of chronic rhinosinusitis	37.3	13	22.2%
Q4: Symptoms of chronic rhinosinusitis	57.7	9.4	0%
Q5: Treatments for chronic rhinosinusitis	30.9	13.4	31.2%
Q6: Postoperative complications	35.2	12.7	13.3%
Mean ± Standard deviation	40.42 ± 9.43	12.13 ± 1.45	18.62 ± 10.85%

TABLE 2 Readability metrics by category for chronic rhinosinusitis.

Category	Flesch reading ease score (mean ± standard deviation)	Flesch–Kincaid grade level (mean ± standard deviation)	Passive sentences (%) (mean ± standard deviation)
Basic information and prevalence	40.7 ± 4.24	12.15 ± 0.64	22.5 ± 3.54%
Causes and symptoms	47.5 ± 14.42	11.2 ± 2.55	11.1 ± 15.7%
Treatment and complications	33.05 ± 3.04	13.05 ± 0.49	22.25 ± 12.66%

complications category. The Flesch–Kincaid Grade Level indicated a consistent upward trajectory, signifying increasing textual complexity. The standard deviations highlighted the variability within each category, particularly pronounced in the causes and symptoms segment for both Flesch Reading Ease and Passive Sentences.

4 | DISCUSSION

ChatGPT was able to respond to all questions, from defining CRS to describing the causes of disease, symptoms, postoperative complications, and even detailing the roles that rehabilitation and patient education may play. Each response's sentences closely adhered to appropriate grammatical rules and sentence structure.

As for the statistical analysis, the average Flesch Reading Ease score for all categories combined was moderately low (Table 1). This score aligns with the style typically found in academic or professional documents, suggesting that, while ChatGPT's responses are informative, they may not be entirely accessible to individuals without a higher educational background.

Interestingly, the readability scores did not vary significantly across different content categories, as evidenced by the ANOVA test results. This consistency in complexity and readability is beneficial in one aspect, as it suggests that ChatGPT maintains a uniform level of language complexity regardless of the topic complexity. However, it also implies that the AI does not automatically adjust its language complexity in response to the varying difficulty levels of the subject matter. For instance, one might expect the language around basic information and prevalence to be more accessible than that regarding treatment and complications, which inherently deals with more complex medical procedures and concepts.

Even though ANOVA found no significant differences in the readability scores, it is important to mention that the increasing trend in the Flesch–Kincaid Grade Level across the categories may reflect the intrinsic complexity of the medical information as it progresses from

basic definitions to detailed medical procedures and potential complications. However, notably, the causes and symptoms category demonstrated a higher Flesch Reading Ease and a lower Flesch–Kincaid Grade Level compared to the other categories (Table 2). The lower percentage of passive sentences in this category may contribute to its relative readability.

It is important to mention that relying solely on readability metrics to determine if medical material is appropriate for patients has significant drawbacks. Readability metrics like the Flesch–Kincaid assess linguistic simplicity but overlook critical aspects such as health literacy and content accuracy.

Health literacy involves understanding medical terms and concepts, which readability metrics do not measure. Even easy-to-read text can be confusing if it contains medical jargon or complex ideas that are not clearly explained. Furthermore, readability metrics do not ensure the accuracy of the information, which is crucial for patient safety and effective health management.

To create truly patient-appropriate medical material, a comprehensive approach is needed. This approach should combine readability assessments with considerations of health literacy and content accuracy. This means using plain language, explaining medical terms, incorporating visual aids, and having medical professionals review the content for accuracy and relevance.

In the context of patient education, it is also crucial to consider the health literacy of the audience. The National Assessment of Adult Literacy reports that only 12% of U.S. adults have proficient health literacy.[52] Given that many adults may struggle with complex health information, the findings of this study suggest that there is a need for further optimization to enhance readability and ensure that the information is comprehensible to all patients, irrespective of their educational background. Future iterations of AI-driven platforms could focus on dynamically adjusting language complexity based on the user's comprehension level, potentially assessed through preliminary questions or interactive dialogue. Furthermore, incorporating visual aids and interactive elements could enhance

understanding and engagement, particularly for topics that are inherently complex.[53,54]

This study has a few limitations, however. The collection of data was performed at a specific moment in time, which poses a challenge in the rapidly changing domain of AI. Furthermore, the qualitative approach of this study inherently carries the potential for some level of investigator bias. The study also acknowledges the impact of variations in ChatGPT's responses due to differences in how questions are phrased, alongside the restricted range of question sources, as potential areas for further exploration. Future studies could benefit from comparing ChatGPT to other AI models to gain a broader perspective on its effectiveness and limitations in medical and patient education contexts. Future studies can also explore the variance in ChatGPT's responses over multiple instances and with follow-up questions. This is a valuable area for future research and could investigate the consistency and reliability of the AI over repeated interactions, providing a more comprehensive understanding of its performance. Nonetheless, we believe that despite these limitations, our study offers valuable insights into an information source that is increasingly prevalent in today's digital age.

ChatGPT offers several notable advantages in providing medical information. Its primary strength lies in accessibility; it allows users to obtain medical information quickly and easily, regardless of their location. This can be particularly beneficial for individuals in remote areas or those who face barriers to accessing health care professionals. The speed at which ChatGPT generates responses is another significant advantage, providing instant answers to medical queries.

However, there are significant disadvantages to using ChatGPT for medical information. One major issue is accuracy; the inaccuracies in some of ChatGPT's answers highlight the limitation that ChatGPT's responses are only as reliable as the data it has been trained on. Another disadvantage is the omission of critical details; for example, ChatGPT failed to mention the need for objective evidence in diagnosing CRS, which is crucial for accurate diagnosis and treatment planning. The readability and comprehensibility of ChatGPT's responses also pose a challenge, as the analysis revealed that its outputs are often at a high reading level, making them unsuitable for individuals with limited health literacy. Additionally, ChatGPT's responses lack the nuance and context-specific advice that human health care providers can offer, limiting its ability to tailor information based on individual patient histories or specific circumstances.

To enhance the quality of AI-generated medical information, several methods can be implemented. Integrating AI systems with verified medical databases like PubMed and Medline ensures reliance on current and reliable sources. Regularly updating training data with the latest medical research and guidelines can reduce outdated information. A human-in-the-loop approach, where medical professionals review AI-generated content, can identify and correct discrepancies. Improving the AI's contextual understanding and prioritization of critical clinical information can enhance response relevance and completeness. Increasing transparency about how responses are generated and sourced can also build trust and reliability.

Even though ChatGPT should never be considered a replacement to medical professionals' advice, it can enhance professional medical information by serving as a supplementary resource to provide preliminary information and answer common questions, preparing patients for consultations. It can also explain complex terms in simple language to improve understanding, but medical professionals should review its responses to ensure accuracy. By adjusting responses based on health literacy levels and incorporating visual aids, ChatGPT can make complex information more accessible.

In summary, while ChatGPT presents a promising tool for enhancing access to medical information and can serve as a useful starting point for patient education and general inquiries, it should not replace professional medical advice. Ensuring the accuracy, completeness, and readability of its responses, and providing contextually relevant and individualized information, are critical areas for future development. Also, the current level of language complexity highlights an area for improvement. To fully harness the educational potential of AI in health care, there must be a concerted effort to tailor the readability of content to the diverse needs of patients, ensuring that information is not only accurate but also accessible to those it intends to serve.

5 | CONCLUSION

ChatGPT is a rapidly developing tool that may, soon, become an invaluable asset to the health care system. As of today, this tool may be useful for patients who have difficulty accessing medical information due to geographic or financial constraints. The AI ChatBot has a user-friendly interface and a unique ability to understand a patient's natural language. Nevertheless, the content generated by the ChatBot may be inaccurate, biased, or hard for many patients to understand so, while promising, it is not yet time for AI to be considered a reliable source for attaining medical information especially for patients with limited health literacy. Moreover, because each patient's case is unique, AI is not yet be able to provide precise recommendations to individual patients in the same way that human physicians are able to provide.

CONFLICT OF INTEREST STATEMENT
The authors declare that they have no conflict of interest.

ORCID
Arsany Yassa https://orcid.org/0009-0009-1474-4346
David Avery Cohen https://orcid.org/0000-0001-8263-6341
Aman M. Patel https://orcid.org/0000-0003-0794-041X
Andrey Filimonov https://orcid.org/0000-0002-4285-5862
Jean Anderson Eloy https://orcid.org/0000-0003-2893-7818

REFERENCES

1. Wen Z, Huang H. The potential for artificial intelligence in healthcare. *J Commer Biotechnol*. 2022;27(4):217-225.
2. Siwach M, Mann S. A compendium of various applications of machine learning. *Int J Res Eng Technol*. 2022;9:1141-1144.
3. Angelov PP, Gu X. *Empirical Approach to Machine Learning*. Springer; 2019.
4. Royal Society (Great Britain). *Machine Learning: the Power and Promise of Computers that Learn by Example: an Introduction*. Royal Society; 2017.
5. Mohammad SM. Ethics sheets for AI tasks. *arXiv preprint arXiv: 2107.01183* 2021.
6. Murray E, Lo B, Pollack L, et al. The impact of health information on the internet on health care and the physician-patient relationship: national US survey among 1.050 US physicians. *J Med Internet Res*. 2003;5(3):e17.
7. Bouton CE, Shaikhouni A, Annetta NV, et al. Restoring cortical control of functional movement in a human with quadriplegia. *Nature*. 2016; 533(7602):247-250.
8. Shrivastava VK, Londhe ND, Sonawane RS, Suri JS. A novel and robust Bayesian approach for segmentation of psoriasis lesions and its risk stratification. *Comput Methods Prog Biomed*. 2017;150:9-22.
9. Han SS, Park GH, Lim W, et al. Deep neural networks show an equivalent and often superior performance to dermatologists in onychomycosis diagnosis: automatic construction of onychomycosis datasets by region-based convolutional deep neural network. *PLoS One*. 2018; 13(1):e0191493.
10. Powell ME, Rodriguez Cancio M, Young D, et al. Decoding phonation with artificial intelligence (DeP AI): proof of concept. *Laryngoscope Investig Otolaryngol*. 2019;4(3):328-334.
11. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316(22):2402-2410.
12. Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. *J Family Med Prim Care*. 2019;8(7):2328-2331.
13. Liévin V, Hother CE, Motzfeldt AG, Winther O. Can Large Language Models Reason about Medical Questions? *Patterns (NY)*. 2023;5: 100943.
14. Shah RK. Anatomical variation of nose and paranasal sinuses among patients with chronic rhinosinusitis on computer tomography. *Birat J Health Sci*. 2022;7(1):1732-1735.
15. Wahid NW, Smith R, Clark A, Salam M, Philpott C. The socioeconomic cost of chronic rhinosinusitis study. *Rhinology*. 2020;58(2): 112-125.
16. Grzegorzek T, Kolebacz B, Stryjewska-Makuch G, Kasperska-Zając A, Misiołek M. The influence of selected preoperative factors on the course of endoscopic surgery in patients with chronic rhinosinusitis. *Adv Clin Exp Med*. 2014;23(1):69-78.
17. Dadabhoy M, Awal D. 152 assessment of quality and readability of online patient education materials relating to dog bites. *Br J Surg*. 2022;109(Supplement_6):znac269.
18. Crabtree L, Lee E. Assessment of the readability and quality of online patient education materials for the medical treatment of open-angle glaucoma. *BMJ Open Ophthalmol*. 2022;7(1):e000966.
19. Gao B, Shamrock AG, Gulbrandsen TR, et al. Can patients read, understand, and act on online resources for anterior cruciate ligament surgery? *Orthop J Sports Med*. 2022;10(7).
20. Flinton D, Singh MK, Haria K. Readability of internet-based patient information for radiotherapy patients. *J Radiother Pract*. 2018;17(2): 142-150.
21. Sedaghat AR. Chronic rhinosinusitis. *Am Fam Physician*. 2017;96(8): 500-506.
22. Rosenfeld RM, Piccirillo JF, Chandrasekhar SS, et al. Clinical practice guideline (update): adult sinusitis. *Otolaryngol Head Neck Surg*. 2015; 152(2_suppl):S1-S39.
23. Orlandi RR, Kingdom TT, Smith TL, et al. International consensus statement on allergy and rhinology: rhinosinusitis 2021. *Int Forum Allergy Rhinol*. 2021;11(3):213-739.
24. Zhou F, Zhang T, Jin Y, et al. Developments and emerging trends in the global treatment of chronic rhinosinusitis from 2001 to 2020: a systematic bibliometric analysis. *Front Surg*. 2022;9:851923.
25. Hastan D, Fokkens WJ, Bachert C, et al. Chronic rhinosinusitis in Europe–an underestimated disease. A GA2LEN study. *Allergy*. 2011; 66(9):1216-1223.
26. Bhattacharyya N. Bacterial infection in chronic rhinosinusitis: a controlled paired analysis. *Am J Rhinol*. 2005;19(6):544-548.
27. Sedaghat AR, Gray ST, Caradonna SD, Caradonna DS. Clustering of chronic rhinosinusitis symptomatology reveals novel associations with objective clinical and demographic characteristics. *Am J Rhinol Allergy*. 2015;29(2):100-105.
28. Busaba NY, Siegel NS, Salman SD. Microbiology of chronic ethmoid sinusitis: is this a bacterial disease? *Am J Otolaryngol*. 2004;25(6): 379-384.
29. Kutluhan A, Çetin H, Kale H, et al. Comparison of natural ostiodilatation and endoscopic sinus surgery in the same patient with chronic sinusitis. *Braz J Otorhinolaryngol*. 2020;86:56-62.
30. Aurora R, Chatterjee D, Hentzleman J, Prasad G, Sindwani R, Sanford T. Contrasting the microbiomes from healthy volunteers and patients with chronic rhinosinusitis. *JAMA Otolaryngol Head Neck Surg*. 2013;139(12):1328-1338.
31. Koh D, Kim H, Han S. The relationship between chronic rhinosinusitis and occupation: the 1998, 2001, and 2005 Korea national health and nutrition examination survey (KNHANES). *Am J Ind Med*. 2009;52(3): 179-184.
32. Feng CH, Miller MD, Simon RA. The united allergic airway: connections between allergic rhinitis, asthma, and chronic sinusitis. *Am J Rhinol Allergy*. 2012;26(3):187-190.
33. Ramadan HH, Fornelli R, Ortiz AO, Rodman S. Correlation of allergy and severity of sinus disease. *Am J Rhinol*. 1999;13(5):345-348.
34. Kasapoğlu F, Onart S, Basut O. Preoperative evaluation of chronic rhinosinusitis patients by conventional radiographies, computed tomography and nasal endoscopy. *Turk J Ear Nose Throat*. 2009;19(4): 184-191.
35. Ghatee MA, Kanannejad Z, Nikaein K, Fallah N, Sabz G. Geo-climatic risk factors for chronic rhinosinusitis in southwest Iran. *PLoS One*. 2023;18(7):e0288101.
36. Alkholaiwi FM, Almutairi RR, Alrajhi DM, Alturki BA, Almutairi AG, Binyousef FH. Occupational and environmental exposures, the association with chronic sinusitis. *Saudi Med J*. 2022;43(2):125-131.
37. Clarhed UK, Johansson H, Svendsen MV, Torén K, Møller Fell AK, Hellgren J. Occupational exposure and the risk of new-onset chronic rhinosinusitis–a prospective study 2013-2018. *Rhinology*. 2020;58: 597-604.
38. Yang H, Paul KC, Cockburn MG, et al. Residential proximity to a commercial pesticide application site and risk of chronic rhinosinusitis. *JAMA Otolaryngol Head Neck Surg*. 2023;149(9):773-780.
39. Savović S, Buljčik-Čupić M, Jovančević L, Kljajić V, Lemajić-Komazec S, Dragičević D. Frequency and intensity of symptoms in patients with chronic rhinosinusitis. *Srp Arh Celok Lek*. 2019;147(1–2): 34-38.
40. Eloy P, Poirrier AL, De Dorlodot C, Van Zele T, Watelet JB, Bertrand B. Actual concepts in rhinosinusitis: a review of clinical presentations, inflammatory pathways, cytokine profiles, remodeling, and management. *Curr Allergy Asthma Rep*. 2011;11:146-162.
41. Ragab S, Scadding GK, Lund VJ, Saleh H. Treatment of chronic rhinosinusitis and its effects on asthma. *Eur Respir J*. 2006;28(1):68-74.
42. Rosenfeld RM, Andes D, Bhattacharyya N, et al. Clinical practice guideline: adult sinusitis. *Otolaryngol Head Neck Surg*. 2007;137(3): S1-S31.

43. Dass K, Peters AT. Diagnosis and management of rhinosinusitis: highlights from the 2015 practice parameter. *Curr Allergy Asthma Rep*. 2016;16:1-7.

44. Tomazic PV, Stammberger H, Braun H, et al. Feasibility of balloon sinuplasty in patients with chronic rhinosinusitis: the Graz experience. *Rhinology*. 2013;51(2):120-127.

45. Koskinen A, Penttilä M, Myller J, et al. Endoscopic sinus surgery might reduce exacerbations and symptoms more than balloon sinuplasty. *Am J Rhinol Allergy*. 2012;26(6):e150-e156.

46. Palmer JN, Messina JC, Biletch R, Grosel K, Mahmoud RA. A cross-sectional, population-based survey of US adults with symptoms of chronic rhinosinusitis. *Allergy Asthma Proc*. 2019;40(1):48-56.

47. Hopkins C, Browne JP, Slack R, et al. Complications of surgery for nasal polyposis and chronic rhinosinusitis: the results of a national audit in England and Wales. *Laryngoscope*. 2006;116(8):1494-1499.

48. Giri VP, Giri OP, Bajracharya S, et al. Risk of acute kidney injury with amikacin versus gentamycin both in combination with metronidazole for surgical prophylaxis. *J Clin Diagn Res*. 2016;10(1):FC09.

49. Bhattacharyya N, Fried MP. The accuracy of computed tomography in the diagnosis of chronic rhinosinusitis. *Laryngoscope*. 2003;113(1): 125-129.

50. Tan BK, Chandra RK. Postoperative prevention and treatment of complications after sinus surgery. *Otolaryngol Clin N Am*. 2010;43(4): 769-779.

51. Jeican II, Trombitas V, Crivii C, et al. Rehabilitation of patients with chronic rhinosinusitis after functional endoscopic sinus surgery. *Balneo PRM Res J*. 2021;12(1):65-72.

52. Office of the Surgeon General. *Office of Disease Prevention and Health Promotion. Proceedings of the Surgeon General's Workshop on Improving Health Literacy*. National Institutes of Health; 2006.

53. Hafner C, Schneider J, Schindler M, Braillard O. Visual aids in ambulatory clinical practice: experiences, perceptions and needs of patients and healthcare professionals. *PLoS One*. 2022;17(2): e0263041.

54. Stewart JA, Wood L, Wiener J, et al. Visual teaching aids improve patient understanding and reduce anxiety prior to a colectomy. *Am J Surg*. 2021;222(4):780-785.