European Archives of Oto-Rhino-Laryngology (2024) 281:2081–2086
https://doi.org/10.1007/s00405-023-08104-8

HEAD AND NECK

Exploring the potential of Chat-GPT as a supportive tool for sialendoscopy clinical decision making and patient information support

Carlos M. Chiesa-Estomba[1,2,6,7,12,13] · Jerome R. Lechien[3,6,7,12] · Luigi A. Vaira[4,5,6,7,12] · Aina Brunet[6,7] · Giovanni Cammaroto[6,7,8,12] · Miguel Mayo-Yanez[6,7,9,12] · Alvaro Sanchez-Barrueco[6,7,10,11] · Carlos Saga-Gutierrez[1,2,6,7]

Received: 19 June 2023 / Accepted: 29 June 2023 / Published online: 5 July 2023

Abstract

Introduction Sialendoscopy has emerged in the last decades as a groundbreaking technique, offering a minimally invasive approach for exploring and managing salivary gland disorders. More recently, the advent of chatbots, powered by advanced natural processing language and artificial intelligence algorithms, has revolutionized the way healthcare professionals and patients access and analyze medical information and potentially will support soon the clinical decision-making process.

Materials and methods A prospective, cross-sectional study was designed to assess the level of agreement between Chat-GPT and 10 expert sialendoscopists aiming the capabilities of Chat-GPT to further improve the management of salivary gland disorders.

Results The mean level of agreement was 3.4 (SD: 0.69; Min: 2, Max: 4) for Chat-GPT's answers while it was 4.1 (SD: 0.56; Min: 3, Max: 5) for the group of EESS ($p < 0.015$). The overall Wilcoxon signed-rank test yielded a significance level of $p < 0.026$ when comparing the level of agreement between Chat-GPT and EESS. The mean number of therapeutic alternatives suggested by Chat-GPT was 3.33 (SD: 1.2; Min: 2, Max: 5), while it was 2.6 (SD: 0.51; Min: 2, Max: 3) for the group of EESS; $p = 0.286$ (95% CI − 0.385 to 1.320).

Conclusion Chat-GPT represents a promising tool in the clinical decision-making process within the salivary gland clinic, particularly for patients who are candidates for sialendoscopy treatment. Additionally, it serves as a valuable source of information for patients. However, further development is necessary to enhance the reliability of these tools and ensure their safety and optimal use in the clinical setting.

Keywords Salivary · Gland · Sialendoscopy · Chat-GPT · Chatbot

Introduction

The field of otolaryngology has witnessed remarkable advancements over the past few years, revolutionizing the diagnosis and treatment of various conditions affecting the head and neck region. Among these developments, sialendoscopy has emerged as a groundbreaking technique, offering a minimally invasive approach for exploring and managing salivary gland disorders [1]. With its ability to visualize the intricate salivary duct system, sialendoscopy has paved the way for precise and targeted interventions.[1]

Traditionally, the management of salivary gland diseases relied heavily on open surgical procedures, often associated with significant morbidity and prolonged recovery periods. However, the advent of sialendoscopy has revolutionized the diagnostic and therapeutic landscape, providing an alternative that is less invasive, more precise, and offers enhanced patient outcomes. By utilizing miniaturized endoscopic instruments, sialendoscopy enables direct visualization and manipulation of the salivary gland ductal system, allowing for targeted interventions while preserving glandular function.

The current key indications for sialendoscopy, include the management of obstructive sialadenitis, salivary stones (basket removal or laser lithotripsy), and strictures or stenosis. Furthermore sialendoscopy, can serve as a valuable adjunct tool in the diagnosing workup of some diseases, such as neoplastic lesions (guided biopsy) or autoimmune disorders (intraductal irrigation), that may

Extended author information available on the last page of the article

affect the salivary glands [2]. And the decision on the approach to be used will be usually defined based on sialoendoscopist's personal experience or previously published algorithms [3].

In recent years, the advent of chatbots, powered by advanced natural processing language (NPL) and artificial intelligence (AI) algorithms, has revolutionized the way healthcare professionals and patients access and analyze medical information and potentially will support soon the clinical decision-making process. Some key advantages of using chatbots are (a) rapid access to information; b) enhanced diagnostic accuracy (assisting healthcare professionals by analyzing symptoms, medical history, and other relevant data); (c) decision support and treatment recommendations; (d) continuous learning and updating; (e) improved workflow efficiency; (f) patient education and engagement; (g) scalability and accessibility (including smartphones, websites, and messaging applications). However, they should not replace human expertise and clinical judgment [4–6].

The aim of this study was to analyze the combination of emerging technologies like Chat-GPT 3.5 (Chat-Generative Pre-trained Transformer 3.5, developed by OpenAI, San Francisco, USA) in medical training. Specifically, we aim to enhance the capabilities of sialendoscopy and further improve the management of salivary gland disorders. Additionally, we aim to know the potential information provided to our patients about these pathologies when they seek it.

Material and methods

A prospective, cross-sectional study was designed to assess the level of agreement between Chat-GPT and 10 expert sialendoscopists (ES) in six potential clinical scenarios. The experts were selected based on their clinical and scientific background in the field.

The six clinical scenarios considered were defined by the main authors according to the most common cases observed in three different sialendoscopy clinics:

(a) What is the best option for treating a lithiasis larger than 1 cm in the submandibular salivary gland hilum?
(b) What is the best option for treating a lithiasis of 3 mm in the submandibular salivary gland duct?
(c) What is the best approach for treating a proximal parotid gland lithiasis larger than 1 cm?
(d) What is the best approach for treating a proximal parotid gland lithiasis of 3 mm?
(e) What is the best option for treating radioiodine-induced sialadenitis?
(f) What is the best option for treating salivary duct stenosis?

Afterward, the questions were sent to the 10 ES via email as an e-questionnaire, which included all the potential treatment strategies (shared through the Google Forms platform). Simultaneously, each clinical case was presented to Chat-GPT on May the 10th [3], and potential treatment strategies were requested, with each answer being collected (Fig. 1).

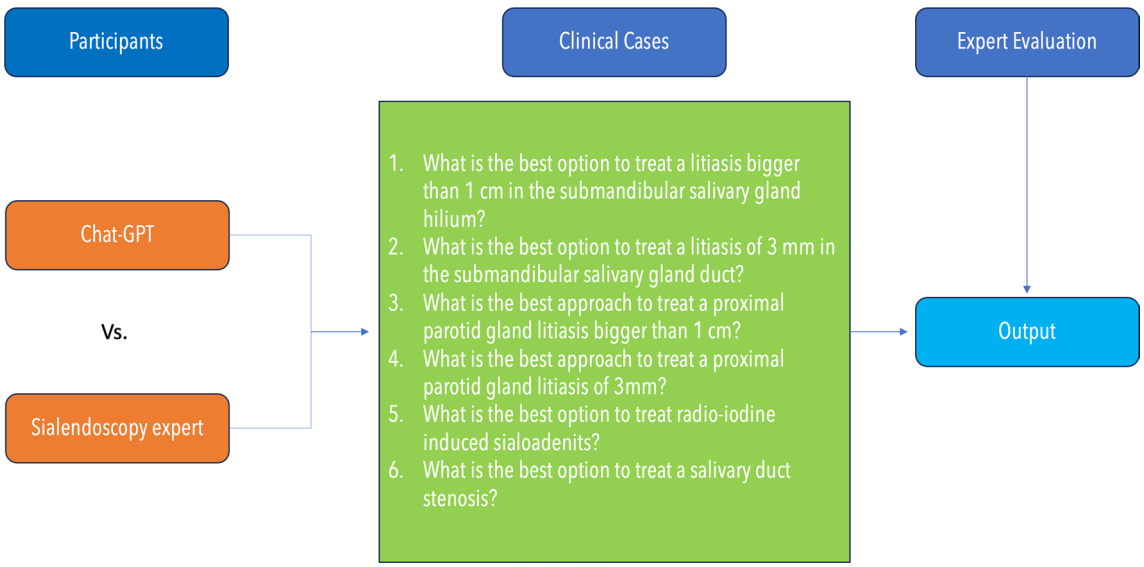

Fig. 1 Study design

To ensure consistency, all questions were entered into the Chat-GPT engine by one investigator.

Subsequently, another set of 10 ES reviewed and rated the overall level of agreement between Chat-GPT and ES clinical decisions, as well as the level of agreement for each question, using a Likert-Scale method ranging from 1 to 5 (1 = Strongly disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, and 5 = Strongly agree). This study did not need ethical approval as no patient-level data were used.

Statistical analysis

Quantitative and continuous variables were expressed as the mean ± standard deviation (SD). Correlation was evaluated using the Wilcoxon signed-rank test to compare the results regarding the level of agreement. Two-tailed *t*-tests were used to compare the mean number of responses between ES and Chat-GPT. The significance threshold used was $p < 0.05$. All statistical analyses were performed using JASP (Version 0.11.1. University of Amsterdam, Netherlands) (https://jasp-stats.org/).

Results

The answers provided by the 10 EESS and Chat-GPT for the six clinical scenarios, each with at least two potential treatment alternatives, were evaluated. The mean level of agreement, as assessed by the evaluators using the Likert scale for Chat-GPT's answers, was 3.4 (SD: 0.69; Min: 2, Max: 4), while it was 4.1 (SD: 0.56; Min: 3, Max: 5) for the group of EESS ($p < 0.015$). The overall Wilcoxon signed-rank test yielded a significance level of $p < 0.026$ when comparing the level of agreement between Chat-GPT and EESS despite sometimes the order of options were not the same, but still consider a gland sparing approach (Fig. 2 and Table 1).

The mean number of therapeutic alternatives suggested by Chat-GPT was 3.33 (SD: 1.2; Min: 2, Max: 5), while it was 2.6 (SD: 0.51; Min: 2, Max: 3) for the group of EESS; $p = 0.286$ (95% CI − 0.385 to 1.320) (Fig. 3).

Discussion

In recent decades, the development of minimally invasive surgical techniques for managing salivary gland pathologies has provided surgeons with the ability to employ conservative and gland-sparing approaches instead of major surgery. However, clinical, and surgical training remains limited globally, highlighting the need for strategies to enhance decision-making support for novice clinicians and increase the information accessible to patients.

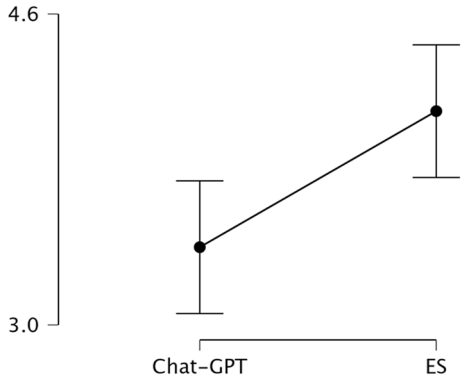

Fig. 2 Descriptive plot of mean level of agreement. *ES* expert sialendoscopist

To the best of our knowledge, this is the first study designed to compare the level of agreement between expert sialendoscopists and Chat-GPT in common clinical scenarios within the salivary gland clinic. Our results indicate that EESS treatment suggestions were rated higher than Chat-GPT suggestions in half of the clinical scenarios, while they were equal in the other half. Regarding the potential treatment options, Chat-GPT suggested a higher number of options without a statistically significant difference. Overall, the information provided by Chat-GPT was considered comprehensive and accurate.

Chat-GPT is considered the most popular chatbot and is built upon a natural language processing (NLP) architecture. It utilizes an autoregressive large language model (LLM) trained on vast amounts of human-generated text from sources such as the Internet. The chatbot generates answers by predicting the next most likely "token" or word/phrase to complete the ongoing response. However, it is important to note that the AI has limitations when it comes to reasoning capabilities [4]. As highlighted by Rao et al., these limitations become evident in the frequent instances where Chat-GPT provides futile recommendations or fails to provide a diagnosis despite having all the necessary information [7].

Due to the widespread availability of chatbots, there has been a growing number of preliminary studies exploring the use of Chat-GPT in clinical decision-making [8–11]. These studies, are motivated by the potential adoption of AI-based chatbots as patient counseling tools and aids in clinical decision-making. As demonstrated by Ayers et al., chatbots can assist clinicians by drafting messages based on patients' queries, providing support to physicians or support staff in messaging with patients [12]. The availability of these medical dialogue systems also addresses workflow imbalances caused by a limited number of experienced healthcare professionals in certain fields [13]. This results in time savings for more complex tasks, consistent responses, and improved communication skills for staff. Additionally, it

Table 1 Comparison among clinical suggestions between Chat-GPT and Expert Sialendoscopist

Case	Question	Chat-GPT	Surgeon	Overall Level of Agreement
Case 1	*1. What is the best option to treat a lithiasis bigger than 1 cm in the submandibular salivary gland hilium?*	1. Sialoadenectomy 2. Transoral duct stone removal	1. Transoral duct stone removal 2. Sialoadenectomy 3. EWSL	Neither agree or disagree = 3
Case 2	*2. What is the best option to treat a lithiasis of 3 mm in the submandibular salivary gland duct?*	1. Conservative management 2. Sialendoscopy approach (Plus basket or laser) 3. ESWL[a] 4. Sialoadenectomy	1. Sialendoscopy approach (Plus basket or laser) 2. Conservative Management	Agree = 4
Case 3	*3. What is the best approach to treat a proximal parotid gland lithiasis bigger than 1 cm?*	1. Parotidectomy 2. Transoral duct stone removal	1. Combined transfacial approach 2. Sialendoscopy approach (Plus basket or laser) 3. ESWL	Neither agree or disagree = 3
Case 4	*4. What is the best approach to treat a proximal parotid gland lithiasis of 3 mm?*	1. Sialendoscopy approach (Plus basket or laser) 2. ESWL* 3. Transfacial combined approach	1. Sialendoscopy approach (Plus basket or laser) 2. Conservative treatment (Hidratation, NSAID, Gland Massage)	Neither agree or disagree = 3
Case 5	*5. What is the best option to treat radioiodine induced sialoadenits?*	1. Symptomatic relief 2. NSAIDs 3. Saliva stimulation (Lemon, candies) 4. Sialogogues 5. Sialendoscopy	1. Symptomatic relief (Manage of pain, swelling, and dry mouth) 2. Sialendoscopy 3. Saliva Stimulation & Sialogogues	Agree = 4
Case 6	*6. What is the best option to treat a salivary duct stenosis?*	1. Conservative management. (Non-surgical management: In cases of mild or moderate salivary duct stenosis) 2. Dilation and/or stenting 3. Sialendoscopy 4. Sialoadenectomy	1. Sialendoscopy + Dilatation + Stenting 2. Sialendoscopy + dilation 3. Non-surgical management: (In cases of mild or moderate salivary duct stenosis)	Agree = 4

ESWL extracorporeal shock wave lithotripsy, [b]NSAIDs nonsteroidal anti-inflammatory drugs

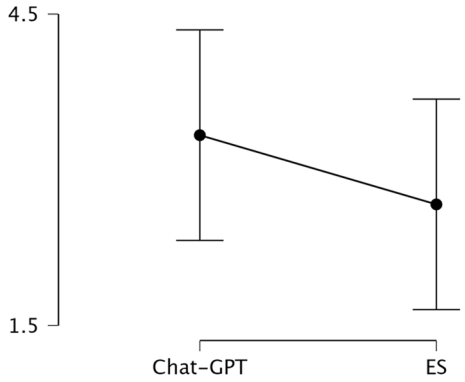

Fig. 3 Descriptive plot of mean number of therapeutic alternatives. *ES* expert sialendoscopist

leads to improved median response time, standards of care, empathy, patient satisfaction, outcomes, and reduced unnecessary clinical visits [14, 15].

In this study, the use of Chat-GPT was limited to specific parts of the clinical workflow and did not provide longitudinal patient or clinician support, which is a notable limitation. Additionally, the lack of patient participation can be considered another limitation. It is important to exercise caution despite the impressive performance of Chat-GPT because small errors in clinical judgment can potentially lead to adverse outcomes or misinformation for patients. Furthermore, it is crucial to validate the results of this study through replication in order to ensure their reliability and generalizability. Finally, we need to consider the cons highlighted by Biswas of using Chat-GPT in the clinical scenario: (a) limited accuracy, (b) bias and limitations of data; (c) lack of clinical context; (d) limited engagement; (e) not direct interaction with health professionals [16].

Conclusion

Chat-GPT represents a promising tool in the clinical decision-making process within the salivary gland clinic, particularly for patients who are candidates for sialendoscopy treatment. Additionally, it serves as a valuable source of information for patients. However, further development is necessary to enhance the reliability of these tools and ensure their safety and optimal use in the clinical setting.

Availability of data and materials The data is available.

Declaration

The author Jerome R. Lechien is also guest editor of the special issue on 'ChatGPT and Artificial Intelligence in Otolaryngology-Head and Neck Surgery'. He was not involved with the peer review process of this article.

References

1. Roland LT, Skillington SA, Ogden MA (2017) Sialendoscopy-assisted transfacial removal of parotid sialoliths: a systematic review and meta-analysis. Laryngoscope 127:2510–2516
2. Desmots F, Chossegros C, Salles F, Gallucci A, Moulin G, Varoquaux A (2014) Lithotripsy for salivary stones with prospective US assessment on our first 25 consecutive patients. J Craniomaxillofac Surg 42:577–582
3. Foletti JM, Graillon N, Avignon S, Guyot L, Chossegros C (2018) Salivary Calculi Removal by minimally invasive techniques: a decision tree based on the diameter of the calculi and their position in the excretory duct. J Oral Maxillofac Surg 76:112–118
4. ChatGPT: Optimizing Language Models for Dialogue (2023) OpenAI. Published November 30, 2022. https://openai.com/blog/chatgpt/. Accessed 10 May 2023
5. Parviainen J, Rantala J (2022) Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care. Med Health Care Philos 25:61–71
6. Javaid M, Haleem A, Pratap SR (2023) ChatGPT for healthcare services: an emerging stage for an innovative perspective. BenchCouncil Trans Benchmarks Stand Eval 3(1):100105
7. Rao A, Pang M, Kim J, Kamineni M, Lie W, Prasad AK, Landman A, Dreyer KJ, Succi MD (2023) Assessing the utility of ChatGPT throughout the entire clinical workflow. medRxiv [Preprint] (**2023 Feb 26:2023.02.21.23285886**)
8. Bates DW, Levine D, Syrowatka A et al (2021) The potential of artificial intelligence to improve patient safety: a scoping review. Npj Digit Med 4(1):1–8. https://doi.org/10.1038/s41746-021-00423-6
9. Rao A, Kim J, Kamineni M, Pang M, Lie W, Succi MD (2013) Evaluating ChatGPT as an adjunct for radiologic decision-making. https://doi.org/10.1101/2023.02.02.23285399
10. Levine DM, Tuwani R, Kompa B, et al (2013) The diagnostic and triage accuracy of the GPT-3 artificial intelligence model. https://doi.org/10.1101/2023.01.30.23285067
11. Johnson D, Goodman R, Patrinely J, Stone C, Zimmerman E, Donald R, et al (2023) Assessing the accuracy and reliability of AI-Generated medical responses: an evaluation of the Chat-GPT Model. Res Sq [Preprint] (**2023 Feb 28:rs.3.rs-2566942.10**)
12. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB et al (2023) Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 183:589–596
13. Tripathy S, Singh R, Ray M (2023) Natural language processing for COVID-19 consulting system. Proc Comput Sci 218:1335–1341
14. Rasu RS, Bawa WA, Suminski R, Snella K, Warady B (2015) Health literacy impact on national healthcare utilization and expenditure. Int J Health Policy Manag 4:747–755. https://doi.org/10.15171/ijhpm.2015.151
15. Rotenstein LS, Holmgren AJ, Healey MJ et al (2022) Association between electronic health record time and quality of care metrics in primary care. JAMA Netw Open 5(10):e2237086
16. Biswas SS (2023) Role of Chat GPT in Public Health. Ann Biomed Eng 51:868–869

Authors and Affiliations

Carlos M. Chiesa-Estomba[1,2,6,7,12,13] · Jerome R. Lechien[3,6,7,12] · Luigi A. Vaira[4,5,6,7,12] · Aina Brunet[6,7] · Giovanni Cammaroto[6,7,8,12] · Miguel Mayo-Yanez[6,7,9,12] · Alvaro Sanchez-Barrueco[6,7,10,11] · Carlos Saga-Gutierrez[1,2,6,7]

✉ Carlos M. Chiesa-Estomba
chiesaestomba86@gmail.com

1 Department of Otorhinolaryngology, Donostia University Hospital, Biodonostia Research Institute, Osakidetza, 20014 San Sebastian, Spain

2 Otorhinolaryngology Department, Faculty of Medicine, Deusto University, Bilbo, Spain

3 Division of Laryngology and Broncho-Esophagology, Department of Otolaryngology and Head and Neck Surgery, EpiCURA Hospital, UMONS Research Institute for Health Sciences and Technology, University of Mons, Mons, Belgium

4 Maxillofacial Surgery Operative Unit, Department of Medicine, Surgery and Pharmacy, University of Sassari, Sassari, Italy

5 Biomedical Sciences Department, School of Biomedical Sciences, University of Sassari, Sassari, Italy

6 Department of Otorhinolaryngology, Head and Neck Surgery, Hospital Universitari Bellvitge, Barcelona, Spain

7 Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), Barcelona, Spain

8 Department of Head-Neck Surgery, Otolaryngology, Head-Neck and Oral Surgery Unit, Morgagni Pierantoni Hospital, 47121 Forlì, Italy

9 Otorhinolaryngology, Head and Neck Surgery Department, Complexo Hospitalario Universitario A Coruña (CHUAC), 15006 A Coruña, Galicia, Spain

10 ENT and Cervicofacial Surgery Department, Hospital Universitario Fundación Jiménez Díaz, Madrid, Spain

11 Otorhinolaryngology Department, Faculty of Medicine, Universidad Alfonso X el Sabio, Madrid, Spain

12 Head & Neck Study Group of Young-Otolaryngologists of the International Federations of Oto-Rhino-Laryngological Societies (YO-IFOS), Paris, France

13 Young Confederation of European Otorhinolaryngology, Head and Neck Surgery, Vienna, Austria