ChatGPT Performance in Laryngology and Head & Neck Surgery: A Clinical Case-Series.

Objective: To study the performance of ChatGPT in the management of laryngology and head and neck (LHN) cases.

Methods: History and clinical examination of patients consulting at the Otolaryngology-Head & Neck Surgery department were presented to ChatGPT, which was interrogated for differential diagnosis, management, and treatment. The ChatGPT performance was assessed by two blinded board-certified otolaryngologists using the following items of a composite score and the Ottawa Clinic Assessment Tool: differential diagnosis; additional examination; and treatment options. The complexity of clinical cases was evaluated with the Amsterdam Clinical Challenge Scale test.

Results: Forty clinical cases were submitted to ChatGPT, accounting for 14 (35%), 12 (30%), and 14 (35%) easy, moderate and difficult cases, respectively. ChatGPT indicated a significant higher number of additional examinations compared to practitioners (p=0.001). There was a significant agreement between practitioners and ChatGPT for the indication of some common examinations (audiometry, ultrasonography, biopsy, gastrointestinal endoscopy or videofluoroscopy). ChatGPT never indicated some important additional examinations (PET-CT, voice quality assessment, or impedance-pH monitoring). ChatGPT reported highest performance in the proposition of the primary (90%) or the most plausible differential diagnoses (65%), and the therapeutic options (60-68%). The ChatGPT performance in the indication of additional examinations was lowest.

Conclusion: ChatGPT is a promising adjunctive tool in LHN practice, providing extensive documentation about disease-related additional examinations, differential diagnoses, and treatments. The ChatGPT is more efficient in diagnosis and treatment, rather than in the selection of the most adequate additional examination.

Introduction:

A chatbot is currently defined as an electronic system that simulates conversations by responding to keywords or phrases. Chatbots are commonly used in various marketing platforms, websites or messaging [1]. The Chatbot Generative Pre-trained Transformer (ChatGPT) was launched November 20, 2022 by OpenAI (Open AI, San Francisco, USA) to use algorithms to respond to simple-to-complicated questions [2]. Some reports have showed that ChatGPT is able to succeed law, business, or medical school exams [3], and should be useful to help the practitioner in clinical practice, research or administrative tasks [4,5]. Because its large access to knowledge database, ChatGPT should be used as an adjunctive clinical tool for practitioners, helping in the establishment of differential diagnoses in clinical situations, and, therefore, the prescription of additional examinations and treatments. To date, there is no publication assessing the ChatGPT performance in the management of clinical cases in laryngology and head and neck surgery.

The objective of this study was to investigate the performance of ChatGPT in the management of laryngology and head and neck surgery cases.

Methods:

Setting

Forty-eight patients consulting in the Laryngology-Head & Neck Surgery division of CHU Saint-Pierre (Department of Otolaryngology, Head Neck Surgery, Brussels, Belgium) and the Dour Medical Center (Dour, Belgium) were recruited in July 2023. The data of patients with complete information regarding history, complaints, comorbidities, medication, clinical and potential additional examinations were presented to ChatGPT (Open AI, San Francisco, USA). ChatGPT was systematically interrogated for differential diagnoses (What are your differential diagnoses?), additional examinations (What are your additional examinations to find the

diagnosis?) and potential therapeutic regimen (What are your treatment(s) for the primary diagnosis?). The responses/propositions of ChatGPT were collected in a database by an independent researcher (B.G.) and compared with the responses/propositions of two senior laryngologists (J.R.L., S.H.) who were blinded regarding the differential diagnosis, check-up and potential therapeutic management of ChatGPT. The two senior laryngologists carefully reassessed the management of all patients prior to submit the data to ChatGPT. Note that all patients benefited from a complete ear, nose, and throat examination, including tympanoscopy, nasofibroscopy (with stroboscopy in cases of dysphonia), oral examination and neck palpation. The study was approved by the institutional review board (reference: XX). Patient consented to participate.

Level of difficulty of the case

A clinical case commonly consists of 4 basic elements: the medical history; the clinical examination; the technical diagnostic findings (additional examinations); and the treatment, including the psychosocial context of the patient [6,7]. The findings of these characteristics may lead to a variation of the degree of complexity, while the clinical case complexity may increase as it contains distracting information. Based on these findings, some scoring systems were developed and available in the literature. In the present study, a modified version of the General Items off the Amsterdam Clinical Challenge Scale test (ACCS) [6] was used to rate the complexity of clinical cases submitted to ChatGPT. The ACCS consisted of six generic items that play a critical part in the challenge posed by an individual consultation: previous history/actual context; problem presented; communication (patient complaints and responses); physical examination (typical *versus* atypical signs); patient management (adequate versus complicated management); and prevention. Each item was evaluated on a 5-point scale, ranging from 1 (easy) to 5 (difficult). According to Gercama *et al*. [6], the extremes of each item were

defined in general terms. For example, the item "problem presented" is scored 1 when the problem is straightforward, not likely to be serious and of a limited nature, and 5 when it is vague, difficult to define, probably serious or complex [6]. For example, patients with atypical presentation of disease or poor therapeutic response to an evidence-based treatment may be assessed as 5/5 in examination and management. Each clinical case has a score ranging from 6 to 30. Scores ranging from 6 to 14, 15 to 23, and 24 to 30 were considered as easy, moderate, and difficult, respectively. At the end of the consultations, two practitioners evaluated the ACCS of each patient case (Figure 1) [6].

ChatGPT performance

The data collected throughout the consultation were presented to ChatGPT. The performance of ChatGPT in the management of clinical cases was scored with items of the Ottawa Clinic Assessment Tool (OCAT) [8], which is a clinical instrument used to evaluate the performance of resident or fellow-in-training. Precisely, two board certified otolaryngologists head and neck surgeons (J.R.L., C.C.) evaluated the accuracy of the responses of ChatGPT with following OCAT outcomes: differential diagnosis; management plan (additional examinations); and therapeutic approaches. For each item, the practitioner used a 5-point Likert scale ranging from 1 (unprepared to do, inappropriate management) to 5 (can be independent, adequate management) (Figure 1). In addition to the OCAT items, practitioners were invited to judge the performance of ChatGPT (yes or no) in the establishment of the most plausible primary diagnosis; the quality of differential diagnoses; the usefulness of proposed additional examinations regarding the clinical case; the necessity of additional examinations; the consideration of all important additional examinations for the clinical situation; the adequacy of therapeutic options, and the consideration of all indispensable therapeutic options (Figure 1). The judgement of ChatGPT management findings by the two otolaryngologists was based on

current clinical guidelines available on websites of scientific societies, including the Confederation of European Otorhinolarynological-Head Neck Surgery/European Laryngological Society, American Academy of Otolaryngology Head Neck Surgery, the International Federation of Otorhinolaryngological Societies, the French Society of Otorhinolaryngology (SFORL), and the Spanish Otorhinolaryngological Society.

Statistical analyses

Statistical analyses were performed with the Statistical Package for the Social Sciences for Windows (SPSS version 22,0; IBM Corp, Armonk, NY, USA). The number of additional examinations prescribed by patients were compared between practitioners and ChatGPT with Mann-Whitney U Test. Additional examinations proposed by practitioners and ChatGPT were coded with a predefined number in a matrix, allowing the evaluation of consistency between findings of the physicians *versus* ChatGPT (*kappa* analysis). The results of the competence scores of ChatGPT provided by the two experienced otolaryngologists were compared with a consistency analysis (Kendall tau). The consistency was considered as low, moderate and strong for k<0.40, 0.40-0.60, and k>0.60, respectively. A level of significance of p<0.05 was used.

Results:

Subjects and Setting

Of the 48 initial recruited patients, the clinical history of 40 patients was completed and presented to ChatGPT. There were 25 females (62.5%) and 15 males (37.5%), respectively. The mean age of patients was 47.7 ± 15.9 years. The clinical cases were considered as easy, moderate or difficult in 14 (35%), 12 (30%), and 14 (35%) patients, respectively. The mean ACCS was 17.5 ± 6.4. The main ear, nose and throat symptoms of patients are reported in Table 1. Dysphonia (N=17, 43%), globus sensation (N=15, 38%), throat sticky mucus/postnasal drip

(N=11, 28%), dysphagia (N=10, 25%), and throat clearing (N=10, 25%) were the most prevalent symptoms. The most common primary or secondary diagnoses of patients included laryngopharyngeal reflux (LPR) (N=20, 50%), vocal fold benign lesions (N=6, 15%), unilateral or bilateral vocal cord paralysis (N=2, 5%), chronic otitis media (N=2, 5%), and Eustachian tube dysfunction (N=2, 5%) (Appendices 1, 2 and 3).

Additional Examination Performance

A total of 74 additional examinations were indicated by the senior otolaryngologists (mean per patient=1.78 ± 1.00), while ChatGPT proposed 108 additional examinations (mean per patient= 2.78 ± 1.3; p=0.001; Mann-Whitney U test). The imaging examinations and bacteriology/cultures are indicated by ChatGPT 2.56 and 8.0 times as much as the practitioners (Table 2). Considering the level of difficulty of cases (ACCS), the mean numbers of additional examinations indicated by the practitioners were 1.71 ± 1.07, 1.75 ± 0.87 and 1.86 ± 1.10 for ACCS easy, moderate or difficult cases, respectively (p=0.378). The mean numbers of additional examinations proposed by ChatGPT were 3.29 ± 1.13, 2.92 ± 0.79, and 2.14 ± 1.51 for clinical cases judged as easy, moderate and difficult, respectively (p=0.010). Additional Pearson analysis reported a negative significant association between the degree of difficulty of clinical cases and the number of additional examinations proposed by ChatGPT (r_s=-0.373, p=0.018).

The consistency analysis data between practitioner and ChatGPT in the indication of additional examinations are described in Table 2. Table 2 includes additional examinations, which were prescribed at least once by practitioners and ChatGPT. Indeed, some additional examinations were prescribed only by ChatGTP (i.e. sinus X-ray (N=3), sinus MRI (N=1), pH study (N=8), thyroid check-up (N=1), polysomnography (N=2), lung volume evaluation (N=3) and electromyography (N=5). In the same way, practitioners prescribed additional examinations

that were never indicated by ChatGPT (i.e. impedance-pH testing (N=9), psychophysical olfactory evaluations (N=1), voice quality assessment (N=16), and Positron Emission tomography and computed tomodensimetry (Pet-CT) (N=3). The consistency analysis of these additional examinations, which were never prescribed by ChatGPT or practitioners, was consequently not significant.

There were significant consistencies between practitioners and ChatGPT for some examinations, such as head and neck ultrasonography, audiometry, tympanometry, ear tomodensitometry, gastrointestinal endoscopy, lesion biopsy or cytology, videofluoroscopy swallowing study, and bacteriology. The indications of audiometry (k=0.643), ear CT (k=0.999) and cytology (k=0.655) were the only additional examinations reporting strong consistency. However, some additional examinations were never considered by ChatGPT, including the voice quality assessment (subjective and objective voice evaluations), impedance-pH monitoring, psychophysical olfactory evaluation, and the Pet-CT, which support the low kappa coefficient value of some items. Note that in the case number 19, ChatGPT proposed head and neck MRI in a patient with a history of pacemaker.

Diagnosis and Treatment Performance

The mean ChatGPT performance scores of additional examination, differential diagnosis, and treatment items are summarized in Table 3. The interrater consistency of both otolaryngologist judges was significant in the assessment of primary diagnosis, useful additional examinations, necessary additional examinations, complete additional examination options, and OCAT scores for diagnosis, management, and treatment (Table 3). The performance score of ChatGPT did not significantly differ in additional examination, differential diagnosis and treatment scores of the OCAT items. According to judges, ChatGPT reported highest performance in the proposition of plausible differential diagnoses (90%), the proposition of the most plausible

diagnosis (65%), and the proposition of a series of adequate therapeutic options (60-68%). In 67% to 90% of cases, judges reported that some additional examinations indicated by ChatGPT are not necessary, while ChatGPT forgot some indispensable additional examinations in 55% to 75% of cases (Table 3). The performance of ChatGPT did not significantly differ according to level of difficulty of clinical cases.

Discussion:

The development of artificial intelligence and chatbot dedicated to healthcare is an important ongoing topic in medicine. The development of ChatGPT was so fast that a collective of more than 1,000 artificial intelligence experts, researchers and backers have joined a call for an immediate pause of at least 6 months on the creation of giant Ais, such as GPT-4, to study and mitigate the capabilities and dangers of these systems [9]. Indeed, to date, the reliability of ChatGPT was not extensively investigated, while it is full available for patients and physicians. In the field of otolaryngology, there are currently less than 10 studies exploring the potential of ChatGPT and no study has investigated its performance and reliability in clinical practice. To the best of our knowledge, this study is the first investigation of the performance of ChatGPT in clinical practice when some real clinical cases are presented to the chatbot.

The primary observation of the present study was the ability of ChatGPT to propose additional examinations to explore potential differential diagnoses related to the presentation a list of symptoms and findings associated with an otolaryngological condition. ChatGPT was found to indicate a significant higher number of additional examinations compared to practitioners but cannot discern the superiority of one over the other (for example, the superiority of sinus CT over sinus X-ray). In Evidence-Based Medicine, the choice of additional examinations needs to consider the patient risk (radiation of some X-ray or CT), the clinical history of patients (pacemaker and MRI), the information obtained through the examination, and the cost of the

examination for the healthcare system [10,11]. The findings of the present study support that ChatGPT is not ready to select the best additional examination considering all these important outcomes and the patient history. The MRI proposition of ChatGPT in a patient with a pacemaker supports this conclusion as well as the systematic proposition of sinus X-ray for diagnosing chronic rhinosinusitis because X-ray is not evidence-based according to guidelines [12,13]. However, according to our consistency analysis, it appears that some ChatGPT propositions significantly matched with the indications of practitioners. The significant strong consistencies concern some common/first-line examinations, which are all frequently prescribed in practice (e.g. head and neck ultrasonography, audiometry, and cytology) for prevalent and easily-diagnosed conditions. Interestingly, ChatGPT never proposed some more specialized additional examinations, including voice quality assessment, psychophysical olfactory evaluations, or impedance pH-monitoring, which are recommended in the management of dysphonia [14,15], chronic olfactory dysfunction [16], and recurrent or chronic LPR [17].

The selection of adequate additional examinations is an important step to confirm the primary diagnosis of a clinical condition and to propose treatment(s). According to two blinded board-certified otolaryngologists, the highest ChatGPT performance concerns the proposition of primary (90%), differential diagnoses (65%), and therapeutic options (60-68%), whereas ChatGPT performance appears lowest in the indication of additional examinations. Judges observed that the performance of ChatGPT may substantially vary regarding the current knowledges of the disease. For example, the ChatGPT management of LPR was not adequate. ChatGPT does not differentiate LPR from gastroesophageal reflux disease (GERD) and systematically proposed pH-study, gastrointestinal endoscopy, and proton pump inhibitors (PPI) or antiH2 blockers in all LPR patients, including those with confirmed nonacid reflux at the pH-impedance monitoring and related resistance to PPI. As for the additional examinations,

ChatGPT functions as an encyclopedia and proposes a list of theoretical treatments that cannot consider the personal history of patient. The theoretical performance of ChatGPT in otolaryngology was recently supported in two original studies. Chiesa-Estomba *et al*. observed a significant agreement between ChatGPT and a college of experts in the theoretical clinical decision-making process within the salivary gland clinic [18]. Hoch *et al*. investigated the accuracy of ChatGPT to 2,576 theoretical questions covering 15 otolaryngology subspecialties [19]. ChatGPT correctly answered to 57% of questions, especially in single-choice questions, which may be commonly judged as easier than multiple questions in which distracting information may increase the level of difficulty [7].

The consideration of some medical history information as important or not is a key human aptitude for the determination of the most plausible primary diagnosis [7]. As it was difficult to objectify this point, the two otolaryngologist judges observed that ChatGPT tried to consider some distracting information (e.g. comorbidities without relation with the otolaryngological condition) in the establishment of diagnosis and treatment. In general, there are two sources of mistake: the lack of knowledge and lack of information processing [7]. In particular the degree of complexity leads to errors due to incorrect information processing or distracting information. This is especially important if in one case several differential diagnoses are conceivable, or the patient has several clinical problems that need to be weighted correctly. To what extent the degree of complexity of a clinical case contributes to the difficulty is unclear [7].

The findings of the present study are important for patients and young practitioners in otolaryngology and head and neck surgery. Indeed, according to the mediatization of ChatGPT, there is an increasing number of patients who consult ChatGPT prior to a physician [20,21]. The access of some theoretical information, such as the list of potential additional examinations or differential diagnoses, may complicate the task of the practitioner when indicating such additional examinations or treatment. In the same vein, young practitioners, e.g. medical

students, residents or fellow-in-training, need to consider the ChatGPT information with precautions, keeping in mind that the human discernment is not yet acquired by chatbot systems. The main strengths of the present study are the originality and the consideration of real clinical cases, which makes this study the first investigation assessing the reliability and performance of ChatGPT in otolaryngology head and neck surgery practice. Because the development of intelligent chatbots is very recent, there is no validated clinical tool for the assessment of IA effectiveness in clinical practice. Thus, we used two modified clinical tools to characterize the difficulty of clinical cases (ACCS), and to judge the performance of ChatGPT (OCAT). The use of items of these validated tools is the main limitation of the present study because they were not developed for chatbot evaluations. Future studies are needed to develop clinical instruments dedicated to the evaluation of the performance of chatbots.

Conclusion:

ChatGPT is a promising adjunctive tool in laryngology and head and neck surgery practice, providing extensive documentation about disease-related additional examinations, differential diagnoses, and treatment. The ChatGPT is more efficient in diagnosis and treatment, rather than in the selection of the most adequate additional examination. Future clinical studies are needed to assess the performance of ChatGPT in other otolaryngology subspecialties.

References:

1.Pernencar C, Saboia I, Dias JC. How Far Can Conversational Agents Contribute to IBD Patient Health Care-A Review of the Literature. *Front Public Health*. 2022; 10:862432. doi: 10.3389/fpubh.2022.862432.

2.Hill-Yardin EL, Hutchinson MR, Laycock R, Spencer SJ. A Chat(GPT) about the future of scientific publishing. *Brain Behav Immun*. 2023; 110:152-154. doi: 10.1016/j.bbi.2023.02.022.

3. Choi JH, Hickman KE, Monahan A, Schwarcz D. ChatGPT Goes to Law School ? *Minnesota Legal Studies Research Paper No. 23-03* ; 2023.

4. Gupta R, Park JB, Bisht C, Herzog I, Weisberger J, Chao J, Chaiyasate K, Lee ES. Expanding Cosmetic Plastic Surgery Research Using ChatGPT. *Aesthet Surg J*. 2023: sjad069. doi: 10.1093/asj/sjad069.

5. Salvagno M, Taccone FS, Gerli AG. Can artificial intelligence help for scientific writing? *Crit Care*. 2023; 27(1):75. doi: 10.1186/s13054-023-04380-2.

6. Gercama AJ, de Haan M, van der Vleuten CPM. Reliability of the Amsterdam Clinical Challenge Scale (ACCS): a new instrument to assess the level of difficulty of patient cases in medical education. *Med Educ. 2000;* 34(7):519–524.

7. Braun LT, Lenzer B, Fischer MR, Schmidmaier R. Complexity of clinical cases in simulated learning environments: proposalfor

a scoring system. *GMS J Med Educ*. 2019; 36(6):Doc80. doi: 10.3205/zma001288.

8. Rekman J, Hamstra SJ, Dudek N, Wood T, Seabrook C, Gofton W. A New Instrument for Assessing Resident Competence in Surgical Clinic:

The Ottawa Clinic Assessment Tool. *J Surg Educ*. 2016; 73(4):575-82. doi: 10.1016/j.jsurg.2016.02.003.

9. https://futureoflife.org/open-letter/pause-giant-ai-experiments/

10. Fine B, Dhanoa D. Imaging appropriateness criteria: why Canadian family physicians should care. *Can Fam Physician*. 2014; 60(3):217-8.

11. Kang SK, Gold HT. How to Perform Economic Evaluation in Implementation Studies: Imaging-Specific Considerations and Comparison of Financial Models. *J Am Coll Radiol*. 2023; 20(3):292-298. doi: 10.1016/j.jacr.2022.11.018.

12. Hellings PW, Fokkens WJ, Orlandi R, Adriaensen GF, Alobid I, Baroody FM, et al. The EUFOREA pocket guide for chronic rhinosinusitis. *Rhinology*. 2023; 61(1):85-89. doi: 10.4193/Rhin22.344.

13. Yoshiyasu Y, Wu F, Dhanda AK, Gorelik D, Takashima M, Ahmed OG. GPT-4 accuracy and completeness against International Consensus Statement on Allergy and Rhinology: Rhinosinusitis. *Int Forum Allergy Rhinol*. 2023. doi: 10.1002/alr.23201.

14. Dejonckere PH, Bradley P, Clemente P, Cornut G, Crevier-Buchman L, Friedrich G, Van De Heyning P, Remacle M, Woisard V; Committee on Phoniatrics of the European Laryngological Society (ELS). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). *Eur Arch Otorhinolaryngol*. 2001; 258(2):77-82. doi: 10.1007/s004050000299.

15. Stachler RJ, Francis DO, Schwartz SR, Damask CC, Digoy GP, Krouse HJ, McCoy SJ, Ouellette DR, Patel RR, Reavis CCW, Smith LJ, Smith M, Strode SW, Woo P, Nnacheta LC. Clinical Practice Guideline: Hoarseness (Dysphonia) (Update). *Otolaryngol Head Neck Surg*. 2018; 158(1_suppl):S1-S42. doi: 10.1177/0194599817751030.

16. Stuck BA, Beule A, Damm M, Gudziol H, Hüttenbrink KB, Landis BN, Renner B, Sommer JU, Uecker FC, Vent J, Hummel T; Committee on Olfaction and Gustation of the German Society of Otorhinolaryngology, Head and Neck Surgery. Position paper "Chemosensory

testing for expert opinion in smell disorders". *Laryngorhinootologie*. 2014; 93(5):327-9. doi: 10.1055/s-0033-1364034.

17. Lechien JR, Vaezi MF, Chan WW, Allen J, Karkos PD, Saussez S, Altman K, Amin MR, Ayad T, Barillari MR et al. 2023. The Dubai Definition and Diagnostic Criteria of Laryngopharyngeal Reflux: The IFOS Consensus. Laryngoscope.

18. Chiesa-Estomba CM, Lechien JR, Vaira LA, Brunet A, Cammaroto G, Mayo-Yanez M, Sanchez-Barrueco A, Saga-Gutierrez C. Exploring the potential of Chat-GPT as a supportive tool for sialendoscopy clinical decision making and patient information support. *Eur Arch Otorhinolaryngol*. 2023. doi: 10.1007/s00405-023-08104-8

19. Hoch CC, Wollenberg B, Lüers JC, Knoedler S, Knoedler L, Frank K, Cotofana S, Alfertshofer M. ChatGPT's quiz skills in different otolaryngology subspecialties: an analysis of 2576 single-choice and multiple-choice board certification preparation questions. *Eur Arch Otorhinolaryngol*. 2023. doi: 10.1007/s00405-023-08051-4.

20. Millstein J, Agarwal A.What can doctors and patients do with ChatGPT? | Expert Opinion. Philadelphia Inquirer. 2023.

21. Chaix B, Bibault JE, Pienkowski A, Delamon G, Guillemassé A, Nectoux P, Brouard B. When Chatbots Meet Patients: One-Year Prospective Study of Conversations Between Patients With Breast Cancer and a Chatbot. *JMIR Cancer*. 2019; 5(1):e12856. doi: 10.2196/12856.

Table 1: Patient symptoms.

Main symptoms	Easy cases ACCS 6-14	Moderate cases ACCS 15-23	Difficult cases ACCS 24-30	Total
Dysphonia, roughness or voice fatigue	7 (50)	4 (33)	6 (43)	17 (43)
Globus sensation	5 (36)	7 (58)	3 (21)	15 (38)
Throat sticky mucus or postnasal drip	5 (36)	5 (42)	1 (7)	11 (28)
Dysphagia	4 (29)	3 (25)	3 (21)	10 (25)
Throat clearing	7 (50)	2 (17)	1 (7)	10 (25)
Throat pain or odynophagia	2 (14)	3 (25)	2 (14)	7 (18)
Cough	1 (7)	4 (33)	1 (7)	6 (15)
Cervical mass	2 (14)	0 (0)	2 (14)	4 (10)
Hearing loss	3 (21)	1 (8)	0 (0)	4 (10)
Anorexia or weight loss	1 (7)	0 (0)	2 (14)	3 (8)
Nasal obstruction	0 (0)	2 (17)	0 (0)	2 (5)
Tinnitus	1 (7)	1 (8)	0 (0)	2 (5)
Dyspnea	0 (0)	0 (0)	2 (14)	2 (5)
Dry mouth	0 (0)	1 (8)	0 (0)	1 (3)
Fever	1 (7)	0 (0)	0 (0)	1 (3)
Heartburn	0 (0)	1 (8)	0 (0)	1 (3)
Smell loss of parosmia	0 (0)	1 (8)	0 (0)	1 (3)
Dry eyes	0 (0)	1 (8)	0 (0)	1 (3)
Aspirations	0 (0)	0 (0)	1 (7)	1 (3)
Tonsil ulceration	0 (0)	0 (0)	1 (7)	1 (3)

Table 1 footnotes: Abbreviations: ACCS= Amsterdam Clinical Challenge Scale.

Table 2: Additional Examination Consistency.

Additional examinations	kappa	p-value
Head & Neck ultrasonography	0.481	0.001
Head & Neck CT	0.158	NS
Head & Neck MRI	0.091	NS
Sinus come beam/CT	0.231	NS
Biology (CRP, Leucocytes, etc.)	0.053	NS
Allergy skin or blood assessments	0.048	NS
Audiometry	0.643	0.001
Tympanometry	0.481	0.001
Ear CT	0.999	0.001
Gastrointestinal endoscopy	0.216	0.028
Esophageal manometry	0.039	NS
Biopsy	0.483	0.002
Cytology	0.655	0.001
Videofluoroscopy (swallowing)	0.362	0.019
Bacteriology (swab, secretions, etc.)	0.186	0.043

Table 2 footnotes: This table presents additional examinations commonly indicated by OTO and ChatGPT at least once time. Abbreviations: CRP=C-reactive protein; CT=computed tomography; MRI=magnetic resonance imaging; NS=non-significant; OTO=otolaryngologists.

Table 3: Performance Analysis of ChatGPT in Diagnostic and Treatment.

Outcomes	Otolaryngologist 1				Otolaryngologist 2				Consistency	
	Easy	Moderate	Difficult	Tot	Easy	Moderate	Difficult	Tot	Kendall	p-value
Composite score (N (%))										
1. The primary diagnosis was the most plausible (N (%))	10 (71)	7 (58)	9 (64)	26 (65)	10 (71)	7 (58)	9 (64)	26 (65)	0.890	0.001
2. The differential diagnoses were all plausible	14 (100)	11 (92)	11 (79)	36 (90)	14 (100)	11 (92)	11 (79)	36 (90)	0.167	NS
3. The additional examinations could be usefull in the situation	8 (57)	7 (58)	10 (71)	25 (63)	8 (57)	5 (42)	7 (50)	20 (50)	0.650	0.001
4. The additional examinations were all necessary	7 (50)	2 (17)	4 (29)	13 (33)	1 (7)	0 (0)	3 (21)	4 (10)	0.375	0.001
5. All additional examination option were presented	6 (43)	7 (58)	5 (36)	18 (45)	3 (21)	1 (8)	6 (43)	10 (25)	0.368	0.010
6. The therapeutic options are all adequate regarding the disease	8 (57)	8 (67)	11 (79)	27 (68)	11 (79)	5 (42)	8 (57)	24 (60)	0.409	0.009
7. All therapeutic indispensable options were presented	9 (64)	6 (50)	7 (50)	22 (55)	4 (29)	0 (0)	6 (43)	10 (25)	0.238	NS
Ottawa Clinic Assessment Tool outcomes (mean (SD))										
1. Differential diagnoses	4.14 (0.95)	3.75 (1.48)	3.71 (1.20)	3.88 (1.20)	3.86 (0.66)	3.42 (0.79)	3.57 (0.94)	3.63 (0.81)	0.365	0.010
2. Management plan	3.71 (1.20)	3.33 (0.78)	3.29 (0.83)	3.45 (0.96)	3.21 (0.98)	2.92 (0.67)	3.14 (0.66)	3.10 (0.78)	0.520	0.001
3. Therapeutic manament	3.79 (1.25)	3.50 (1.17)	3.57 (1.16)	3.63 (1.17)	3.71 (0.91)	3.33 (0.78)	3.21 (1.31)	3.43 (1.04)	0.664	0.001

Table 3 footnotes: The interrater reliability analysis was carried out with Kendall tau. Abbreviations: NS=non significant.

Figure 1: Chart flow.

Figure 1 footnotes: Abbreviations: ACCS=Amsterdam Clinical Challenge Scale test; OCAT:

Ottawa Clinic Assessment Tool: OTO=otolaryngologist.

Appendix 1: Clinical case features and ChatGPT results.

| | | | | | | Otolaryngologist consultation findings | | |
N	G	Age	Symptoms	History/medication	Clinical examination	Additional examinations	Diagnosis	Treatment
37	F	30	Recurrent throat pain, fever and lymphadenopathy, chronic dysphagia (5-y)	Tonsil abcess (2 times) treated with antibiotics	Grade III tonsils	-	Recurrent tonsil infections	Tonsillectomy
7	F	24	Globus, throat clearing, Abdominal pain, postnasal drip/sticky mucus (2-y)	None	Tongue tonsil hypertrophy, laryngo-pharyngeal inflammation	HEMII-pH testing *Negative allergy test*	LPR	Diet, stress reduction, PPI/alginate
8	F	40	Dysphonia, globus, throat pain (6-mo)	Suspected LPR	Vocal fold erythema Laryngeal inflammation	Voice quality assessment	Suspected LPR	Diet, stress reduction, PPI/alginate
9	F	53	Dysphonia, dysphagia, throat clearing, throat mucus (>1-y)	Ehlers Danlos	Coated/tongue, tonsil hypertrophy, laryngo-pharyngeal inflammation	Voice quality assessment	Suspected LPR	Diet, stress reduction, PPI/alginate
21	M	56	Right neck mass, weight loss (10 kg) dysphagia (6-mo)	Alcohol/tobacco overuses (30 years)	Right piriform sinus exophytic mass	Neck CT, PetCT, biopsy, biology & nutrition check-up	Hypopharyngeal primary carcinoma	Oncological board discussion
23	F	32	Dysphonia post-thyroidectomy (1 mo)	Thyroidectomy for goiter	Right vocal cord paralysis	Voice quality assessment	Vocal cord paralysis	Medialization, speech therapy
3	M	22	Left hearing loss, tinnitus, throat clearing, globus, cough (6-mo)	Recurrent LPR Recurrent otitis media	Bilateral ear retraction pocket, laryngo-pharyngeal inflammation	Audiometry, Tympanometry, naso-pharyngeal pH testing	Chronic otitis media, recurrent suspected LPR	Nasal saline irrigation, corticoids, diet, stress reduction, PPI/alginate
17	F	53	Chronic hoarseness, throat clearing, globus, sticky mucus (>4-y)	Tobacco overuse (30 PY)	Bilateral Reinke edema (grade III), laryngo-pharyngeal inflammation	Voice quality assessment	Reinke edema	Stop tobacco, In-office laser surgery, speech therapy
5	M	39	Recurrent throat clearing, postnasal drip, sticky mucus (>3-year)	Nasopharyngeal reflux (Restech)	Mulberry turbinate, & hypertrophy Laryngeal inflammation	*Normal sinus CT Nasopharyngeal Reflux*	Recurrent/ chronic LPR	Drug change: Magaldrate to alginate, continue diet and stress reduction.

#	Sex	Age	Symptoms					
1	F	33	Left cervical painful mass (3-mo)	Asthma	Submandibular mass	Neck US, MRI and biology	Salivary lithiasis	NSAID, pilocarpin, sialadenoscopy
10	F	24	Dysphonia, dysphagia, throat sticky mucus (>12-mo)	Tonsillectomy Vocal cord nodules	Vocal cord nodules, Laryngopharyngeal inflammation	Voice quality assessment	Vocal cord nodules Suspected chronic LPR	Diet, stress reduction, PPI/alginate, Speech therapy
20	F	32	Sudden dysphonia after crying (1-w)	Voice professional	Right vocal cord hemorrhage	Voice quality assessment	Vocal cord hemorrhage	In-office laser cauterization
14	M	53	Dysphonia, cough, sticky mucus, throat clearing (24-mo)	Septoplasty, Nonacid naso-pharyngeal reflux	Postnasal drip Laryngopharyngeal inflammation	*Nasopharyngeal pH testing: nonacid nasopharyngeal reflux*	LPR	Diet, stress reduction, alginate only
2	M	65	Hearing loss Throat clearing, globus (6-mo)	External ear stenosis, GERD	Bilateral total EED stenosis, laryngeal inflammation	Audiometry (bone) Ear CT	Bilateral EED stenosisn acute suspected LPR	Canaloplasty Diet, stress reduction, PPI/alginate
12	F	54	Dysphagia, globus, heartburn tinnitus (>15-mo)	Breast cancer, COPD, hypo-thyroidism	Inferior turbinate hypertrophy, laryngo-pharyngeal inflammation	Voice quality assessment, audiometry, Tympanometry	Eustachian tube dysfunction, suspected LPR	Diet, stress reduction, PPI/alginate
13	M	67	Cough, throat pain, postnasal drip, globus (7-mo)	Nonacid LPR (HEMII-pH)	Coated tongue, tonsil erythema, laryngeal inflammation	*HEMII-pH: nonacid LPR*	LPR	Diet, stress reduction, alginate only
15	F	62	Dry mouth, sticky mucus, cough, globus follow-up(>6-mo)	Recurrent suspected LPR Aspecific laryngitis	Sticky mucus, tongue tonsil edema Laryngeal inflammation	Biology: positive for Chlamydia Pneumonia	Resistant LPR to PPI, infectious laryngitis	Diet, stress reduction, alginate, antibiotics (clarithromycin)
16	M	27	Globus, dysphonia, sitcky mucus, left nasal obstruction, halitosis (>19-mo)	Hearth insufficiensy Ineffective PPI-therapy	Left septal deviation Laryngopharyngeal inflammation	*Normal sinus CT Nonacid naso-pharyngeal reflux*	Recurrent/ chronic nonacid LPR	Diet, stress reduction, alginate only
18	M	51	Dysphonia, suspicion of vocal fold paralysis, globus, throat clearing (6-mo)	Crohn, COVID-19 Suspected LPR	Left vocal fold polyp Laryngopharyngeal inflammation	Voice quality assessment	Left vocal fold polyp Suspected LPR	In-office laser polyp surgery, speech therapy, diet/stress, alginate
24	M	56	Recurrent laryngeal cancer	Alcohol/tobacco	Persistent carcinoma	PetCT and biopsy:	Laryngeal carcinoma	Salvage laryngectomy

21

			after primary chemoradiation (cT3 carcinoma)	overuses	5-mo after the treatment	resistant carcinoma	resistant to chemoradiation	
11	F	65	Hypoacusis, dysphonia, dysphagia, Sticky mucus (>9-mo)	Recurrent chronic otitis media	Adenoid hypertrophy, chronic otitis media, laryngeal inflammation	Audiometry, Tympanometry, voice quality assessment	Chronic otitis media, LPR, Eustachian tube dysfunction	Diet, stress reduction, PPI/alginate, nasal saline irrigation & corticoids
30	M	70	Bilateral odynophagia, otalgia (6-mo)	None	Bilateral stylo-hyoid calcified ligaments	Neck CT scan	Eagle syndrome	Transoral robotic styloidectomy
27	F	50	Chronic cough, negative pH testing, normal pulmonary examinations	None	Laryngopharyngeal hypersensitivity	*HEMII-pH testing:* negative	Laryngeal hypersensitivity	Amitryptilin, GABA pentin, or superior laryngeal nerve infiltration
6	M	75	Nasal Congestion, heartburn, dysphonia (>12-mo)	Nasopharyngeal reflux, (Restech)	Laryngopharyngeal hypersensitivity & inflammation.	*Normal sinus CT Nasopharyngeal reflux*	Nasopharyngeal reflux	Diet, stress reduction, PPI/alginate, nasal saline irrigation & corticoids.
39	F	38	Dysphonia, dysphagia, cough, globus, sticky mucus (4-y)	Thyroidectomy Diabetes, arthrosis	Normal vocal cord mobility, laryngeal inflammation	HEMII-pH testing Voice quality assessment	Suspected LPR	Diet, stress reduction, PPI/alginate
4	F	71	Sudden smell loss, globus, dry eyes, sticky mucus, throat clearing (7-mo)	COVID-19	Dry eyes, coated tongue, Laryngopharyngeal inflammation	Psychophysical evaluations	Postviral OD Suspected LPR	Olfactory cleft PRP injection, diet, stress reduction, PPI/alginate
25	F	66	cT3 supraglottic cancer, Weigt loss (6 kg), Dysphagia	Radiotherapy for supraglottic cancer (10-y), hypertension	Epiglottis carcinoma	Neck CT, PetCT *Biopsy: carcinoma*	Second supraglottis carcinoma	Salvage surgery
26	F	49	Aspirations, cough, globus, throat, sticky mucus (9-mo)	None	Coated tongue, normal FEES, laryngeal inflammation	Videofluoroscopy	Suspected LPR	Diet, stress reduction, PPI/alginate
40	F	45	Singer with difficulty to reach high-pitch sounds (6-mo)	Thyroidectomy (12-mo), hip prosthesis (2-y)	Normal vocal cord mobility, hyposensitivity right tongue base	Voice quality assessment	Superior laryngeal nerve injury during surgery	Speech therapy
19	F	61	Right parotid tumor,	Gastritis	Right parotid mass	Neck MRI and CT	Parotid lympho-	Imaging and cytology

			Symptoms (duration)	History	Examination	Investigation	Diagnosis	Treatment
			progressive growth (6-mo)	HIV, pacemaker		Cytology (US)	epithelial cyst	
22	F	36	20 kg loss after a diet, dysphonia, voice fatigue (3-mo)	None	Glottal insufficiency	Voice quality assessment	Glottis insufficiency	Speech therapy, vocal cord augmentation
28	F	36	Dysphonia, voice fatigue (6-mo)	Asthma, inhaled corticosteroids (9-mo)	Vocal fold dryness, sticky mucus	Voice quality assessment	Laryngitis post-inhaled corticosteroids	Stop inhaled corticoids/ change drugs
29	M	66	Bilateral vocal cord paralysis postthyroidectomy, tracheotomy, Wish for decannulation	Thyroid cancer Thyroidectomy Tracheotomy	Bilateral vocal cord paralysis in adduction	Neck CT scan	Bilateral vocal cord paralysis	Bilateral CO2 anterior crico-arytenoidectomy
33	M	40	Progressive dyspnea when playing trumpet, neck mass, dysphagia (9-mo)	None	Left laryngeal ventricle hypertrophy, left neck mass	Neck CT	Laryngocele	Surgery
34	M	70	Dysphagia, globus, throat pain (1-y)	Cervical arthro-desis (1-y), diabetes, hypertension	FEES: normal	Videofluoroscopy Neck CT	Arthrodesis-related dysphagia (iatrogenic)	Speech therapy (swallowing)
38	M	20	Left tonsil ulceration (3-mo)	Oral sexual practice	Left tonsil ulceration	Biology (sexual diseases), biopsy & culture	Syphilis	Antibiotics
35	F	36	Dysphonia, throat pain Voice professional (12 mo)	Vocal cord nodule surgery (12 mo)	Lack of vibration of vocal cord	Voice quality assessment	Vocal fold scars	Speech therapy, resection of scars, PRP injection
36	F	41	Sudden dysphonia (12-mo)	Diabetes, burnout	Normal cough, aphonia, NFN	Voice quality assessment	Psychogenic dysphonia	Speech therapy, psychotherapy
31	F	66	Recurrent dysphagia, globus, weight loss, telangiectasia (3-y)	Resistant LPR to PPI, alginate, magaldrate	Telangiectasia (fingers), laryngeal inflammation	Manometry, GI, biology (immun), biopsy	CREST syndrome Esophageal scleroderma	Vasodilators, immunosuppressant
32	F	34	Dysphonia, arthralgia,	None	Orange nodules	Voice quality	Bamboo nodes	Corticoids,

voice professional (>12 mo)	on vocal cord	assessment, biology (autoimmun), biopsy	Rheumatoid polyarthritis	speech therapy

Appendix 1 footnotes: The additional examinations in italics consisted of results of examination at the consultation time. Abbreviations: COPD=chronic obstructive pulmonary disease; COVID-19=coronavirus disease 2019; CRS(w)NP=chronic rhinosinusitis (without) nasal polyposis; CT=computed tomography; EMG=electromyography; FEES=fiberoptic endoscopic evaluation of swallowing; FESS=functional endoscopic sinus surgery; EED=external ear duct; ETD=Eustachian tube dysfunction; GERD=gastroesophageal reflux disease; GI=gastrointestinal endoscopy; LPRD=laryngopharyngeal reflux disease; MRI=magnetic resonance imaging; NFN=normal nasofibroscopy; NSAID=non-steroidal anti-inflammatory drug; OD=olfactory dysfunction; OSAS=obstructive sleep apnea syndrome; PPI=proton pump inhibitors; PRP=platelet-rich plasma; PY=pack/year; US=ultrasonography.

Appendix 2: ChatGPT findings regarding clinical cases.

	Otolaeryngologist		ChatGPT	
N	Diagnosis	Additional examination	Differential diagnosis	Treatment regarding the clinical presentation
37	Recurrent tonsil infections	Throat culture, Neck CT or MRI, PSG	Chronic tonsillitis, tonsillar hypertrophy, tonsilloliths	Antibiotics, tonsillectomy
7	LPR	Sinus CT, sinus X-Ray, biology (CRP)	Acute sinusitis, GERD, upper aerodigestive tract infection, COPD, viral bronchitis.	PPI and H2 blockers Diet and stress management
8	Suspected LPR	Sinus CT, X-Ray, Biology (CRP)	Acute sinusitis, GERD, laryngopharyngeal infection, COPD, viral bronchitis.	PPI and H2 blockers Diet and stress management
9	Suspected LPR	Neck CT, prick test, Genetics for Ehlers Danlos swallowing study, GI	LPR, chronic pharyngitis, vocal cord dysfunction, eosinophilic esophagitis, oral candidosis, tongue tonsillitis	Speech therapy, PPI and H2 blockers Diet and stress management
21	Hypopharyngeal primary carcinoma	Biopsy, neck CT or MRI.	Laryngeal or pharyngeal carcinoma, metastasis, benign mass, granulomatosis, pseudotumor	Biopsy, neck CT or MRI.
23	Vocal cord paralysis	EMG, Neck CT and MRI	Recurrent laryngeal nerve injury	Speech therapy, medialization, thyroplasty, reinnervation
3	Chronic otitis media, recurrent suspected LPR	Audiometry, Tympanometry pH metry, GI, throat bacteriology	Chronic otitis media (effusion/suppurative), cholesteatoma, tympanosclerosis	Nasal corticoids or transtympanic tube, Allergy checkup.
17	Reinke edema	Vocal cord biopsy, stroboscopy, lung testing (spirometry)	Reinke edema, vocal cord dysfunction, vocal cord polyps, bilateral vocal cord nodules.	Vocal hygiene, speech therapy, smoking cessation, surgery
5	Recurrent/ chronic LPR	pH metry, GI, neck CT, sinus/throat cultures	GERD, unspecified CRS, allergic rhinitis, LPR, Chronic tonsilitis	PPI and H2 blockers Diet and stress management

1	Salivary lithiasis	Neck US, MRI, Biology, Prick skin tests	Adenitis, Abcess, Adenopathy, Parotidis, Thyroiditis.	Hot compress, pilocarpin, Surgery
10	Vocal cord nodules Suspected chronic LPR	Neck CT, voice quality assessment, allergy testing	Vocal cord nodules, vocal cord polyps, GERD, LPR, chronic laryngitis.	Corticoids, voice rest, PPI and H2 blockers Diet and stress management
20	Vocal cord hemorrhage	None	Vocal cord hemorrhage	Voice rest, anti-inflammatory drugs, vocal cord abuse reduction
14	LPR	Neck CT, throat cultures, allergy tests	Acute laryngitis, chronic laryngitis, GERD, COPD, vocal cord nodules.	PPI and H2 blockers Diet and stress management
2	Bilateral EED Stenosis, acute suspected LPR	Ear CT. pH metry, GI, throat bacteriology	EED stenosis, Chronic otitis media, Presbycusis, ETD, GERD, CRS, allergic rhinitis, LPR, chronic tonsilitis	Canaloplasty, balloon dilatation, saline irrigation, Skin flap, PPIs, H2 blockers, Diet/Stress management.
12	Eustachian tube dysfunction, suspected LPR	Neck CT, thyroid function tests	LPR, unspecified CRS, hypothyroidism, bronchitis, cancer related symptoms (unspecified ?)	PPI and H2 blockers, diet and stress management, nasal decongestants, antibiotics and corticoids
13	LPR	Neck CT, throat swab, allergy testing	Chronic tonsilitis, chronic pharyngitis, GERD, LPR	PPI and H2 blockers Diet and stress management
15	Resistant LPR to PPI, infectious laryngitis	Neck CT, salivary gland function test, Allergy, polysomnography	CRSwNP, chronic tonsilitis, GERD, OSAS, Sjögren syndrome	PPI and H2 blockers, stress reduction, diet, antibiotics
16	Recurrent/ chronic nonacid LPR	Esophageal manometry, pH Metry, allergy	LPR, GERD, CRSwNP, chronic tonsilitis, postnasal drip	PPI and H2 blockers, stress reduction, diet, nasal saline irrigation and corticosteroids
18	Left vocal fold polyp Suspected LPR	Neck CT Laryngeal biopsy	LPR, CRSwNP, Allergic rhinitis, vocal fold polyp, vocal fold nodules	Polypectomy/resection of mass, corticoids, speech therapy
24	Laryngeal carcinoma resistant to chemoradiation	Neck CT or MRI, oncological board assessment	Persistent carcinoma	Salvage surgery, immunotherapy, clinical trials, palliative

11	Chronic otitis media, LPR, Eustachian tube dysfunction	Throat swab culture, audiometry, swallowing study, GI	Laryngitis, chronic pharyngitis, otitis media, GERD,	Speech therapy, PPI and H2 blockers, diet and stress management, GERD-surgery.
30	Eagle syndrome	Neck CT or MRI	Tonsilloliths, tonsillar hypertrophy, chronic pharyngitis, GERD	PPIs, H2 blockers, diet/Stress management, saltwater gargles, good oral hygiene/hydration, tonsillectomy
27	Laryngeal hypersensitivity	pH testing, manometry, allergy, inhaled broncho-dilators, methacholine test	Cough variant asthma, GERD, LPR	PPIs, inhaled corticoids, speech therapy
6	Nasopharyngeal reflux	RAST (IgE), nasal cytology (eosinophiles), sinus CT	LPR, allergic rhinitis, CRSNP	speech therapy PPI and H2 blockers
39	Suspected LPR	pH study, GI, allergy evaluation	LPR, chronic laryngitis, allergic rhinitis	PPIs, H2 blockers, diet/Stress management, speech therapy
4	Postviral OD Suspected LPR	Psychophysical evaluations, sinus X-Ray, Neck CT, Biology (B12 level, CRP)	Postviral OD, CRSNP, CRSwNP, Neurodegenerative disease, Autoimmun disease	PPIs, H2 blockers, Diet/Stress management.
25	Second supraglottic carcinoma	Neck CT or MRI, biopsy	Supraglottic carcinoma	Radiation, chemotherapy or chemoradiotherapy, surgery, immunotherapy
26	Suspected LPR	pH testing, EMG	LPR, chronic laryngitis, laryngeal hypersensitivity, postnasal drip, chronic tonsilitis.	Diet, stress reduction, mucolytics, nasal corticoids, antihistamine, PPIs, H2 blockers, laryngeal desensitization (breath)
40	Superior laryngeal nerve injury during surgery	EMG, Neck CT and MRI, neurological consultation	Superior laryngeal nerve dysfunction, hypoglossal dysfunction, vocal cord muscle atrophy	Speech therapy, nerve reconstruction
19	Parotid lympho-	Neck US, CT, cytology,	Peiomorphic or Whartin tumor,	Surgery

	epithelial cyst	biology (CRP)	malignancy, lymph node, metastasis	
22	Glottis insufficiency	None	Muscle atrophy, vocal cord paresis, psychogenic dysphonia, vocal fold lesions	Biology (autoimmun diseases), speech therapy, psychological support.
28	Laryngitis post-inhaled corticosteroids	Lung assessment	Vocal cord dysfunction, Reinke edema, vocal cord nodules, vocal cord polyps, muscle tension dysphonia	Continue asthma treatment
29	Bilateral vocal cord paralysis	-	Bilateral vocal cord paralysis (adduction) and permanent tracheotomy.	Vocal cord lateralization, speech therapy, electrolarynx,
33	Laryngocele	Neck CT or MRI, biopsy	Laryngeal papillomatosis, laryngeal cancer, laryngocele, lymphadenopathy	Surgery, radiation, voice therapy
34	Arthrodesis-related dysphagia (iatrogenic)	Esophageal manometry, videofluoroscopy, endoscopic US, biopsy	Achalasia, esophageal stricture, esophageal spasm, esophageal cancer	Calcium channel blockers, nitrates, or botulinum toxin injections, esophageal dilatation, surgery or chemoradiation
38	Syphilis	Throat culture, Neck CT or MRI, biopsy	Bacterial or viral tonsillitis, tonsil abscess, HPV infection, cancer	Antibiotics, analgesics
35	Vocal fold scars	Neck CT or MRI, EMG	Vocal fold scar, granuloma, Reinke edema, vocal cord paresis,	Speech therapy, corticoids, surgery of the lesion
36	Psychogenic dysphonia	EMG and psychological evaluation	Functional dysphonia, muscle tension dysphonia, conversion disorder, vocal fold paresis or spasmodic dystonia	Speech therapy, psychotherapy, muscle relaxants
31	CREST syndrome scleroderma	GI, biology	Scleroderma, GERD	HE blockers, prokinetic, esophageal dilatation
32	Bamboo nodes Rheumatoid arthritis	-	Reinke edema, vocal cord polyps	Speech therapy, vocal cord surgery (removal of lesion)

Appendix 2 footnotes: Abbreviations: COPD=chronic obstructive pulmonary disease; COVID-19=coronavirus disease 2019; CRS(w)NP=chronic rhinosinusitis (without) nasal polyposis; CT=computed tomography; EMG=electromyography; FEES=fiberoptic endoscopic evaluation of swallowing; FESS=functional endoscopic sinus surgery; EED=external ear duct; ETD=Eustachian tube dysfunction; GERD=gastroesophageal reflux disease; GI=gastrointestinal endoscopy; LPRD=laryngopharyngeal reflux disease; MRI=magnetic resonance imaging; NFN=normal nasofibroscopy; NSAID=non-steroidal anti-inflammatory drug; OD=olfactory dysfunction; OSAS=obstructive sleep apnea syndrome; PPI=proton pump inhibitors; PRP=platelet-rich plasma; PY=pack/year; US=ultrasonography.

Appendix 3: Main Diagnoses.

Primary diagnoses	Easy ACCS 6-14	Moderate ACCS 15-23	Difficult ACCS 24-30	Total
Laryngeal disorders				
Chronic suspected or confirmed laryngopharyngeal reflux	6	7	1	14
Vocal fold polyp, hemorrhage, nodules or scarring	2	1	1	4
Acute suspected or confirmed laryngopharyngeal reflux	2	1	0	3
Recurrent suspected or confirmed laryngopharyngeal reflux	1	2	0	3
Unilateral or bilateral vocal cord paralysis	1	0	1	2
Bacterial laryngitis	0	1	0	1
Reinke edema	1	0	0	1
Bamboo nodes (vocal folds)	0	0	1	1
Glottis insufficiency	0	0	1	1
Laryngeal primary hypersensitivity	0	1	0	1
Iatrogenic laryngitis	0	0	1	1
Laryngocele	0	0	1	1
Iatrogenic laryngeal superior nerve injury	0	0	1	1
Psychogenic dysphonia	0	0	1	1
Swallowing disorders				
Cervical arthrodesis inducing iatrogenic dysphagia	0	0	1	1
Eagle syndrome	0	1	0	1
Esophageal scleroderma (CREST syndrome)	0	0	1	1
Recurrent tonsil infection	1	0	0	1
Head & Neck disorders				
Salivary lymphoepithelial cyst	0	0	1	1
Salivary lithiasis	1	0	0	1
Hypopharyngeal primary carcinoma	1	0	0	1
Supraglottic laryngeal carcinoma (resistant to radiation)	0	1	0	1
Second laryngeal carcinoma	0	0	1	1
Pharyngeal syphilitic ulceration	0	0	1	1
Associated otological or rhinological conditions				
Chronic otitis media	1	1	0	2
Eustachian tube dysfunction	0	2	0	2
Postviral olfactory dysfunction	0	1	0	1
Rheumatoid polyarthritis	0	0	1	1
Bilateral ear external duct stenosis	1	0	0	1

Appendix 3 footnotes: Abbreviations: ACCS=Amsterdam Clinical Challenge Scale test.