Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# AI-Assisted Voice Analysis for Laryngeal Cancer: Primary Care Perspectives on Clinical Implementation

**Cindy C. Ejindu**
Morehouse School of Medicine
**Meghana Darla**
Emory University School of Medicine
**Karim Asi**
Emory University School of Medicine
**Anthony Law**
anthony.law@emory.edu

Emory University School of Medicine

**Research Article**

**Keywords:** Artificial intelligence, Laryngeal cancer, Dysphonia, Primary care, Voice analysis, Clinical decision support, Implementation science, Cancer risk stratification, Mixed-methods research

**Posted Date:** November 19th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11223624/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

**Additional Declarations:** No competing interests reported.

## Abstract

# Background

Voice-based artificial intelligence may offer a noninvasive approach to identifying patients at elevated risk for laryngeal cancer, but successful implementation depends on its acceptability and fit within primary care. This study examined primary care clinicians' approaches to persistent dysphonia, perceptions of artificial intelligence–assisted voice-based risk stratification, and factors influencing clinical implementation.

# Methods

We conducted a cross-sectional mixed-methods study using semi-structured interviews and an anonymous electronic survey that could be completed independently. Interviews explored dysphonia evaluation, referral practices, technology use, and perceived implementation considerations. Survey items assessed clinical reasoning using two dysphonia vignettes, technology readiness, application acceptability, and perceived barriers. Survey responses were summarized using descriptive statistics; no inferential statistical testing was performed. Interviews were analyzed using reflexive thematic analysis, and qualitative and quantitative findings were subsequently integrated at the aggregate level.

# Results

Fifteen clinicians participated in interviews, and 16 anonymous survey responses were received. In a clinical vignette of a patient with a smoking history who presented for persistent dysphonia, all respondents identified laryngeal or other head and neck malignancy in the differential diagnosis, compared with 12 of 15 (80.0%) for a similar clinical vignette without a smoking history. Thirteen of 16 survey respondents (81.3%) reported a positive or very positive initial impression of the proposed application, and 11 (68.8%) were comfortable or very comfortable integrating it into clinical practice. Interviews revealed that clinicians generally lacked a standardized pathway for escalating persistent dysphonia and viewed artificial intelligence primarily as decision support rather than a replacement for clinical judgment. Acceptance was conditional on demonstrated accuracy, minimal workflow disruption, appropriate clinician oversight, efficient referral pathways, and attention to equity, privacy, and accountability.

# Conclusions

Primary care clinicians demonstrated interest in voice-based artificial intelligence for laryngeal cancer risk stratification, but willingness to adopt the technology depended on more than diagnostic performance. Successful implementation will require integration into existing workflows, preservation of

clinical oversight, clear pathways from risk identification to definitive evaluation, and attention to patient trust and equitable access.

## Introduction

Laryngeal cancer is the second most common head and neck cancer, accounting for 211,000 new cases and 126,000 deaths per year worldwide [1]. When diagnosed and treated early, laryngeal cancer is associated with favorable oncologic outcomes and preservation of voice and swallowing function [2–4]. However, many patients present with advanced-stage disease, which is associated with substantially worse survival, greater morbidity, and reduced quality of life [5, 6]. Dysphonia is a common presenting symptom of laryngeal cancer and an important clinical feature prompting evaluation [7]. Current clinical guidelines recommend referral for laryngoscopy when dysphonia fails to resolve or improve within 4 weeks, or earlier when a serious underlying cause is suspected [8].

However, primary care providers (PCPs) must still determine which symptomatic patients require more urgent escalation, often without direct visualization of the larynx. This assessment may be particularly challenging when classic risk factors or other concerning features are absent. Delays can occur along this pathway: Smith et al. reported an average of approximately 40 days between a patient's primary care visit and first otolaryngology visit among patients ultimately diagnosed with laryngeal cancer [9]. Additionally, an analysis of patients with laryngeal cancer within the National Cancer Database found that delays in treatment are associated with worse survival outcomes, becoming statistically significant at 46 days [10]. Longer diagnostic delays may have a detrimental effect on survival in laryngeal cancer compared with other head and neck malignancies [11]. These findings highlight the need for strategies that support timely risk stratification and escalation of symptomatic patients before definitive laryngeal evaluation.

Artificial intelligence (AI) offers the potential to provide an additional voice-derived risk signal earlier in this clinical pathway. Across several malignancies, AI-based technologies have demonstrated the ability to assist with risk assessment and diagnostic triage, including applications in colorectal, lung, breast, cervical, and skin cancers [12–16]. Prior studies have surveyed PCPs on their preferences for implementation of these types of tools, with most PCPs demonstrating openness to using them in their practices [17, 18]. Despite growing interest in AI-assisted cancer risk stratification, evidence specific to laryngeal cancer remains limited.

Most AI research has focused on analysis of laryngeal imaging to predict malignancy [19, 20]. However, this is unlikely to be applicable to the primary care setting, where laryngeal imaging may not be readily available. Voice-based AI may offer a more accessible approach to symptomatic risk stratification. Emerging studies suggest that machine-learning models can distinguish voice characteristics associated with laryngeal malignancy and may therefore provide a noninvasive estimate of cancer risk in patients presenting with voice changes [20–22]. Although these studies demonstrate encouraging diagnostic performance, nearly all have focused on algorithm development and validation rather than

implementation within the clinical pathway. The clinical value of voice-based AI therefore depends not only on diagnostic performance, but on how its risk estimates inform clinician decision-making and referral urgency within the workflow.

We conducted a mixed-methods pre-implementation study examining the primary-care-to-otolaryngology pathway for dysphonia, including current evaluation and referral practices, perceptions of AI-assisted voice-based risk stratification for laryngeal cancer, and clinical, workflow, ethical, and systems-level implementation considerations. We further sought to characterize how an AI-derived risk signal could be incorporated into routine primary care and connected to definitive laryngeal evaluation.

# Methods

### Study Design

We conducted a cross-sectional convergent mixed-methods study to characterize PCPs' perceptions of a voice-based AI risk-stratification tool designed to identify patients at elevated risk for laryngeal cancer. The qualitative and quantitative strands were collected independently through semi-structured interviews and an anonymous electronic survey, respectively. Both approaches were used to pair in-depth exploration of clinicians' experiences and implementation considerations with a structured descriptive assessment of clinical reasoning, technology readiness, and application acceptability. Participants could complete either component independently. Qualitative and quantitative data were analyzed separately using their respective analytic approaches and subsequently integrated at the aggregate level to identify areas of convergence, complementarity, and divergence between qualitative themes and survey findings. Mixed-methods reporting followed the Good Reporting of a Mixed Methods Study (GRAMMS) recommendations [23], with the completed checklist provided in Appendix A.

### Participants and Recruitment

Clinicians with current outpatient primary care responsibilities were recruited using purposive and convenience sampling from academic practices, federally qualified and community health centers, private or direct-primary-care practices, and general outpatient practices. Recruitment emails were distributed from January through December 2025 to clinicians identified through practice websites, the Georgia Primary Care Association directory, institutional contacts, and professional networks. Recruitment materials included invitations to participate in a semi-structured interview and a link to an anonymous electronic survey. Due to the survey being anonymous, responses could not be linked to individual interview participants. Recruitment occurred through multiple distribution channels; therefore, the total number of clinicians who received or viewed the study invitation could not be reliably determined and a precise survey response rate could not be calculated. Participation was voluntary, and no compensation was provided.

### Data Collection

## Interview Procedures

After providing informed consent, participants completed an approximately 30-minute research visit via Zoom. Interviews were conducted by four members of the research team: two laryngologists, a medical student, and a research assistant. A semi-structured interview guide was used to explore the frequency and evaluation of voice complaints, perceived laryngeal cancer risk factors, previous experience with symptomatic risk stratification for early laryngeal cancer, access to otolaryngology, technology and AI use, workflow requirements, privacy, equity, and regulatory oversight. The interview guide standardized the core domains discussed; however, the description of the proposed tool was not scripted verbatim across interviews. One interviewer (AL) was involved in development of the proposed voice-analysis technology. Another interviewer (CE) had prior academic relationships with four participants through prior clinical training and/or institutional academic roles.

Participants were verbally introduced to a proposed voice-based AI risk-stratification tool intended for use in patients with voice changes. The tool was described as a smartphone- or tablet-based tool that would analyze a brief voice sample and provide an estimate of laryngeal cancer risk to support clinical decision-making and referral. No prototype or mock-up was shown, and participants did not interact with the tool during the study. Preliminary testing demonstrating approximately 95% accuracy in detecting laryngeal mass under controlled conditions was discussed during interviews; sensitivity and specificity were not provided. Potential downstream workflows, including expedited or automatic referral to otolaryngology and varying degrees of clinician involvement, were explored as hypothetical implementation options rather than fixed features of the risk-stratification tool.

Interviews were recorded using Zoom (Zoom Communications, Inc) and automatically transcribed. Transcripts were subsequently reviewed for obvious transcription errors while preserving participants' intended meaning and were de-identified prior to qualitative analysis. The complete interview guide is provided in Appendix B. Reporting of the qualitative interview component followed the Consolidated Criteria for Reporting Qualitative Research (COREQ) [24], with the completed checklist provided in Appendix C.

## Survey procedures

An electronic survey was administered using Qualtrics and could be completed independently of the interview. The survey began with two parallel clinical vignettes involving persistent dysphonia and asked respondents to identify their leading differential diagnoses and proposed evaluation. Additional items assessed demographics, baseline technology use and AI familiarity, perceptions of the proposed tool, comfort with clinical integration, preferred device and personnel, and perceived implementation barriers. Items included Likert-type, multiple-choice, and open-ended response formats. The complete survey instrument, along with the clinical vignettes, is provided in Appendix D. Reporting of the electronic survey followed the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [25], with the completed checklist provided in Appendix E.

Sample Size

Because this was an exploratory mixed-methods study, no formal power calculation was performed. The quantitative component was intended to characterize implementation considerations rather than test hypotheses, while the qualitative sampling strategy sought information-rich perspectives across differing outpatient primary care environments and practice structures. Sample adequacy was considered using principles of information power described by Malterud et al. [26]. Initial recruitment yielded a predominantly academic sample; recruitment was subsequently expanded to include clinicians from rural and other nonacademic practice settings to increase variation in practice context and obtain diverse information-rich perspectives. Recruitment sought variation in clinician experience and practice context rather than a predetermined numerical sample size.

# Data Analysis

Quantitative analysis

The quantitative analysis was exploratory and descriptive. Survey responses were summarized descriptively using frequencies and percentages, and 95% Wilson score confidence intervals. Analyses were performed using Microsoft Excel. Exploratory descriptive analyses examined comfort with integration across PCP characteristics, including age, years in practice, technology use, and AI familiarity, as well as reported conditions influencing potential integration. For multiple-response items, each option was calculated using the total number of survey respondents as the denominator; percentages therefore were not expected to total 100%. Free-text responses regarding preferred personnel were manually categorized as physician, registered nurse, medical assistant, any trained team member, or nonspecific clinician. Responses could receive more than one category when participants identified multiple acceptable roles.

Open-ended vignette responses were manually coded by one investigator (CE) into non-mutually exclusive diagnostic and evaluation/management categories. Diagnostic categories included laryngeal or other head and neck malignancy; benign laryngeal or vocal fold pathology, including laryngitis, vocal fold nodules or polyps, dysfunction, and paralysis; reflux or esophageal disease; pulmonary disease or lung malignancy; neurologic causes; thyroid pathology; and other considerations. Evaluation and management categories included otolaryngology referral, additional history and/or focused examination, imaging, laboratory testing, laryngoscopy or direct visualization, and empiric medical therapy. Responses could receive multiple codes when more than one diagnosis or management step was identified. Coding was based conservatively on explicitly stated content; for example, an otolaryngology referral was not assumed to include laryngoscopy unless direct visualization was specifically mentioned. Responses stating “same as above” were interpreted using the paired response to the preceding vignette, whereas non-substantive responses that did not provide clinical information were excluded. Ambiguous or nonspecific responses were reviewed in the context of the full response and were assigned a category only when the intended meaning could be determined from the participant’s wording. This structured

categorization was used solely to summarize vignette responses and was conducted separately from the reflexive thematic analysis of interview transcripts. Due to the small sample size and exploratory study objective, findings were summarized descriptively and no inferential statistical or hypothesis testing was performed.

### Qualitative Analysis

Interview transcripts were analyzed using reflexive thematic analysis following Braun and Clarke's six-phase approach [27]. The analysis was primarily semantic and used a hybrid deductive–inductive orientation. Deductive attention was given to domains established by the study objectives, including clinical recognition, workflow, trust, referral access, privacy, and equity, while inductive coding was used to identify unanticipated meanings and relationships within those domains.

ATLAS.ti Web Version 9.24.2 (ATLAS.ti Scientific Software Development GmbH) was used to facilitate organization and coding of the qualitative data. The lead qualitative analyst (MD) reviewed the transcripts, generated and iteratively refined codes, and developed candidate themes through repeated examination of coded extracts and the full dataset. Candidate themes were reviewed and refined for internal coherence, conceptual distinctiveness, and relevance to the research questions and were subsequently defined and named. A second researcher (CE) subsequently reviewed the candidate thematic structure, theme definitions, discrepant or potentially contradictory cases, supporting data extracts, and selected quotations. Differences in interpretation were discussed and used to further examine and refine the final thematic analysis. This collaborative analytic review was intended to support reflexive engagement with the data rather than to establish coding agreement; therefore, an intercoder-reliability statistic was not calculated.

### Mixed-methods integration

Quantitative and qualitative findings were integrated after the two components had been analyzed separately. Survey findings were compared with the qualitative themes to determine whether interview data confirmed, explained, expanded, or contradicted the quantitative patterns. Attention was given to acceptability, workflow integration, accuracy and reliability, patient acceptance, staffing, referral pathways, and governance.

### Ethical considerations

The study was approved by the Emory University Institutional Review Board (STUDY00007975). Participants provided informed consent before completing any study activities. Names and other directly identifying information were excluded from reported findings.

### Reflexivity Statement

We acknowledge that the positionality of the research team, including prior knowledge and expertise in AI, laryngology, voice disorders, and clinical research, may have influenced the framing, conduct, and

interpretation of the study. Interviews were conducted by research team members with differing clinical and research backgrounds using a common semi-structured interview guide. To support reflexive interpretation, the lead qualitative analyst (MD) conducted the primary coding and theme development, while a second qualitative reviewer (CE) subsequently reviewed the thematic structure, discrepant cases, and supporting quotations. Differences in interpretation were discussed and used to refine the final analysis. We have attempted to represent both supportive and critical clinician perspectives throughout the analysis and reporting.

# Results

Participant Characteristics

Fifteen clinicians participated in semi-structured interviews, and 16 anonymous survey responses were received. Characteristics of the two unlinked samples are presented in Table 1. Interview participants represented varied clinician types, practice environments, and geographic settings, while demographic information from the anonymous survey included age, gender, and years in practice. Several interview participants also held overlapping roles as clinical educators, preceptors, or medical directors. All quantitative survey analyses were exploratory and hypothesis-generating.

Table 1
Characteristics of the interview and anonymous survey samples.

| **A. Interview Sample, N = 15** | |
|---|---|
| **Characteristic** | **n (%)** |
| **Clinician Type** | |
| Physician | 14 (93.3%) |
| Physician Assistant | 1 (6.7%) |
| **Gender** | |
| Women | 10 (66.7%) |
| Men | 5 (33.3%) |
| **Practice Type** | |
| Academic/teaching practice | 10 (66.7%) |
| Federally qualified health center | 4 (26.7%) |
| Direct primary care/private practice | 1 (6.7%) |
| **Geographic Setting** | |
| Urban/Suburban | 11 (73.3%) |
| Rural | 4 (26.7%) |
| **B. Survey Sample, N = 16** | |
| **Demographics** | **n (%)** |
| **Age, years** | |
| 30–39 | 3 (18.8%) |
| 40–49 | 7 (43.8%) |
| 50–59 | 3 (18.8%) |
| ≥ 60 | 3 (18.8%) |
| **Gender** | |
| Women | 12 (75.0%) |
| Men | 4 (25.0%) |
| **Years in practice** | |
| < 5 | 1 (6.3%) |

| A. Interview Sample, N = 15 | |
|---|---|
| **Characteristic** | **n (%)** |
| **Clinician Type** | |
| 5–10 | 2 (12.5%) |
| 11–20 | 7 (43.8%) |
| > 20 | 6 (37.5%) |

Interview and anonymous survey data represent unlinked samples; the extent of overlap could not be determined. A. Characteristics of participants that completed the semi-structured interview. B. Characteristics of participants who completed the anonymous online survey.

Fourteen interviews were available as verbatim transcripts. For one additional interview, the original transcript was unavailable and only an automatically generated Zoom meeting summary remained. The summary was not coded, used to generate themes, or used for verbatim quotation. Five themes were constructed from the interview data and are summarized in Table 2.

Table 2
Qualitative themes, meanings, and representative quotations.

| Theme | Description | Representative quotation |
|---|---|---|
| Recognition and escalation of persistent dysphonia | Clinicians relied on symptom persistence, treatment response, and combinations of risk factors rather than a standardized pathway. | *“I don’t know of any… risk… rule or anything like that.”* —P03 |
| Conditional acceptance of AI-assisted risk stratification | Participants generally viewed AI as decision support and varied in their willingness to rely on automated recommendations. | *"I would trust but verify."* —P05 |
| Workflow requirements for implementation | Adoption depended on minimal workflow disruption, flexible staffing, and straightforward documentation. | *"Technology that doesn't help you, but costs you time..."* —P09 |
| From risk-stratification to definitive laryngeal evaluation | Clinicians viewed the tool as more useful when elevated-risk results facilitated timely referral and follow-up. | *"Links to care [are] probably more helpful than… tools."* —P12 |
| Equity, privacy, and accountability | Clinicians emphasized fairness, transparency, privacy, and accountability in addition to technical performance. | *"We don't own your voice; you own your voice."* —P08 |

Five themes were developed from the interview data. Representative quotations are provided to illustrate each theme.

Recognition and Escalation of Persistent Dysphonia

Participants responded to two parallel clinical vignettes describing a 63-year-old man with progressively worsening “airy” hoarseness for 3 months, increasing difficulty producing sufficient volume, and no symptoms of an acute upper respiratory infection. The first patient had Chronic obstructive pulmonary disease (COPD) and a 35 pack-year smoking history, whereas the second was a never-smoker with no significant medical history. Fourteen respondents provided analyzable responses to vignette 1, and 15 provided analyzable responses to vignette 2. Responses stating “same as above” were interpreted using the paired vignette 1 response. Non-substantive responses that did not provide clinical information were excluded from analysis. One vignette 2 response stating “duplicate question” was excluded from analysis (Table 3).

Table 3
Clinician responses to two dysphonia vignettes.

| **Differential diagnosis** | **Vignette 1, n (%; 95% CI)[a]** | **Vignette 2, n (%; 95% CI)[b]** |
|---|---|---|
| Laryngeal or other head and neck malignancy | 14 (100.0%; 78.5–100.0) | 12 (80.0%; 54.8–93.0) |
| Benign laryngeal or vocal fold pathology | 11 (78.6%; 52.4–92.4) | 13 (86.7%; 62.1–96.3) |
| Reflux or esophageal disease | 5 (35.7%; 16.3–61.2) | 9 (60.0%; 35.7–80.2) |
| Pulmonary disease or lung malignancy | 3 (21.4%; 7.6–47.6) | 1 (6.7%; 1.2–29.8) |
| Neurologic cause | 2 (14.3%; 4.0–39.9) | 0 (0.0%; 0.0–20.4) |
| Thyroid pathology | 1 (7.1%; 1.3–31.5) | 1 (6.7%; 1.2–29.8) |
| Other considerations[c] | 1 (7.1%; 1.3–31.5) | 2 (13.3%; 3.7–37.9) |
| **Evaluation/management step** | | |
| Otolaryngology referral | 11 (78.6%; 52.4–92.4) | 11 (73.3%; 48.0–89.1) |
| Additional history and/or focused examination | 7 (50.0%; 26.8–73.2) | 6 (40.0%; 19.8–64.3) |
| Imaging considered | 8 (57.1%; 32.6–78.6) | 6 (40.0%; 19.8–64.3) |
| Laboratory testing | 2 (14.3%; 4.0–39.9) | 3 (20.0%; 7.0–45.2) |
| Laryngoscopy or direct visualization explicitly identified | 3 (21.4%; 7.6–47.6) | 4 (26.7%; 10.9–52.0) |
| Empiric medical therapy | 0 (0.0%; 0.0–21.5) | 2 (13.3%; 3.7–37.9) |

Responses are summarized by differential diagnosis and proposed evaluation/management step for parallel clinical scenarios differing primarily by smoking history. a. n = 14. b. n = 15. c. Other considerations included postnasal drainage in Vignette 1 and medication-induced dysphonia and human papillomavirus in Vignette 2.

Clinicians generally recognized malignancy as a potential cause of persistent dysphonia in standardized clinical scenarios, particularly when classic risk factors were present. All 14 respondents to the vignette involving a 35 pack-year smoking history included laryngeal or another head and neck malignancy in the differential diagnosis (100.0%; 95% CI, 78.5%–100.0%), compared with 12 of 15 respondents in the otherwise identical never-smoker vignette (80.0%; 95% CI, 54.8%–93.0%). Otolaryngology referral was proposed by 11 respondents in each scenario (78.6% and 73.3%, respectively).

Interview findings demonstrated greater variability in real-world recognition and escalation. Voice complaints were generally understood first within the context of benign conditions such as viral illness, laryngitis, allergies, postnasal symptoms, or reflux. Their frequency varied across practices and seasons, ranging from rare encounters to several patients per week during respiratory-illness periods. Concern increased with symptom persistence or worsening, failure to improve with treatment, change from baseline, and accumulated red flags such as smoking, older age, weight loss, dysphagia, lymphadenopathy, hemoptysis, or prior head and neck cancer. Direct experience with laryngeal cancer was generally limited.

*“If they are older and smokers, you obviously think about malignancy as a potential cause.”* —P05

*“If they are younger or middle-aged and don't have a history of smoking or many major risk factors… if they don't get better, I will send them to see ENT.”* —P04

However, participants generally described no formal primary-care risk stratification or escalation pathway. Evaluation relied on clinical history, basic head and neck examination, and treatment response. One participant explained:

*“I don't know of any… risk, kind of rule, or anything like that, and so may well exist… I don’t use that. I’m not familiar.”* —P03

Recognition itself could also be difficult when clinicians lacked longitudinal familiarity with patients. Another participant noted that a trainee first had to identify that the voice had changed and that, without sufficient continuity, symptoms might otherwise be attributed to a viral condition. Some participants viewed a standardized aid as potentially useful for promoting more consistent assessment. One participant suggested:

*“Probably having some kind of standardized tool that I could look at just as a key to remind me… to do that.”* —P02

Clinicians recognized obvious cancer risk in standardized scenarios, but real-world escalation remained dependent on individual clinical judgment, symptom persistence, and the accumulation of concerning features.

Conditional Acceptance of AI-Assisted Risk Stratification

The survey cohort was relatively technology-experienced: 14/16 (87.5%) reported being familiar or very familiar with AI in healthcare, and 13/16 (81.3%) reported using technology frequently or always in clinical practice. Interview participants nevertheless varied in their prior clinical use of AI, ranging from little or no direct use to experience with Electronic Health Record (EHR)-based decision support, AI-assisted documentation, risk calculators, and other AI-enabled applications. 13/16 respondents (81.3%) reported a positive or very positive initial impression, and 11/16 (68.8%) reported being comfortable or very comfortable integrating the tool into practice. Although no respondent indicated that the tool would not be beneficial, only 7/16 (43.8%) selected "yes," whereas 9/16 (56.3%) selected "maybe" (Table 4).

Table 4
Survey findings on acceptability and implementation preferences.

| **Survey outcome** | **n (%; 95% CI)** |
|---|---|
| **Initial impression** | |
| Positive/very positive | 13 (81.3%; 57.0–93.4%) |
| Neutral | 3 (18.8%; 6.6–43.0%) |
| **Believed the tool could benefit patients** | |
| Yes | 7 (43.8%; 23.1–66.8%) |
| Maybe | 9 (56.3%; 33.2–76.9%) |
| **Comfort integrating the tool** | |
| Comfortable/very comfortable | 11 (68.8%; 44.4–85.8%) |
| Neutral | 5 (31.3%; 14.2–55.6%) |
| **Preferred device** | |
| Smartphone | 15 (93.8%; 71.7–98.9%) |
| Tablet | 1 (6.3%; 1.1–28.3%) |
| **Preferred personnel for voice assessment**[a] | |
| Physician | 10 (62.5%; 38.6–81.5%) |
| Registered nurse | 5 (31.3%; 14.2–55.6%) |
| Medical assistant | 3 (18.8%; 6.6–43.0%) |
| Any trained team member | 2 (12.5%; 3.5–36.0%) |
| Clinician, unspecified | 1 (6.3%; 1.1–28.3%) |

Findings summarize respondents' perceptions of the proposed application and preferences for its clinical implementation. [a]Manually coded from open-ended responses. Respondents could identify more

than one acceptable role; therefore, percentages do not total 100%.

In exploratory subgroup descriptive analyses, similar patterns were observed by age and years in practice: comfort with tool integration was reported by 8/10 respondents (80.0%) who were younger than 50 years or had ≤ 20 years in practice, compared with 3/6 (50.0%) who were aged ≥ 50 years or had > 20 years in practice. Interestingly, respondents who reported neutral familiarity with healthcare AI (2/2, 100.0%) were more comfortable with integration compared to 9/14 (64.3%) who were familiar or very familiar. Comfort was similar across frequency of clinical technology use. Given the small subgroup sizes, these patterns were considered descriptive and hypothesis-generating (Fig. 1).

Interview findings clarified the conditions underlying this receptivity. Participants generally viewed AI as potentially helpful for identifying subtle patterns, reducing uncertainty, improving efficiency, and prompting earlier action. Many positioned AI as a source of decision support rather than an unquestioned authority. One participant summarized this directly:

*"I would trust but verify. So, if it tells me that's what they think, then I'll refer to [a] specialist to see if it's actually accurate or not."* —P01

Concerns included hallucinations, inaccurate recommendations, missed disease, excessive reliance on automation, and erosion of clinical judgment. Participants emphasized that AI output should be interpreted alongside the patient's history, physical examination, and clinical circumstances.

*"I trust but verify, because AI can make up things, it can hallucinate."* —P07

However, expectations for oversight were not uniform. Some clinicians were comfortable with automated referral when performance had been sufficiently demonstrated. One participant contrasted the model's proposed accuracy with personal diagnostic ability, which was contingent on the expected frequency and consequences of erroneous recommendations:

*"If you're going to do automatic referral… the error rate [has to be] so low that it's a very occasional thing."* —P07

*"If the machine's like, 'Hey, this is a concern,' and you've built a machine that is 90% or 95% accurate… I'm not going to be 95% accurate."* —P08

Clinicians demonstrated conditional acceptance of AI-assisted voice-based risk stratification. Receptivity depended not simply on whether the technology was available, but on confidence in clinical performance, appropriate clinician oversight, the consequences of false-positives, and a clear understanding of how AI-generated risk estimates would inform patient care.

Workflow Requirements for Implementation

Workflow and EHR integration were the most frequently cited implementation barrier, selected by 12/16 respondents (75.0%; 95% CI, 50.5%–89.8%). Accuracy and reliability and lack of time were each

identified by 9/16 respondents (56.3%; 95% CI, 33.2%–76.9%), followed by patient acceptance and engagement (7/16, 43.8%), cost (5/16, 31.3%), and insufficient training or support (4/16, 25.0%) (Fig. 2). 15/16 respondents preferred a smartphone rather than a tablet for application use (93.8%; 95% CI, 71.7%–98.9%). Physicians were the most frequently identified acceptable personnel for voice assessment (10/16, 62.5%; 95% CI, 38.6%–81.5%), although respondents also identified registered nurses, medical assistants, and other trained team members as acceptable personnel (Table 4).

Further exploratory subgroup analyses suggested that several implementation concerns were more frequently reported among respondents who remained neutral about integrating the tool. EHR/workflow integration was identified by all 5 (100%) neutral respondents compared to 7/11 (63.3%) respondents who were comfortable or very comfortable with integration. Similar patterns were observed for accuracy and reliability, cost, and insufficient training or support (Fig. 3). Given the small subgroup sizes, these comparisons were considered descriptive and hypothesis-generating only.

Interview findings further clarified how these barriers operated in practice. Participants evaluated feasibility according to whether risk stratification could be incorporated into existing care rather than operate as another clinical obligation. Participants generally described acceptable durations ranging from approximately 30 seconds to 5 minutes; they were more tolerant of longer recordings when staff or patients could complete them independently or when clinical suspicion was high. Participants also frequently proposed staff-led workflows in which nurses or medical assistants administered the risk stratification tool, and the clinician reviewed the result.

*“If the voice sample requires… 2, 3, 4 minutes of voice recording… I can obviously leave the room, have my nurse record it, and then I can come back in the room after the recording is finished.”* —P05

*“If it's user-friendly enough that the nurse or the medical assistant does it prior to [the visit]… then you kind of have to just investigate, or… figure out next steps.”* —P01

Workflow concerns included separate portals, duplicate documentation, additional logins, excessive clicking, loading delays, manual transfer of results, and shared devices that could create bottlenecks. Participants distinguished beneficial technology from those that increase workload:

*“What I hate is bad technology—technology that doesn’t help you but costs you time and makes you have to do more.”* —P03

*“If it's clunky, and if it's not seamless, and if it's not a quick thing for me to integrate, I'm done.”* —P14

Most participants preferred EHR integration because it could improve documentation, result visibility, referral initiation, and continuity of care. However, EHR integration was not an absolute requirement for every participant. One clinician stated:

*“If it is not built in, I think that that’s okay… not being integrated within the EHR… would not be a detriment for me to use or not use the app.”* —P15

Across both strands, adoption depended on minimizing workflow disruption, enabling flexible team-based administration, and ensuring straightforward documentation and efficient result review.

From Risk Stratification to Definitive Laryngeal Evaluation

Although otolaryngology referral was proposed by most respondents in both dysphonia vignettes (Table 3), interviews demonstrated that identifying a patient as higher risk did not necessarily ensure timely definitive laryngeal evaluation. Participants judged the value of the application partly by what happened after an elevated-risk result. Referral environments differed markedly across practice settings. Clinicians practicing within integrated systems or with established otolaryngology contacts described secure messaging, internal referral coordinators, and relatively rapid access.

*"It's the joys of secure chat… I'll just… send a secure chat in Epic and say, 'Is this a patient that… could be seen sooner?'"* —P01

*"Links to care [are] probably more helpful for me than… tools, you know, if I identify something that's concerning."* —P12

Others reported specialist shortages, uncertain scheduling, insurance restrictions, transportation challenges, rural distance, affordability concerns, and patient reluctance. Participants emphasized that recognizing risk was insufficient when downstream care remained inaccessible:

*"The biggest problem...there… is… getting them to a specialist."* —P11

*"Usually if there's a delay, it's usually just the insurance."* —P08

*"It may require driving… 30 minutes, 45 minutes to get to the bigger city to see that specialist*." —P02

Another described limited visibility after placing an order:

*"It's more of a black box… we're [going to] put an ambulatory referral to ENT and just kind of hope it happens."* —P10

Participants suggested expedited or risk-stratified referrals, e-consultation, direct specialist messaging, automatic referral triggers, and one-click pathways. At the same time, clinicians emphasized that automated referral would not ensure completed care. Patients might still require education, reassurance, financial assistance, transportation support, or active navigation. Risk-stratification was therefore viewed as most clinically useful when an elevated-risk result could be translated into an accountable and actionable pathway to definitive laryngeal evaluation. A study-informed conceptual pathway illustrating how voice-based AI risk stratification could support risk-appropriate referral urgency is presented in Fig. 4.

Equity, Privacy, and Accountability

Concerns about implementation extended beyond accuracy. 9/16 survey respondents (56.3%) identified accuracy and reliability as a potential barrier, while 7/16 (43.8%) identified patient acceptance and engagement (Fig. 2). The interview data provided additional context for these survey findings, showing that concerns about accuracy and patient acceptance encompasses broader questions of fairness, transparency, privacy, and accountability. Participants questioned whether the tool would perform consistently across racial and ethnic populations, dialects, languages, and different voice characteristics.

*"Has the program been designed with different… races, cultures, accents… and has that been considered in the design of the technology and the comparison of the voices?"* —P13

*"Does it work for all people? Does it work for some people… and do we end up underdiagnosing people because of their differences?"* —P09

Equity also involved access. Participants questioned whether uninsured patients, rural populations, and individuals without smartphones or reliable technology would benefit equally. Concerns about false-positive results included unnecessary workup, patient anxiety, physician burden, and increased demand on already limited specialty services. Voice recordings generated specific concerns regarding consent, ownership, storage, and reuse. One participant explained:

*"We don't own your voice; you own your voice. We will listen to it. We will analyze it… but it's still your voice."* —P08

Privacy concerns varied according to what information the tool would collect. Some clinicians expressed little concern if recordings were de-identified and contained no protected health information, while others emphasized secure storage, breach prevention, and clear disclosure regarding access and retention.

*"Well, usually I don't put any patient data in it. I'm just putting in things. There are no identifiers."* —P01

Participants generally supported regulatory or institutional oversight of AI, including monitoring accuracy, effectiveness, safety, data security, and performance across patient populations. Participants also emphasized patient awareness, informed consent, and clinically informed governance.

*"I think it should be regulated… just to make sure it's accurate. And just to make sure that the effectiveness and the efficiency is monitored by someone."* —P05

Together, these perspectives indicated that trust depended not only on model performance. Participants emphasized equitable performance across patient populations, clear communication regarding data use and ownership, and defined responsibility for errors or harms.

Integrated Mixed-Methods Findings

Across the five domains, quantitative and qualitative findings were broadly convergent and complementary. Survey findings identified the frequency of key clinical and implementation concerns,

while interview data provided additional context regarding how those concerns shaped clinician decision-making, workflow feasibility, referral pathways, and trust in AI-assisted risk stratification. The integrated findings and corresponding implications for implementation of the voice-based AI risk-stratification tool are summarized in Table 5.

Table 5
Joint display of quantitative findings, qualitative themes, integrated interpretations, and implementation implications.

| Quantitative finding | Qualitative theme | Integrated interpretation | Implementation implication |
|---|---|---|---|
| Malignancy was identified by 100.0% of respondents in Vignette 1 and 80.0% in Vignette 2 | Recognition and escalation of persistent dysphonia | Clinicians recognized malignancy in standardized cases, particularly when smoking history was present, but interviews showed that real-world escalation remained dependent on symptom persistence, accumulated risk factors, and individual recognition. | The tool may be most useful as a risk-stratification aid, particularly when classic risk factors are absent or voice changes are subtle. |
| ENT referral was proposed by 78.6% and 73.3%, respectively. | | | |
| Accuracy/reliability selected by 56.3% | Conditional acceptance of AI-assisted risk stratification | Accuracy concerns included hallucinations, false positives, real-world performance, and generalizability across populations. | Conduct prospective validation studies and clearly communicate performance and limitations. |
| EHR/workflow integration selected by 75.0% | Workflow requirements for implementation | Participants explained that separate portals, extra clicks, shared devices, and duplicate documentation could prevent use. | Design for low-friction documentation and flexible EHR or secure stand-alone integration. |
| Lack of time selected by 56.3% | | Time tolerance depended on delegation and whether clinicians had to remain present. | Permit MA/RN administration and provide rapid results, followed by clinician review. |
| Physicians were the most frequently preferred personnel for voice assessment (62.5%) | | Physicians were often considered appropriate, but participants also supported nurse or MA administration with clinician review. | |
| All respondents selected “yes” or “maybe” regarding benefit to patients | Risk-stratification to definitive laryngeal evaluation | Favorability was conditional on clinical validation, usability, and timely referral access. | Pilot implementation should evaluate the complete clinical pathway, not only model accuracy. |
| Patient acceptance selected by 43.8% | Equity, privacy, and accountability | Interviews identified consent, voice ownership, privacy, education, and fear as determinants of acceptance. | Use transparent consent language and clear explanations of data use. |

Survey and interview findings were integrated at the aggregate level to identify convergent and complementary patterns. Implications for implementation of AI-assisted voice-based risk stratification in

primary care are described.

## Discussion

While laryngeal cancer often presents with voice changes early in the disease course, delays in diagnosis and treatment are common and have a substantial impact on outcomes [5, 6, 9, 10]. Current American Academy of Otolaryngology–Head and Neck Surgery (AAO-HNS) clinical practice guidelines recommend referral for laryngoscopy when dysphonia fails to resolve or improve within 4 weeks, or earlier when a serious underlying cause is suspected [8]. In primary care, clinicians evaluating persistent dysphonia must determine which symptomatic patients require urgent escalation, often before definitive laryngeal visualization is available. With this in mind, AI-assisted voice analysis may offer an objective risk-stratification signal that helps prioritize patients for specialty evaluation.

A growing body of literature supports the potential of AI-assisted technologies for the detection of laryngeal lesions. A recent systematic review and meta-analysis by Marrero-Gonzalez et al. found that AI-assisted endoscopy is the most extensively studied modality, while evidence for AI-assisted voice analysis remains limited but promising [20]. Included voice-based studies reported diagnostic accuracies approaching 96% for distinguishing laryngeal cancer or glottic neoplasms from healthy controls or benign voice disorders [22, 28]. More recently, analysis of the Bridge2AI-Voice dataset demonstrated the feasibility of a high-sensitivity voice-based model for identifying laryngeal lesions while considering its translation into telemedicine and triage workflows [29]. Although these findings demonstrate the technical feasibility of voice-based AI, most published work has focused on algorithm development and diagnostic performance rather than implementation in primary care. Our study addressed this gap by examining PCPs perceptions of AI-assisted voice-based symptomatic risk stratification and the factors that may influence its implementation in routine clinical practice.

Recognition and Escalation of Persistent Dysphonia.

A finding of this study was that the unmet need was not simply recognition of obvious cancer risk, but consistent risk stratification and escalation among patients presenting with persistent dysphonia. Malignancy was included in the differential diagnosis by 100% of respondents to the smoking-associated vignette and 80% of respondents to the otherwise identical never-smoker vignette. Most also recommended otolaryngology referral in both scenarios. This suggests that participants were generally able to recognize when persistent dysphonia could represent malignancy, even when a classic risk factor such as smoking was absent.

However, interviews revealed that real-world evaluation remained highly individualized. Decisions to escalate care depended on symptom persistence or progression, lack of improvement with empiric treatment, accumulated warning signs, and individual clinical judgment. Gradual voice changes could also be overlooked when continuity was limited, or symptoms were initially attributed to benign conditions. These findings suggest that implementation is not for a lack of clinician awareness, but for the absence of a consistent method for identifying which patients need expedited evaluation. Therefore,

the potential role of voice-based AI may be less about prompting PCPs to consider malignancy, and more about providing an objective risk signal to provide classification and prioritization of symptomatic patients.

Risk Stratification Within the Referral Pathway

Risk stratification is clinically meaningful if the result changes downstream care. An elevated-risk classification was considered useful by participants if it could alter what happened next, especially when referral access varies across practice settings. PCPs with established otolaryngology relationships, internal referral coordinators, or integrated health systems described relatively efficient pathways. In contrast, others reported specialist shortages, insurance restrictions, transportation barriers, rural distance, patient reluctance, and limited follow-up after a referral was placed.

This finding suggests that improved risk-stratification alone may not translate into earlier action when downstream capacity remains limited. A tool that generates additional referrals without a clear process could increase patient anxiety and place further demand on constrained specialty services. Participants therefore favored features such as risk-stratified referrals, e-consultation, direct specialist communication, and a streamlined process. However, they also emphasized that automated referral would not replace patient education and navigation. Overall, the relevant unit of implementation is the entire care pathway, from risk stratification to completed laryngeal evaluation.

Disease- and Modality-Specific Considerations for Voice-Based AI Implementation

In our study, participants generally favored AI as a decision-support tool rather than a replacement for clinical judgment. They emphasized the importance of accuracy, appropriate clinician oversight, and minimal workflow disruption, while raising concerns about transparency, accountability, privacy, bias, and patient acceptance. Notably, interview participants varied in their prior clinical use of AI, suggesting that conditional receptivity to the proposed tool may not be limited to PCPs with prior AI experience. Several considerations identified in this study are consistent with the broader literature on AI implementation in primary care. A recent qualitative systematic review found recurring concerns regarding technical reliability, workflow integration, human oversight, privacy, liability, and bias among PCPs [30]. Similarly, a 2026 systematic review of clinical decision-support systems for early cancer detection in primary care found that workflow integration, skills requirements, interference with clinical decision-making, and medicolegal uncertainty were relevant to systems used for symptomatic assessment [31]. Prior studies across primary and acute care have similarly emphasized reliable and evidence-based performance, clear and transparent outputs, appropriate clinician involvement, and compatibility with clinical workflow [17, 18, 32–36].

Our study extends this literature by identifying several disease- and modality-specific considerations for voice-based AI risk stratification. Voice acquisition is a discrete task that may not require clinician participation. Participants identified team-based workflows in which nurses or medical assistants could obtain the voice sample, while PCPs reviewed the results and determined the next steps. The use of

recorded voice introduced privacy considerations, including patient awareness of recording, ownership of voice data, storage, retention, reuse, and access. These concerns may require specific consent and data-governance approaches. Additionally, participants questioned whether model performance would remain consistent across accents, dialects, languages, racial and ethnic groups, and other voice characteristics. This concern is particularly relevant for a voice-based tool, because differences in speech patterns and vocal characteristics could influence how accurately the model performs across patient populations. Overall, our findings extend broader concerns about AI implementation by identifying issues that are relevant to voice-based risk stratification.

### Implications for Prospective Clinical Trials and Evaluation

Our findings define the next stage of clinical trials. A prospective study should test the voice-based risk-stratification tool within the complete clinical pathway rather than evaluating performance alone. In addition to sensitivity and specificity, future studies should measure whether AI-generated risk estimates alter referral urgency, the proportion of referrals successfully completed, and time from primary care assessment to laryngeal evaluation and diagnosis. Outcomes should also include staff time required for voice acquisition, clinician time, workflow completion, specialty referral volume, and otolaryngology workload. False-positive results should be evaluated not only as a performance metric but also in terms of unnecessary referrals, patient anxiety, and specialty-resource utilization. Finally, model performance should be examined across clinically relevant demographic, linguistic, accent, and dialect subgroups without aggregate performance metrics that may conceal clinically meaningful variation in who is correctly or incorrectly classified. A trial designed around these outcomes would determine whether implementing the voice-based risk stratification tool into primary care actually shortens and improves the pathway to definitive laryngeal evaluation.

### Limitations

This study has several limitations. The small, purposively and conveniently recruited sample limits the generalizability of the findings. The survey analyses were exploratory and hypothesis-generating, as reflected by the wide confidence intervals, and should not be interpreted as confirmatory population estimates. Participation was voluntary, creating the potential for self-selection bias; clinicians with greater familiarity with or interest in AI may have been more likely to participate. Women comprised 75.0% of the anonymous survey sample, a higher proportion than reported across most primary care physician specialties nationally [37]. The anonymous survey and interview datasets could not be linked at the participant level, limiting mixed-methods integration to aggregate comparisons. The qualitative findings were intended to characterize patterns rather than estimating prevalence in the broader primary care workforce. Participants were verbally introduced to the concept but did not interact with a prototype or use the tool in clinical practice. One interview transcript was unavailable, and the qualitative analysis was therefore based on 14 verbatim transcripts. The proposed tool was described verbally, and preliminary performance was presented to participants as approximately 95% accurate under controlled conditions without accompanying sensitivity or specificity estimates. This may have influenced

participants' perceptions of the tool and limited their ability to evaluate the potential tradeoffs associated with false-positive and false-negative results. Finally, the open-ended interview format supported in-depth exploration but introduced variation in the questions and follow-up prompts received by participants.

# Conclusion

In this exploratory mixed-methods study, primary care clinicians described conditional receptivity to a proposed voice-based AI tool for laryngeal cancer risk stratification. Adoption depended on workflow integration, timely access to otolaryngology care, clinician trust, equitable performance across diverse populations, transparent data practices, and protection of patient privacy. These findings provide guidance for developing and prospectively evaluating voice-based AI symptomatic risk stratification within the complete primary-care-to-otolaryngology pathway.

# Abbreviations

AI
Artificial Intelligence
PCP
Primary Care Providers
EHR
Electronic Health Record
GRAMMS
Good Reporting of a Mixed Methods Study
COREQ
Consolidated Criteria for Reporting Qualitative Research
CHERRIES
Checklist for Reporting Results of Internet E-Surveys

# Declarations

**Ethical approval and consent to participate:**

The study was approved by the Emory University Institutional Review Board (STUDY00007975). All participants provided consent prior to participation and data collection.

**Consent for Publication:**

Participants provided informed consent before completing any study activities. Names and other directly identifying information were excluded from reported findings.

# Funding:

This study was not supported by any funding.


# Author Contribution

CE: Formal analysis, Investigation, Data Curation, Visualization, Writing - Original Draft, Writing - Review & Editing. MD: Formal analysis, Investigation, Data Curation, Visualization, Writing - Review & Editing. KA: Investigation, Writing - Original Draft. AL: Conceptualization, Methodology, Investigation, Writing - Review & Editing, SupervisionAll authors reviewed the final manuscript.


# Acknowledgements:

Not Applicable


# Data Availability

Deidentified survey data underlying this study are available from the corresponding author on reasonable request. To maintain participant confidentiality, interview transcripts cannot be publicly shared.

## References


1. Global Burden of Disease Cancer Collaboration. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: a systematic analysis for the Global Burden of Disease Study. JAMA Oncol. 2019;5(12):1749–68. 10.1001/jamaoncol.2019.2996.
2. Boyle K, Jones S. Functional outcomes of early laryngeal cancer—endoscopic laser surgery versus external beam radiotherapy: a systematic review. J Laryngol Otol. 2022;136(10):898–908. 10.1017/S0022215121002887.
3. Ozturk K, Turhal G, Durusoy D, Sahin E, Akagunduz O, Eyigor S, et al. Long-term swallowing outcomes of radiotherapy and transoral laser microsurgery for T1 glottic cancer treatment. Clin Otolaryngol. 2021;46(2):340–6.
4. Vasudevan SS, Bryan E, Ericksen E, Alla A, Asarkar AA, Olinde L, et al. Transoral laser microsurgery versus radiotherapy for T1-T2 glottic cancer with anterior commissure involvement: a systematic review and meta-analysis. Laryngoscope. 2025;135(6):1861–71.
5. Li MM, Zhao S, Eskander A, Rygalski C, Brock G, Parikh AS, et al. Stage migration and survival trends in laryngeal cancer. Ann Surg Oncol. 2021;28(12):7300–9.

6. Pfuetzenreiter EG Jr, Ferreron GF, Sadka JZ, Souza ABP, Matos LL, Kowalski LP, et al. Total laryngectomy vs. non-surgical organ preservation in advanced laryngeal cancer: a metanalysis. Braz J Otorhinolaryngol. 2024;90(3):101404.
7. Shephard EA, Parkinson MAL, Hamilton WT. Recognising laryngeal cancer in primary care: a large case-control study using electronic records. Br J Gen Pract. 2019;69:679.
8. Stachler RJ, Francis DO, Schwartz SR, Damask CC, Digoy GP, Krouse HJ et al. Clinical practice guideline: hoarseness (dysphonia) (update). Otolaryngol Head Neck Surg. 2018;158(1 Suppl).
9. Smith MM, Abrol A, Gardner GM. Assessing delays in laryngeal cancer treatment. Laryngoscope. 2016;126(7):1612–5.
10. Kompelli AR, Li H, Neskey DM. Impact of delay in treatment initiation on overall survival in laryngeal cancers. Otolaryngol Head Neck Surg. 2019;160(4):651–7.
11. Teppo H, Alho OP. Relative importance of diagnostic delays in different head and neck cancers. Clin Otolaryngol. 2008;33(4):325–30.
12. Díaz O, Rodríguez-Ruíz A, Sechopoulos I. Artificial intelligence for breast cancer detection: technology, challenges, and prospects. Eur J Radiol. 2024;175:111457.
13. Nartowt BJ, Hart GR, Muhammad W, Liang Y, Stark GF, Deng J. Robust machine learning for colorectal cancer risk prediction and stratification. Front Big Data. 2020;3:6.
14. Wu Q, Huang Y, Wang S, Qi L, Zhang Z, Hou D et al. Artificial intelligence in lung cancer screening: detection, classification, prediction, and prognosis. Cancer Med. 2024;13(7).
15. Wu T, Lucas E, Zhao F, Basu P, Qiao Y. Artificial intelligence strengthens cervical cancer screening—present and future. Cancer Biol Med. 2024;21(10):864–79.
16. Giavina-Bianchi M, de Sousa RM, Paciello VZA, Vitor WG, Okita AL, Prôa R et al. Implementation of artificial intelligence algorithms for melanoma screening in a primary care setting. PLoS ONE. 2021;16(9).
17. Ahluwalia VS, Schapira MM, Weissman GE, Parikh RB. Primary care provider preferences regarding artificial intelligence in point-of-care cancer screening. MDM Policy Pract. 2025;10(1):23814683251329007.
18. Hendrix N, Hauber B, Lee CI, Bansal A, Veenstra DL. Artificial intelligence in breast cancer screening: primary care provider preferences. J Am Med Inf Assoc. 2021;28(6):1117–24.
19. Kwon I, Wang SG, Shin SC, Cheon YI, Lee BJ, Lee JC, et al. Diagnosis of early glottic cancer using laryngeal image and voice based on ensemble learning of convolutional neural network classifiers. J Voice. 2025;39(1):245–57.
20. Marrero-Gonzalez AR, Diemer TJ, Nguyen SA, Camilon TJM, Meenan K, O'Rourke A. Application of artificial intelligence in laryngeal lesions: a systematic review and meta-analysis. Eur Arch Otorhinolaryngol. 2025;282(3):1543–55.
21. Kim HB, Song J, Park S, Lee YO. Classification of laryngeal diseases including laryngeal cancer, benign mucosal disease, and vocal cord paralysis by artificial intelligence using voice analysis. Sci

Rep. 2024;14:9297.

22. Wang CT, Chuang ZY, Hung CH, Tsao Y, Fang SH. Detection of glottic neoplasm based on voice signals using deep neural networks. IEEE Sens Lett. 2022;6(3):1–4.
23. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy. 2008;13(2):92–8.
24. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19(6):349–57.
25. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res. 2004;6(3).
26. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res. 2016;26(13):1753–60.
27. Braun V, Clarke V. Thematic analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA handbook of research methods in psychology. Vol. 2, Research designs: quantitative, qualitative, neuropsychological, and biological. Washington, DC: American Psychological Association; 2012. pp. 57–71.
28. Kim H, Jeon J, Han YJ, Joo Y, Lee J, Lee S, et al. Convolutional neural network classifies pathological voice change in laryngeal cancer with high accuracy. J Clin Med. 2020;9(11):3415.
29. Jenkins PD, Bedrick S, Karstens L, Hersh W, Bridge, 2AI-Voice Consortium, Dorr DA. From voice biomarkers to telemedicine screening: developing and evaluating a voice-based AI model for laryngeal lesion detection using the Bridge2AI-Voice dataset. Front Digit Health. 2026;8:1846369.
30. Bogdanffy R, Mundzic A, Nymberg P, Sundemo D, Moberg A, Wikberg C et al. Exploring clinician perspectives on artificial intelligence in primary care: qualitative systematic review and meta-synthesis. JMIR AI. 2026;5.
31. Derksen C, Le Noury KA, Akbar AB, Round T, Walter FM, Scott SE. Why aren't they used? Systematic review of barriers to implementation of clinical decision support systems for early cancer detection in primary care. Br J Gen Pract. 2026;76(765).
32. Goldberg DG, Soylu TG, Hoffman CF, Kishton RE, Cronholm PF. Clinicians' perspectives on the adoption and implementation of EMR-integrated clinical decision support tools in primary care. Digit Health. 2025;11:20552076251334043.
33. Allen MR, Webb S, Mandvi A, Frieden M, Tai-Seale M, Kallenberg G. Navigating the doctor-patient-AI relationship—a mixed-methods study of physician attitudes toward artificial intelligence in primary care. BMC Prim Care. 2024;25:42.
34. Barry B, Zhu X, Behnken E, Inselman J, Schaepe K, McCoy R et al. Provider perspectives on artificial intelligence-guided screening for low ejection fraction in primary care: qualitative study. JMIR AI. 2022;1(1).
35. Nash DM, Thorpe C, Brown JB, Kueper JK, Rayner J, Lizotte DJ, et al. Perceptions of artificial intelligence use in primary care: a qualitative study with providers and staff of Ontario community health centres. J Am Board Fam Med. 2023;36(2):221–8.

36. Darla M, Miltz D, Chandnani K, Purkayastha S, Diehl JW, Bhavani SV. Clinician trust and human factors in artificial intelligence-enabled clinical decision support in acute care: mixed methods study. JMIR Hum Factors. 08/09/2026:95472 (forthcoming/in press); https://preprints.jmir.org/preprint/95472
37. National Center for Health Workforce Analysis. State of the primary care workforce. 2025. Rockville, MD: Health Resources and Services Administration, Bureau of Health Workforce; 2025. https://bhw.hrsa.gov/sites/default/files/bureau-health-workforce/data-research/State-of-the-Primary-Care-Workforce-2025.pdf. Accessed 3 Aug 2026.

# Figures

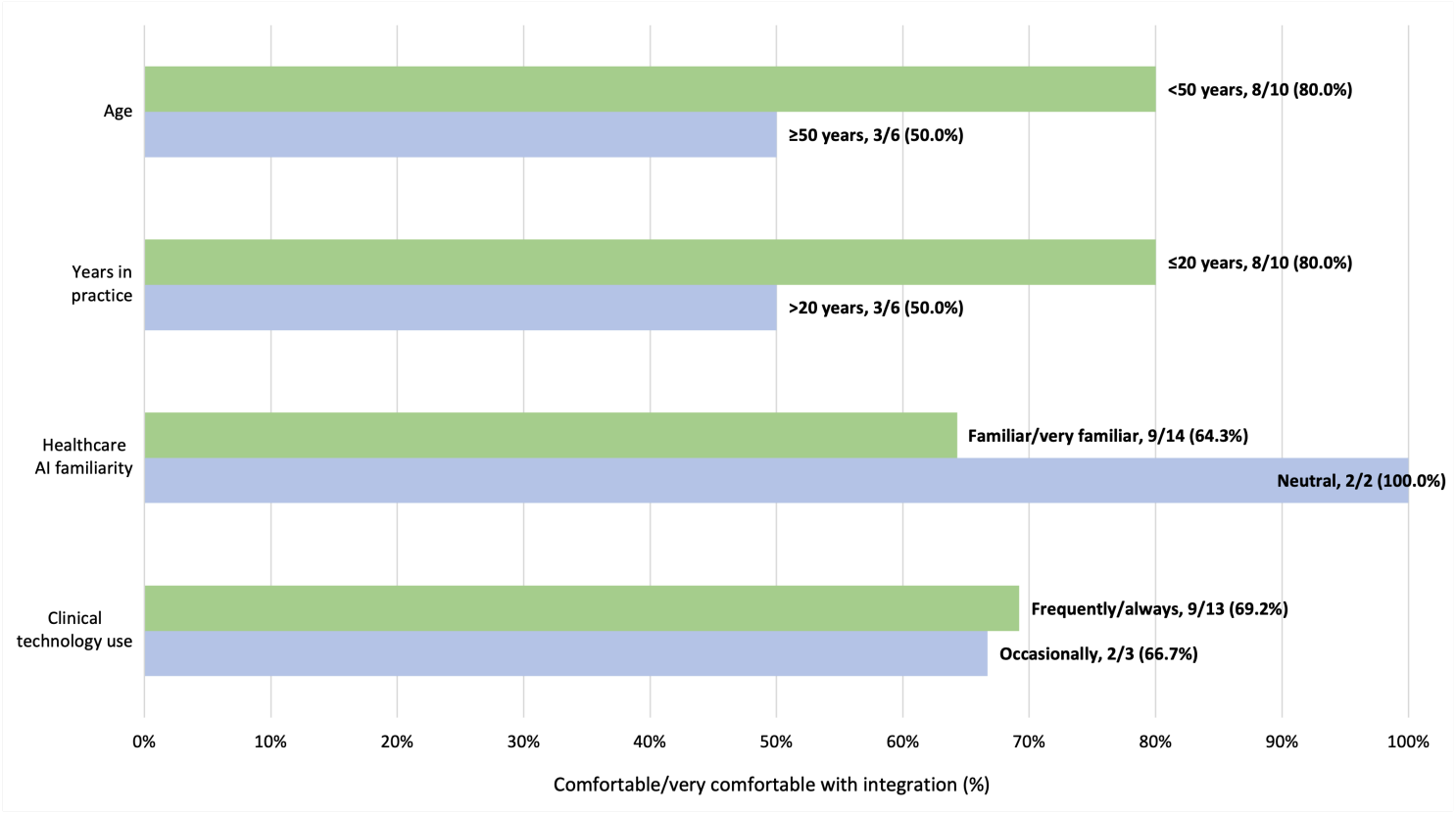
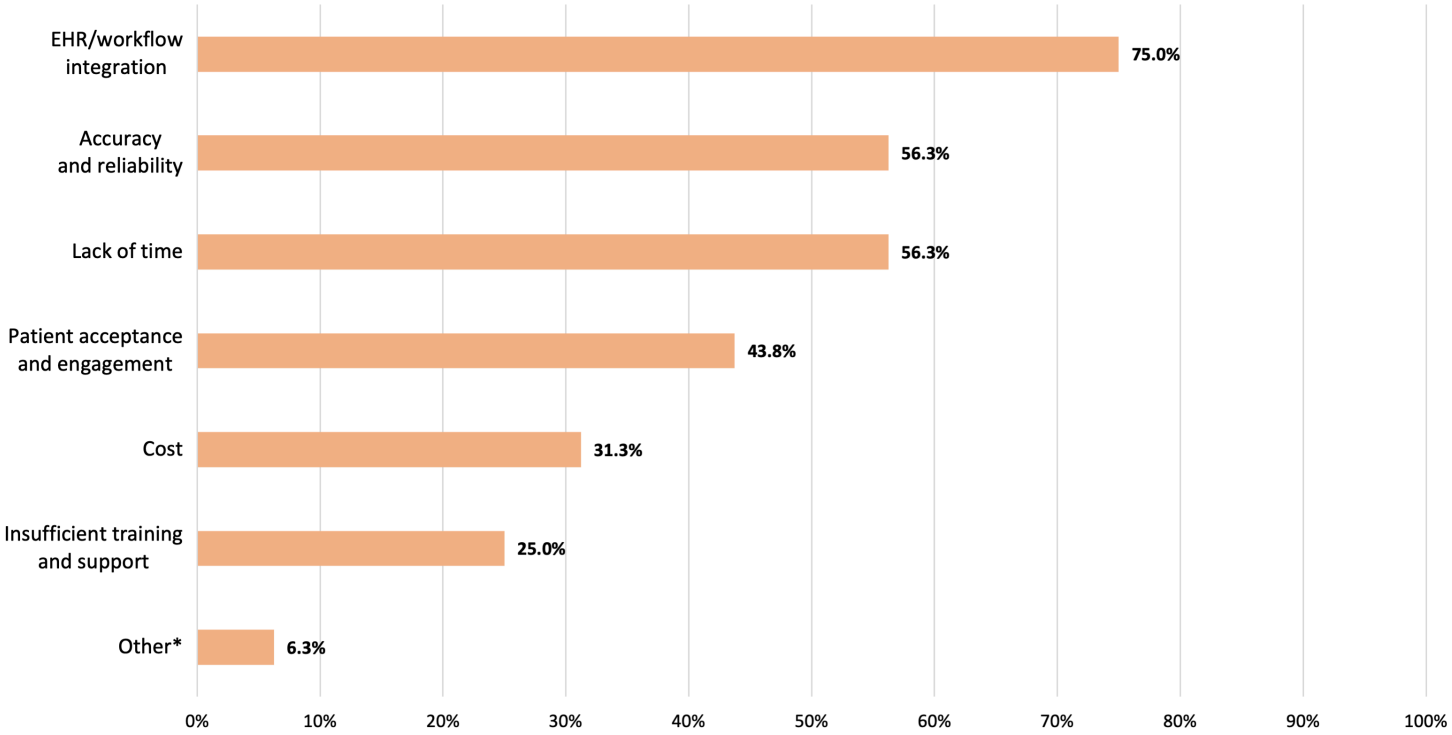



**Figure 1**

Exploratory subgroup analysis of comfort with integration by respondent characteristics.

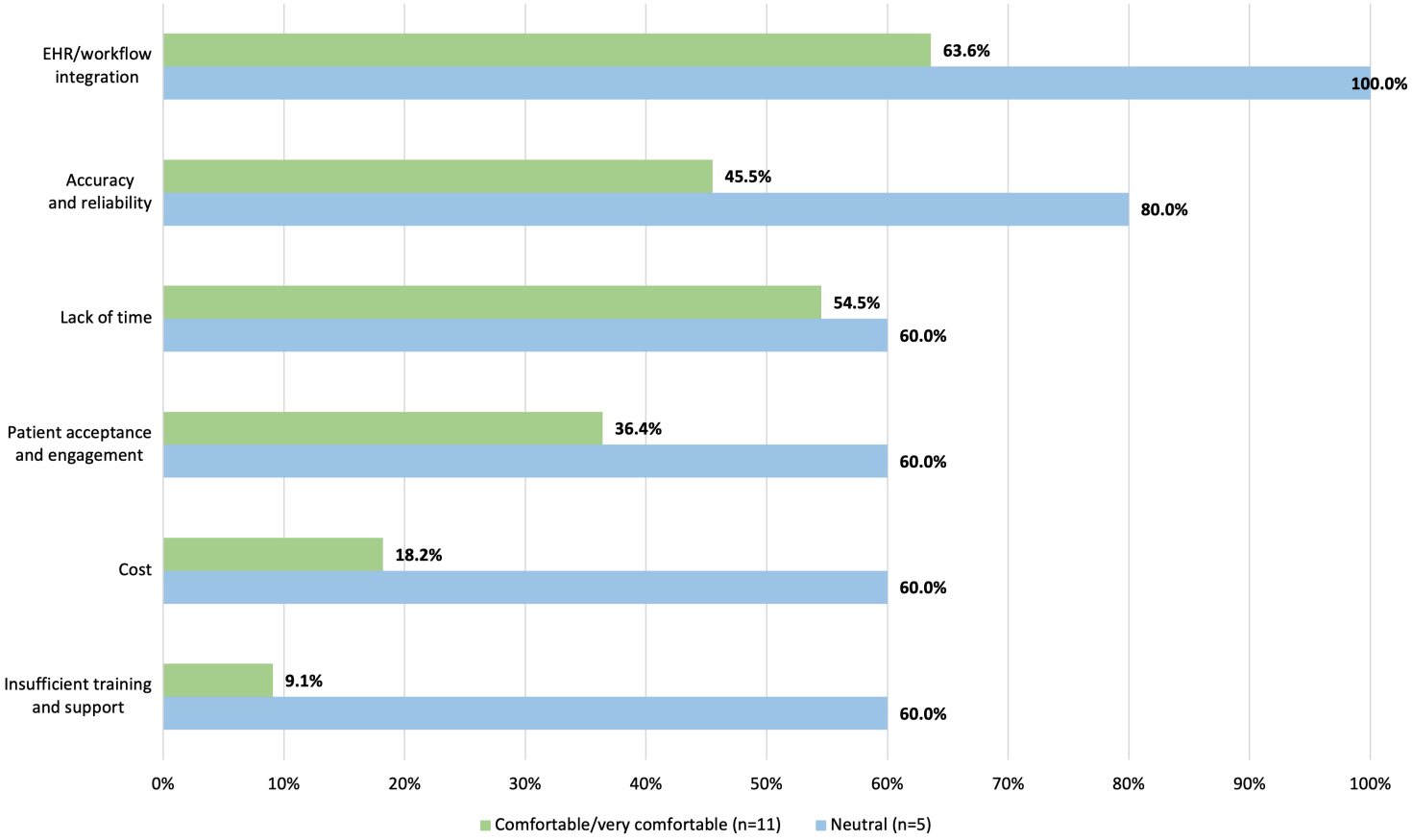
Caption: Bars represent the proportion of respondents within each clinician characteristic subgroup who reported being comfortable or very comfortable integrating the application into clinical practice.




## Figure 2

Perceived barriers to implementing the voice-based risk-stratification tool in primary care.

Caption: Participants could select more than one response; therefore, percentages do not total 100%. *Free-text response reported other competing needs, including screening for other body systems.




## Figure 3

Perceived implementation barriers by comfort with integration.

Caption: Percentages represent the proportion of respondents within each comfort group who selected each barrier. Participants could select more than one barrier; therefore, percentages do not total 100%.

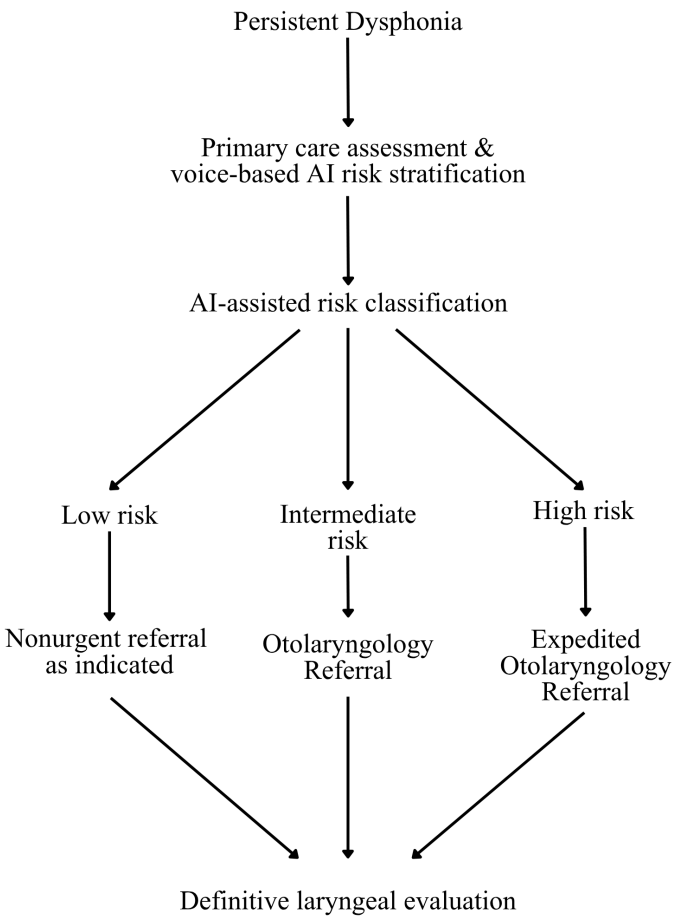



**Figure 4**

Proposed primary-care-to-otolaryngology pathway for AI-assisted risk stratification of persistent dysphonia.

Caption: The conceptual pathway illustrates how voice-based AI risk stratification could be incorporated into primary care in order to support appropriate referral urgency and connection to definitive laryngeal evaluation.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendixA.GoodReportingofaMixedMethodsStudyGRAMMS.docx