



THE 42ND ANNUAL

**BAY
AREA
RESIDENTS'
RESEARCH
SYMPOSIUM**

IN OTOLARYNGOLOGY--
HEAD & NECK SURGERY

BROCHURE

KAISER PERMANENTE
OAKLAND, CALIFORNIA

MAY 8, 2026

DISTINGUISHED FACULTY



Natalie Krane, MD

**ASSISTANT PROFESSOR IN THE DEPARTMENT OF OTOLARYNGOLOGY
AT OREGON HEALTH & SCIENCE UNIVERSITY**

KEYNOTE SPEAKER & JUDGE

Dr. Natalie Krane is a double board-certified facial plastic and reconstructive surgeon and Assistant Professor at Oregon Health & Science University in Portland, Oregon. Her clinical practice primarily focuses on rhinoplasty and complex facial reconstruction. Her grant-funded research focuses on the impact of mindfulness-based interventions on perioperative outcome measures, such as anxiety, sleep quality, pain, and narcotic use. She has also delivered invited lectures in the U.S. and internationally on physician health and wellness, focusing on performance and recovery optimization.

MODERATOR



NIKOLAS BLOCK-WHEELER, MD
THE BLOCK-WHEELER CLINIC
SAN FRANCISCO, CA

FACULTY JUDGES



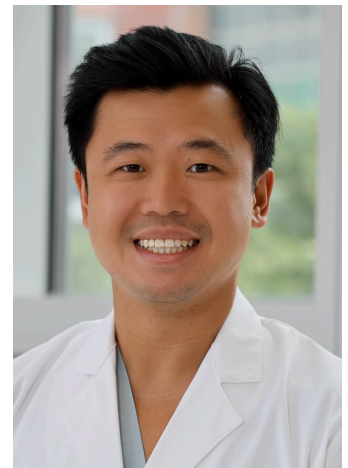
CASSANDRA PUCCINELLI, MD
ENDOCRINE SURGERY
KAISER PERMANENTE



NNEOMA WAMKPAH, MD
FACIAL PLASTICS
STANFORD



KATHY WAI, MD
HEAD & NECK ONCOLOGY
UCSF



MICHAEL LI, MD
HEAD & NECK ONCOLOGY
UC DAVIS

ORGANIZING COMMITTEE

JONATHAN LIANG, MD, MPH, CHAIR, KAISER PERMANENTE OAKLAND
JOLIE CHANG, MD, UNIVERSITY OF CALIFORNIA SAN FRANCISCO
BENJAMIN MALKIN, MD, KAISER PERMANENTE OAKLAND
KARA MEISTER, MD, STANFORD UNIVERSITY
TOBY STEELE, MD, UNIVERSITY OF CALIFORNIA DAVIS
KEVIN WANG, MD, KAISER PERMANENTE OAKLAND
NORIKO YOSHIKAWA, MD, KAISER PERMANENTE OAKLAND

OVERVIEW

THE BAY AREA RESIDENTS' RESEARCH SYMPOSIUM (BARRS) PROVIDES A UNIQUE FORUM FOR OTOLARYNGOLOGY RESIDENTS TO PRESENT THEIR ORIGINAL RESEARCH PROJECTS. A PANEL OF RESPECTED OTOLARYNGOLOGISTS AND THE AUDIENCE OFFER THE RESIDENTS CONSTRUCTIVE CRITICISM ON THE DESIGN, VALIDITY, AND PRESENTATION OF THEIR RESEARCH. CLINICIANS, RESEARCHERS, AND RESIDENTS BENEFIT FROM THE INTERACTIVE DISCUSSION BASED ON EACH PRESENTED TOPIC.

SYMPOSIUM OBJECTIVES

BY THE CONCLUSION OF THIS SYMPOSIUM, ATTENDEES WILL BE ABLE TO:

1. CRITIQUE THE SCIENTIFIC VALIDITY OF THE PRESENTATIONS.
2. SUMMARIZE THE ELEMENTS OF A CLEAR, CONCISE SCIENTIFIC PRESENTATION.
3. DESCRIBE AT LEAST THREE AREAS OF CUTTING-EDGE RESEARCH DEVELOPMENT IN OUR SPECIALTY.

SYMPOSIUM OUTLINE

8:30 - REGISTRATION

9:10 - WELCOME

BARRS CHAIR: JONATHAN LIANG, MD, MPH

9:15 - MORNING INTRODUCTION

MODERATOR: PETER DEBBANEH, MD

Session A - Rhinology, Allergy, & Skull Base Surgery

9:20 - **HENRY ZHENG, MD (KP)**

OUTCOMES OF DUPILUMAB IN SURGICALLY REFRACTORY CHRONIC RHINOSINUSITIS WITH NASAL POLYPS: A MATCHED COHORT STUDY

Background: Dupilumab is an effective biologic therapy for chronic rhinosinusitis with nasal polyps (CRSwNP), but its real-world impact on healthcare utilization following endoscopic sinus surgery (ESS) remains incompletely characterized. This study evaluated the association between dupilumab therapy and revision surgery, systemic antibiotic use, and systemic corticosteroid use in a large healthcare system.

Methods: We conducted a retrospective matched cohort study of adult patients with eosinophilic CRSwNP who underwent ESS between 2017 and 2023. Patients receiving dupilumab for at least 4 months were matched 1:3 to CRSwNP controls based on age category, asthma comorbidity, and Neighborhood Deprivation Index. Cox proportional hazards models assessed the risk of revision ESS. Multivariable logistic regression and negative binomial regression evaluated one-year systemic antibiotic and corticosteroid use.

Results: The cohort included 644 patients (161 dupilumab-treated and 483 controls) with a median follow-up of 36 months. Revision ESS occurred in 25 patients. Dupilumab therapy was associated with a significantly lower hazard of revision surgery (adjusted hazard ratio 0.31, 95% CI 0.11–0.71). Dupilumab was not associated with reduced systemic antibiotic use (adjusted one-year odds ratio 1.54, 95% CI 0.96–2.47; incidence rate ratio 1.14, 95% CI 0.79–1.65). Systemic corticosteroid use was numerically lower but not statistically significant (adjusted odds ratio 0.85, 95% CI 0.48–1.48; incidence rate ratio 0.70, 95% CI 0.45–1.09).

Discussion: In real-world practice, dupilumab use following ESS was associated with a substantially reduced risk of revision surgery but did not significantly reduce systemic antibiotic or corticosteroid utilization.

Conclusion: Dupilumab was associated with a 69% reduction in revision ESS hazard. Dupilumab was not associated with reductions in antibiotic use, and further exploration is needed to examine its effect on corticosteroid use. These findings support dupilumab's real-world clinical value and impact on healthcare utilization.

SYMPOSIUM OUTLINE

9:28 - **ANUJ PATEL, MD (UC DAVIS)**

EFFECT OF PRE-OPERATIVE APREPITANT ON POST-OPERATIVE NAUSEA AND VOMITING IN SINONASAL SURGERY

Background: Post-operative nausea and vomiting (PONV) is common following endoscopic sinonasal surgery (ESS) despite routine multimodal antiemetic prophylaxis. Aprepitant, a neurokinin-1 (NK-1) receptor antagonist, reduces PONV in several surgical populations, including endoscopic trans-sphenoidal surgery. However, the prophylactic role of aprepitant in non-skull base sinonasal surgery is unclear.

Methods: Patients (n=176) undergoing endoscopic sinonasal surgery were retrospectively enrolled. Demographics, anesthetic technique, co-administration antiemetics, and relevant clinical variables were collected. The primary outcome was a composite PONV endpoint, which was positive if rescue antiemetics were used or if there was documented nausea or vomiting in PACU or in post-operative phone calls. Secondary outcomes included PACU length of stay and post-operative bleeding requiring emergency department presentation. Multivariate regression controlling for established PONV risk factors and matched cohort analyses were performed.

Results: Seventy patients (39.8%) received pre-operative aprepitant. Overall, PONV incidence was similar between aprepitant and non-aprepitant groups (18.6% vs 19.9%, OR 0.98, p = 1.00). On multivariate analysis, aprepitant was not associated with reduced odds of PONV (OR 0.90, p = 0.81). Age was independently associated with PONV (OR 0.97, p = 0.02). There were no significant differences between groups in vomiting (OR 0.8, p = 0.86), post-operative bleeding, or PACU length of stay. Findings were consistent in matched cohort analysis.

Conclusion: Pre-operative aprepitant was not associated with reduced PONV following non-skull base ESS. NK-1 receptor antagonists may have limited benefit in this population, though prospective studies are needed to further elucidate their role alongside multimodal prophylaxis.

9:36 - **MAIMUNA AHAMD, MD (UCSF)**

MULBERRY-LIKE CHANGES OF THE POSTERIOR INFERIOR TURBINATE ARE ASSOCIATED WITH DUST MITE ALLERGY

Objective: To evaluate the association between dust mite allergy and mulberry-like changes of the posterior inferior turbinate (MPINT), and to assess the relationship between MPINT and sinonasal symptom burden.

Study Design: Retrospective cohort study.

Setting: Tertiary care center.

Methods: Adult patients who underwent dust mite allergen testing and nasal endoscopy were included. Patients without adequate visualization of the posterior inferior turbinate or without pre-treatment endoscopy were excluded. The primary exposure was dust mite allergy. The primary outcome was the presence of MPINT on nasal endoscopy. Secondary analyses evaluated associations with asthma, gastroesophageal reflux disease (GERD), and pre-treatment Sinonasal Outcome Test-22 (SNOT-22) scores. Associations were assessed using Fisher's exact test and multivariable logistic and linear regression.

Results: Of 143 patient records, 89 met inclusion criteria, including 41 (46.1%) with MPINT. Dust mite allergy was more prevalent among patients with MPINT and was associated with increased odds of MPINT on univariate analysis (OR 3.63, 95% CI 1.40–9.91; $p = 0.005$). This association remained significant after adjustment for age, sex, asthma, and GERD (adjusted OR 3.37, 95% CI 1.31–9.08). Asthma was also associated with MPINT (OR 3.52, 95% CI 1.14–12.11), while GERD was not. On multivariable linear regression, MPINT was not associated with pre-treatment SNOT-22 scores ($\beta = 2.64$, $p = 0.59$), and no other variables were significantly associated with symptom burden.

Conclusions: Dust mite allergy is associated with mulberry turbinate morphology, supporting a potential role for allergic or atopic inflammation. MPINT was not associated with increased symptom burden, suggesting it may represent an endoscopic marker of underlying disease rather than a clinically symptomatic entity.

9:44 - **LISA CHIONIS BUHLER, MD (KP)**

FACTORS ASSOCIATED WITH DUPILUMAB USE IN SURGICALLY TREATED CHRONIC RHINOSINUSITIS WITH NASAL POLYPS

Introduction: Dupilumab is an effective biologic for chronic rhinosinusitis with nasal polyps (CRSwNP), but its high cost and insurance authorization requirements may create barriers to access. Data on sociodemographic and clinical differences in dupilumab utilization among surgically treated patients remain limited.

Methods: We performed a retrospective cohort study of 6,387 adults with CRSwNP who underwent sinus surgery from 2019 to 2024. Patients were stratified by dupilumab prescription status, and univariate comparisons were performed across demographic, insurance, neighborhood deprivation, geographic, and clinical variables.

Results: Overall, 479 patients (7.5%) received dupilumab. Age differed by treatment status ($P < .01$); dupilumab-treated patients had a lower median age (48.1 vs 54.5 years). Race/ethnicity differed ($P < .01$) with Black patients comprising 13.4% of the dupilumab group vs 6.7% of the non-dupilumab group, while White patients comprised 53.0% vs 59.7%. Insurance also differed ($P < .01$), with higher commercial coverage among dupilumab-treated patients (81.0% vs 70.9%). Geography differed by treatment status ($P < .01$): dupilumab-treated patients were more often from Greater San Francisco (18.8%) and East Bay (13.8%) and less often from Central Valley (0.4%). Clinically, dupilumab-treated patients were more likely to have asthma (76.2% treated vs 37.7% untreated), aspirin sensitivity (1.7% vs 0.3%), aspirin/NSAID allergy (6.9% vs 2.3%), and more prior sinus surgeries (mean 1.6 vs 1.2; all $P < .01$).

Discussion: Dupilumab receipt may reflect both greater type 2 inflammatory disease burden and variation in insurance, race/ethnicity, and geography. Further adjusted analyses are needed to clarify whether these differences are driven by clinical indication, access-related factors, or both.

Conclusions: Dupilumab use was uncommon in this surgical CRSwNP cohort and was associated with differences in demographic, insurance, geographic, and clinical factors. Additional multivariable analyses are needed to determine whether these patterns reflect differences in disease severity, treatment access, or both.

9:52 - DISCUSSION / Q&A (15 MIN)

10:07 - **BREAK (15 MIN)**

Session B - Head & Neck Oncology

10:23 - **THOMAS HAUPT, MD (KP)**

SURGEON-PERFORMED POINT-OF-CARE ULTRASOUND-GUIDED VERSUS PALPATION-GUIDED CORE NEEDLE BIOPSY FOR CERVICAL LYMPHADENOPATHY: A MULTICENTER COHORT STUDY

Importance: Non-diagnostic cervical lymph node core needle biopsy (CNB) delays cancer diagnosis and increases patient burden. Whether point-of-care ultrasound (POCUS) guidance by the operating surgeon improves outcomes over a palpation-guided technique has not been established.

Objective: To compare diagnostic success, repeat biopsy rates, and time to diagnosis between POCUS-guided and palpation-guided surgeon-performed cervical CNB.

Design, Setting, and Participants: Retrospective cohort study of 545 cervical CNBs performed by otolaryngology attendings across 27 Kaiser Permanente Northern California facilities from January 2016 through December 2020. Interventional radiology-performed and fine-needle aspiration biopsies were excluded.

Exposures: POCUS-guided (n=126) versus palpation-guided (n=419) surgeon-performed CNB. **Main Outcomes and Measures:** Primary outcome: successful biopsy rate. Secondary outcomes: non-diagnostic rate, repeat biopsy rate, time to diagnosis, and diagnostic accuracy for cancer detection.

Results: POCUS guidance was associated with higher successful biopsy rates (88.1% vs 69.7%; OR, 3.22; 95% CI, 1.81–5.74; $P<.001$) and lower repeat biopsy rates (11.1% vs 27.4%; OR, 0.33; $P<.001$). On multivariable analysis, POCUS guidance was the strongest independent predictor of success (aOR, 7.98; 95% CI, 3.31–19.22; $P<.001$). Benefit was concentrated in nodes 4 cm or smaller (91% with guidance vs 49% to 64% without; $P<.001$). No benefit was observed for larger nodes, where necrosis was more prevalent (48% vs 32%; $P=.02$). For nodes under 4 cm, POCUS-guided sensitivity for cancer was 93.9% versus 68.3% with palpation, and negative predictive value was 91.4% versus 67.0%. Among 129 patients requiring repeat biopsy, 51 (40%) underwent excisional or open surgical biopsy. POCUS adoption increased from 16.8% to 33.3% over the study period (P -trend $<.001$).

Conclusion: In this multicenter surgeon-performed cohort, POCUS-guided cervical CNB was the dominant predictor of biopsy success, with a nearly 8-fold improvement in odds of diagnostic yield and the greatest benefit for nodes 4 cm or smaller. For larger nodes, where higher rates of necrosis render failure pathologic rather than technical, alternative strategies such as POCUS-guided sampling of a smaller satellite node warrant further investigation. These findings support POCUS-guided cervical CNB as the preferred technique and a core procedural competency for head and neck surgeons, with direct implications for training and quality benchmarking.

10:31 - ALYSSA CIVANTOS, MD (UCSF)

INCORPORATING POST-RADIATION PREVENTIVE FLUORIDE INTO HEAD AND NECK CANCER SURVEILLANCE IN A SAFETY-NET OTOLARYNGOLOGY CLINIC

Introduction: Despite an established relationship between head and neck cancer (HNC) survivorship and oral health, adherence to recommended oral care is challenging, particularly for underserved patients. This study aimed to integrate preventative fluoride treatments into routine cancer surveillance in a public safety-net hospital clinic.

Methods: This is a prospective cohort study with a hybrid type III design, including HNC survivors who completed radiation. Patients were recruited during their clinic visits over one year, completing an oral health quality of life survey (OHIP-14), salivary pH samples, and oral photographs. Fluoride varnish treatments were then administered. An Acceptability, Appropriateness, and Feasibility Intervention Measures questionnaire was provided to patients and clinic providers. Descriptive analyses summarized baseline implementation responses. Repeated measures were analyzed using linear mixed-effects models with estimated marginal means (EMM) for time-point comparisons for oral health outcomes.

Results: Forty-one patients completed the baseline visit, with 63% male and a median age of 59 years (IQR 48-64). Eighteen patients had a first follow-up visit, and nine had a second. Amongst participants, >89% agreed that the intervention was feasible, acceptable, and appropriate, with high levels sustained at follow-up. Amongst the 16 providers, >81% agreed that it was feasible, appropriate, and acceptable. Baseline EMM for salivary pH was 6.8 (95% CI 6.6-7.0) and for OHIP-14 scores was 12.8 (95% CI 8.6-17.1), with no significant differences across follow-up visits. Baseline EMM for DMFT scores was 13.8 (95% CI 11.5-16.1), with a statistically significant increase at follow-up (mean difference -0.39, $p=0.025$).

Conclusions: Implementation of fluoride treatments into an otolaryngology clinic is feasible and acceptable to both patients and providers, with stable qualitative and quantitative outcomes during one year. This study identifies a novel, interdisciplinary approach to enhancing equitable access to care and could be adapted to broader practice settings.

10:39 - SUHKARAN SINGH AULAKH, MD (UC DAVIS)

CHEMOTHERAPY AND RADIATION DOSE CORRELATE WITH HYPOTHYROIDISM FOLLOWING RADIATION THERAPY FOR HEAD AND NECK CANCERS

Purpose/Objectives: Hypothyroidism is a common late complication of head and neck radiotherapy. This study aimed to identify clinical and treatment-related risk factors associated with hypothyroidism following intensity-modulated radiation therapy (IMRT).

Materials/Methods: We conducted an IRB-approved retrospective review of patients with locally advanced head and neck cancer treated with radiotherapy between January 2015 and March 2021. Hypothyroidism was defined as initiation of thyroid hormone replacement or a TSH level greater than twice the upper limit of normal. Baseline characteristics (age, gender, tumor site, stage, histology, thyroid volume) and treatment factors (radiation dose, chemotherapy, surgery) were analyzed. Statistical significance was assessed using the Pearson chi-square test.

Results: Of 491 identified patients, 275 met inclusion criteria. Median follow-up was 25.4 months; 77.5% were male, with a mean age of 63 years. Hypothyroidism developed in 38.2% of patients at a median of 13.9 months post-treatment. No significant associations were found with age, gender, histology, or surgery. Higher tumor stage correlated with increased risk ($p = 0.001$). Smaller thyroid volume ($<13 \text{ cm}^3$) was associated with a higher incidence (51.2% vs. 27.2%). Tumor site was significant ($p = 0.007$), with higher rates in oropharyngeal and laryngeal cancers. Mean thyroid radiation dose strongly correlated with hypothyroidism: 45.5% incidence with $>40 \text{ Gy}$ vs. 20.5% with $<40 \text{ Gy}$ ($p < 0.001$). Chemotherapy was also associated ($p = 0.005$), though no specific agent or timing was significant.

Conclusion: Mean thyroid radiation dose and chemotherapy use are associated with increased risk of hypothyroidism following IMRT. Further dosimetric and multivariate analyses are underway to refine risk prediction and optimize screening strategies.

10:47 - SHANNON WU, MD (STANFORD)

PROTECTIVE AND EXACERBATING FACTORS FOR FINANCIAL HARDSHIP IN HEAD AND NECK CANCER

Introduction: Financial hardship presents a devastating challenge for patients with cancer. However, financial hardship in head and neck cancer (HNC) is not well understood nor routinely assessed. This study explored the experiences regarding financial burden in HNC and identified potential gaps from the existing financial hardship screening tool “COMprehensive Score for financial Toxicity” (COST).

Methods: This multi-institutional, prospective, qualitative study recruited 24 patients between two hospital sites: a tertiary-care academic center and a county safety-net hospital. Qualitative interviews were conducted among patients who completed HNC treatment. Thematic analysis was performed to develop themes relating to financial burden in HNC. Themes were evaluated to identify missing gaps in COST.

Results: Twenty-four patients (median 66 years [range, 29-77], 83% male, 29% non-English speakers) were interviewed. Tumor sites included the oral cavity, larynx/hypopharynx, oropharynx, nasopharynx, sinonasal, cutaneous, thyroid, and parotid. Social support, stable baseline financial status, and insurance coverage helped protect against financial hardship. Conversely, unstable baseline financial status, difficulty accessing resources, inability to return to work, high out-of-pocket costs, and transportation barriers intensified financial burden. Although COST captured some elements of financial burden, key domains pertinent to HNC were missed, such as unknown out-of-pocket costs that emerged only after treatment, transportation burdens, caregiver sacrifices, and mismatched expectations about timing and feasibility of return to employment.

Conclusion: Themes of financial challenges relating to HNC treatment emerged, revealing gaps not captured by the existing COST tool. These findings may inform methods to more accurately identify and measure financial hardship unique to the HNC patient population.

10:55 - MONICA BODD, MD (STANDFORD)

“I HAD NO CHOICE”: A MIXED METHODS ANALYSIS OF PATIENT DECISION-MAKING FOR TOTAL LARYNGECTOMY

Chemoradiation and total laryngectomy (TL) offer comparable survival for advanced laryngeal cancer but differ markedly in long-term quality-of-life (QOL) outcomes. Patient decision-making between these options is poorly understood. Understanding the experiences of TL survivors may improve counseling for patients facing this decision. This mixed-methods study describes patient values, expectations, and treatment preferences for advanced laryngeal cancer and their relationship to prior healthcare behavior, post-treatment QOL, and decision-related distress.

We conducted a mixed-methods study from January-October 2025 at a single academic center. Patients with advanced laryngeal cancer who underwent primary or salvage TL were recruited. Using a convergent parallel design, participants completed surveys and semi-structured interviews. Surveys included validated measures of maximizer–minimizer treatment preferences, QOL, decision satisfaction, and decisional regret. Interviews were analyzed using a two-rater iterative interpretive approach for thematic description ($k=0.96$). Data analysis occurred from September-October 2025, with frequent team meetings and member checking (author GG) to ensure rigor.

Fifteen patients completed surveys (15 [100%] men; median age, 75 years; 5 [33.3%] primary TL; median 5.34 years post-TL; 15 [100%] tolerating oral intake), and ten completed interviews. Participants exhibited wide variation in maximizer–minimizer preferences without differences in healthcare utilization. Overall QOL was favorable (median=6; 53% rating 6–7/7), with high decision satisfaction (median=4/5) and low decisional regret (median=20/100). Nonetheless, participants reported substantial social limitations related to voice loss, eating in public, and stoma care. TL was commonly perceived as the only alternative to death after failed prior treatments, with cancer-free survival prioritized over voice or swallowing. Decision-making was strongly shaped by trust in surgeons and the treating institution.

TL survivors reported good adjustment and minimal decisional regret but emphasized social loss, constrained choice, and reliance on clinician guidance, highlighting the importance of eliciting patient values before recommending TL.

11:03 - DISCUSSION / Q&A (15 MIN)

11:20 - **KEYNOTE SPEAKER**
NATALIE KRANE, MD - THE AUGMENTED SURGEON (AI IN OTOLARYNGOLOGY)

12:20 - LUNCH / BREAK / PHOTOS

1:40 - AFTERNOON INTRODUCTION
MODERATOR: NIKOLAS BLOCK-WHEELER, MD

Session C - Facial Plastics Reconstructive Surgery & General

1:45 - **ALICE SU, MD (UC DAVIS)**

FACTORS ASSOCIATED WITH PATIENT SATISFACTION FOLLOWING MANDIBULAR FRACTURE MANAGEMENT

Introduction: Mandible fractures may have severe functional and psychosocial sequelae, but there is a paucity of patient-reported outcome data to understand these symptoms. This study was designed to characterize factors influencing levels of satisfaction in this population as measured by responses from the validated Integrated Modular Patient-Reported Outcome Assessment for Craniomaxillofacial Trauma (IMPACT).

Methods: A prospective, multi-institutional survey study was performed on patients with mandible fractures presenting for follow-up at two level 1 trauma centers in Sacramento, CA, and Memphis, TN between June and November 2025. Patients completed the initial general survey module (IMPACT-G) at 1-2 week follow-up. The primary outcome was patient happiness, categorized as “happy” or “unhappy”. Independent variables included sex, age, race, tobacco usage, insurance status, fracture patterns (unilateral or bilateral), type of management, time from injury to intervention, and individual IMPACT-G item scores. Effect sizes are reported as odds ratios, correlation coefficients, and mean difference in item scores with 95% confidence intervals (CI). Significance was set at $p < 0.05$.

Results: 35 patients were included (median age = 35 years [range: 18 – 84], male = 30 [85.7%]). Most underwent surgical repair (maxillomandibular fixation alone = 6 [17.1%], open reduction internal fixation alone = 9 [25.7%], or both = 17 [48.6%]) while 3 (8.6%) were observed. There were no significant associations between happiness and sociodemographic factors, fracture patterns, or treatment ($p > 0.05$) at primary follow-up. IMPACT-G mean scores were significantly worse in unhappy patients compared to happy patients for facial appearance (-0.9 [95% CI = -1.6 to -0.3], $p = 0.005$), cuts or scars (-1.0 [95% CI = -1.7 to -0.2], $p = 0.012$), and facial movement (-1.3 [95% CI = -2.0 to -0.5], $p = 0.002$).

Conclusion: According to results from the IMPACT-G survey, overall facial appearance, facial movement, and cuts or scars demonstrate an association with initial levels of satisfaction as measured from responses to an overall happiness question following mandibular fracture management in adult patients.

1:53 - **BRYAN LE, MD (KP)**

ASSESSING THE EFFECT OF BOTULINUMTOXIN A INJECTIONS ON PERCEIVED AGE USING AMAZON RECOGNITION SOFTWARE

Background: Artificial intelligence (AI) based facial analysis tools, such as Amazon Rekognition, are increasingly used in aesthetic medicine to provide objective assessments. However, it remains unclear whether these tools align with human perception or capture the qualitative effects of aesthetic interventions. This study compares AI-based and human-perceived age estimates before and after Botox injection and explores whether aesthetic outcomes are better understood as qualitative rather than quantitative changes.

Methods/Design: This study was designed as a prospective observational comparative study. Standardized, publicly available pre- and post-Botox photographs from 23 individuals were utilized for the analysis. These photographs were first processed using Amazon Rekognition to generate an AI-predicted age for each subject. Subsequently, ten human evaluators at a single center, consisting of five medical students and five ENT residents, assessed the same photographs for perceived age, attractiveness, and youthfulness. To evaluate the data, median scores before and after Botox were compared using the Wilcoxon signed-rank test, and the correlation between the AI and human age predictions was assessed using Spearman's rho.

Results: Amazon Rekognition did not detect a statistically significant change in predicted age ($p = 0.36$). Similarly, human evaluators found no significant differences in perceived age ($p = 0.23$) or attractiveness ($p = 0.11$); however, they reported a significant increase in perceived youthfulness ($p = 0.04$). Correlation between AI and human-perceived age estimates was strong (Spearman's $\rho = 0.7$).

Discussion: These preliminary findings highlight a critical distinction between objective metrics and subjective aesthetic success. While AI demonstrates quantitative accuracy in estimating age—evidenced by its strong correlation with human estimates—it remains limited to objective measures. Human evaluators were able to perceive subtle, qualitative improvements in youthfulness following Botox that the AI could not detect. This suggests that aesthetic outcomes are inherently qualitative in nature, and current AI tools may not fully capture the complete picture of aesthetic improvement despite their technical accuracy.

Conclusion: These preliminary findings suggest that while human evaluators perceive subtle qualitative improvements following Botox, AI analysis remains limited to objective measures and may not reflect the inherently qualitative nature of aesthetic outcomes despite its quantitative accuracy. Future studies will aim to increase power and incorporate evaluations from facial plastic experts further along in their training course.

2:01 - NICHOLAS DEL MUNDO, MD (KP)

DO GLP-1 RECEPTOR AGONISTS INCREASE THYROID CANCER RISK? REASSURING EVIDENCE FROM OVER TWO MILLION PATIENTS

Introduction: Our goal is to investigate if the use of glucagon-like peptide 1 receptor agonists (GLP-1 RA) is associated with an increased risk of thyroid cancer.

Methods: This is a retrospective, cohort study with a nested case-control study of adult patients (18 years and older) within Kaiser Permanente Northern California (KPNC) between 1/1/2018 and 12/31/2022. The cohort included patients with at least 1 year of continuous membership before and after the index date who did not have a previous history of GLP-1 RA use, thyroid cancer, or thyroid surgery. These patients were then categorized into exposure groups: GLP-1 RA users and non-GLP-1 RA users. Thyroid cancers were identified through the KPNC Cancer Registry. Demographic information, relevant clinical covariates, and specific GLP-1 RA medication use were collected. In the cohort study, incidence rate and risk for each exposure group were calculated. For the nested case-control study, patients who have thyroid cancer will be matched to non-cancer patients from the cohort on characteristics to determine the proportions of GLP-1 use in the cases and controls.

Results: Our study cohort consisted of 2,500,185 non-GLP-1 RA users and 14,057 GLP-1 RA users. The mean age of the study cohort was 49.2 years, and 52.7% of the cohort were female. The most used GLP-1 RA medications were liraglutide (n=9,515) and semaglutide (n=3,529). The incidence of thyroid cancer in the GLP1-RA users was 10 per 100,000 patients per year, and the incidence amongst non-GLP1-RA users was 13.2 per 100,000 patients per year. There was no significant association of incident thyroid cancer in GLP-1 RA users compared to non-GLP-1 RA users (OR 0.75; p=0.46).

Conclusions: In this large retrospective cohort study within KPNC, GLP-1 RA use was not associated with an increased risk of incident thyroid cancer. Further studies may help clarify associations with specific GLP-1 RA medications or thyroid cancer subtypes.

2:09 - TAYLOR S. ERICKSON, MD (UCSF)

TOWARD PRECISION SURGICAL EDUCATION: A BAYESIAN FRAMEWORK FOR ESTIMATING OPERATIVE COMPETENCE IN OTOLARYNGOLOGY TRAINEES

Background: Competency-based medical education requires robust evidence that surgical trainees have achieved defined proficiency thresholds before advancing. Workplace-based assessments (WBAs) collected via platforms such as SIMPL have increased the volume of trainee performance data available, yet standard approaches to summarizing ratings fail to account for procedure difficulty, rater variability, or the dynamic nature of developing competence.

Objective: To develop and evaluate a Bayesian measurement framework for estimating the operative competence of otolaryngology (OHNS) trainees from workplace-based assessments.

Methods Assessment data were obtained from the Society for Improving Medical Professional Learning (SIMPL) platform for OHNS trainees (PGY 2–5) rated by 156 attending surgeons across 14 key indicator procedures. A Bayesian cumulative ordinal item response theory model was fit to calibrate procedure-specific difficulty and discrimination parameters. Trainee proficiency was then estimated using a sequential Bayesian updating pipeline with exponential decay of older evidence. Model performance was evaluated on a held-out test set using accuracy, Cohen's kappa, and area under the receiver operating characteristic curve.

Results: The dataset included 2,241 observations from 142 trainees rated by 156 attending surgeons. The predicting Practice-Ready ratings had strong overall discrimination (area under the receiver operating characteristic curve, 0.765). Predicted probabilities of achieving a Practice-Ready rating varied substantially across the 14 surgical EPAs at equivalent proficiency levels.

Conclusions: A Bayesian measurement framework integrating item response theory calibration with sequential proficiency updating provides a principled and interpretable approach to estimating OHNS trainee operative competence from workplace-based assessment data. By extracting greater informational value from each individual assessment while accounting for procedure difficulty and trainee-specific performance trajectories, this framework may reduce the total number of ratings required to make meaningful competency determinations — representing a step towards precision medical education.

2:17 - DISCUSSION / Q&A (15 MIN)

2:32 - BREAK (15 MIN)

Session D: Laryngology, Pediatrics, & Otology

2:47 - **NIKHIL ARORA, MD (UCSF)**

THREE-DIMENSIONAL GEOMETRY OF GLOTTIC WEBS IN WENDLER GLOTTOPLASTY: IMPACT OF INFRAGLOTTIC CAPTURE ON PITCH ELEVATION IN AN EX VIVO PORCINE MODEL

Objective: Modified Wendler glottoplasty (MWG) research typically focuses on the anterior-posterior (AP) web dimension as a proportion of vocal fold (VF) length. We investigated the relevance of its craniocaudal dimension to determine whether thick infraglottic web geometry produces greater fundamental frequency (F0) elevation than a thin web of equivalent AP dimension.

Methods: Eight porcine larynges were prepared by removal of supraglottic structures. Compressed air was passed through the glottis to achieve phonation under three conditions: baseline, thin web, and thick web. For the thin web, an anterior suture was placed at a mean of 37% of VF length, capturing only the superior free edge. For the thick web, a suture at the same AP position captured the entirety of the infraglottic vibratory surface. F0 was extracted from audio recordings, and Wilcoxon signed-rank tests with Bonferroni correction compared F0 elevation between conditions.

Results: Median F0 at baseline was 361.5 Hz (IQR: 308.7–392.0 Hz). The thin web did not produce a statistically significant F0 elevation relative to baseline (Bonferroni-adjusted $p = 0.445$; Hodges-Lehmann [HL] shift: 162.6 Hz, 95% CI: -39.3–382.5 Hz). The thick web produced a statistically significant elevation in median F0 to 652.0 Hz (IQR: 491.0–779.7 Hz) from baseline (Bonferroni-adjusted $p = 0.023$; HL shift: 255.4 Hz, 95% CI: 68.1–452.9 Hz). Direct comparison between web geometries was significant, non-adjusted, but did not survive Bonferroni correction ($p = 0.039$; adjusted $p = 0.117$).

Conclusion: Vocal fold vibration is a three-dimensional (3D) phenomenon driven by mucosal wave propagation from the infraglottic surface. This model demonstrates that MWG web geometry impacts F0 elevation not only through its AP dimension but also through the incorporation of the infraglottic vibratory surface of the glottis. Deliberate infraglottic capture may represent a technically modifiable surgical variable to optimize pitch elevation outcomes in gender-affirming voice surgery.

2:55 - PAULINE HUYNH, MD (KP)

A STRUCTURED REVIEW OF COMMERCIALY AVAILABLE GENDER AFFIRMING VOICE CARE MOBILE APPLICATIONS

Background: Mobile health (mHealth) applications are increasingly used by individuals seeking gender-affirming voice care (GAVC) in hopes to supplement professional therapeutic vocal regimens or address this gap where access to GAVC is limited. However, the quality, functionality, and adherence of these apps to professional guidelines remain unclear.

Objective: To systematically evaluate the quality and functionalities of commercially available patient-facing mobile mHealth applications for gender-affirming voice care (GAVC).

Methods: Both the Apple iTunes Appstore and Google Play Store were queried for commercially available, patient-facing mHealth GAVC applications under the search terms “gender voice therapy,” “transgender voice training,” “voice feminization apps,” and “gender-affirming voice apps.” Each application was then evaluated by at least two reviewers for quality utilizing the Mobile Application Rating Scale (MARS) and functionality through the IQVIA Functionality Scale. The applications’ content was then evaluated against guidance from the World Professional Association for Transgender Health (WPATH) and American Speech-Language-Hearing Association (ASHA).

Results: Of the 306 unique applications, 7 met inclusion criteria for evaluation and analysis. On average, applications scored 3.3 out of 5 on quality and 5.1 out of 10 on functionality. The top-ranking mHealth GAVC application varied depending on overall application quality (VoiceShift), functionality (Christella Voice Up), or exposure to professional GAVC principles and guidance (PRYDE Voice and Speech Therapy). One application received the lowest score from all reviewers across all criteria, highlighting the utility of standardized evaluations when counseling patients on these mobile applications.

Conclusion: The quality of existing GAVC mHealth applications remains modest and highly variable, with their efficacy yet to be studied. Knowledge of the general market for these applications may guide patient counseling, especially for those seeking mHealth apps to supplement vocal gender expression practices or address poor access to GAVC. As these applications become increasingly commercially available and utilized by patients, it remains imperative that medical professionals provide input to ensure they contain content that is consistent with professional guidance while remaining functional and user-friendly.

3:03 - KIRSTEN WONG, MD (UC DAVIS)

SOCIAL VULNERABILITY INDEX AS A PREDICTOR FOR EMERGENCY DEPARTMENT UTILIZATION AMONG PEDIATRIC TRACHEOSTOMY PATIENTS

Background/Introduction: Children with tracheostomies are medically complex and frequently present to the emergency department (ED). Studies show social vulnerability is associated with higher ED utilization. However, the association specific to children with tracheostomies remains unclear. This study investigates how the social vulnerability index (SVI) predicts ED utilization among pediatric tracheostomy patients.

Methods/Design: We conducted a retrospective analysis and chart review of pediatric tracheostomy patients treated at a single tertiary care center between March 2007 and November 2025. Tracheostomy-related ED visits were divided into subcategories: wound issues, respiratory issues, accidental decannulation, sepsis or ACLS, and equipment failure. SVI and four subdomains (socioeconomic status, household characteristics, racial & ethnic minority status, and housing type & transportation) were calculated in accordance with CDC methodology.

Results: 664 ED visits from 104 pediatric tracheostomy patients were included. Overall, SVI was associated with higher total ED visits ($p < 0.05$) and hospital admissions ($p < 0.05$); however, subdomains of SVI did not show statistical significance as predictors of ED utilization. Ventilator dependence ($p < 0.01$), epilepsy ($p < 0.05$), sleep disorders ($p < 0.05$), chromosomal abnormalities ($p < 0.02$), asthma ($p < 0.05$), developmental delays ($p < 0.01$), and gastrointestinal conditions ($p < 0.05$) were associated with both higher total and ED visits per year.

Discussion: These findings suggest SVI and neighborhood-level sociodemographic factors may play a role in ED utilization in the pediatric tracheostomy population, but medical complexity appears to be a stronger predictor. Limitations to this study include a small sample size and a single-institution study.

Conclusion: Patient comorbidities and level of medical complexity, as well as SVI, should be considered for improving tracheostomy outcomes and ED utilization. Future directions for this study include examining time to decannulation and other trach-specific outcomes in relation to SVI.

3:11 - KETAN JAIN-POSTER, MD (KP)

A BLINDED COMPARATIVE EVALUATION OF CLINICAL AND AI-GENERATED RESPONSES TO OTOLOGIC PATIENT QUERIES

Background: The objective of this study is to assess the quality, empathy, and readability of large language model (LLM) responses regarding otologic questions from patients as they compare to verified physician responses in other patient-driven forums. This study aims to predict the potential utility of LLMs in patient-centered communication.

Methods: A sample of 49 otology-related questions posted on Reddit r/AskDocs1 between January 2020 and June 2025 were selected using search terms including “hearing loss,” “ear infection,” “tinnitus,” “ear pain,” and “vertigo.” Posts were retrieved using Reddit’s “Top” filter. Each question was answered by a verified doctor on Reddit and three AI LLMs (ChatGPT-4o, ClaudeAI, Google Gemini). Responses were scored by five evaluators.

Results: Common otologic concerns posed in patient questions were otalgia (38.7%), vertigo (28.6%), tinnitus (24.5%), hearing loss (22.4%), and aural fullness (20.4%). LLM responses were longer than physician responses (mean 145 vs 67 words; $p < .05$) and rated higher in quality (10.95 vs 9.58), empathy (7.26 vs 5.18), and readability (4.00 vs 3.73); (all $p < .05$). Evaluators correctly identified AI versus physician responses in 89.4% of cases with higher sensitivity for detecting physician responses (93.5%). By Flesch-Kincaid grade level, ChatGPT produced the most readable content (mean 7.25), while ClaudeAI responses were more complex (11.86; $p < .05$).

Conclusion: LLM responses received higher ratings in quality, empathy, and readability than those of physicians in response to a variety of otologic concerns. When appropriately implemented, such systems may enhance access to understandable otologic information and complement clinician-delivered care.

3:19 - DISCUSSION / Q&A (15 MIN)

3:34 - CONCLUSION

4:00 - RECEPTION & AWARD ANNOUNCEMENT

TENI EAST - 3770 PIEDMONT AVE, OAKLAND

THANK YOU

TO THE RESIDENTS FOR THEIR HARD WORK.
TO THE GUEST SPEAKER AND MODERATORS FOR THEIR SUPPORT.
AND TO THE PROGRAM DIRECTORS, RESEARCH MENTORS & FACULTY
FOR THEIR CONTINUED COMMITMENT TO RESIDENT EDUCATION.

PLEASE SAVE THE DATE

THE 43RD ANNUAL BAY AREA RESIDENTS' RESEARCH SYMPOSIUM IN
OTOLARYNGOLOGY—HEAD & NECK SURGERY

FRIDAY, MAY 14, 2027

COVER ARTWORK BY MICHELANGELO
CREDIT: SCALA / ART RESOURCE

**SCAN THE QR CODE
TO JOIN THE SESSION AND
SUBMIT QUESTIONS**

Join at:
vevox.app

ID:
124-399-679



THIS PAGE INTENTIONALLY BLANK

THIS PAGE INTENTIONALLY BLANK