Leveraging Machine Learning to Identify Risk Factors for HPV-Related Oropharyngeal Cancer and Sex-Based Differences

Austin Coppinger, BA[1], Kaden Bunch, MBA[1], Surya Khatri, BA[1], Audrey Su[1], Yolanda Troublefield, MD, JD[1]

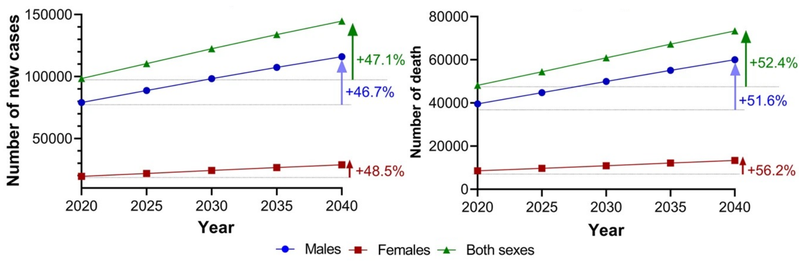
[1] Warren Alpert Medical School, Brown University, Providence, RI

THE WARREN ALPERT Medical School BROWN UNIVERSITY

Background

Globally, the incidence and mortality rates of **oropharyngeal cancers** are expected to rise by ~50% in the next 20 years. In the US, the incidence of HPV-associated oropharyngeal cancer has surpassed that of cervical cancer, with a marked and unexplained sex disparity favoring higher rates in men.

20-Year Projection of Global Oropharyngeal Cases and Deaths[1]



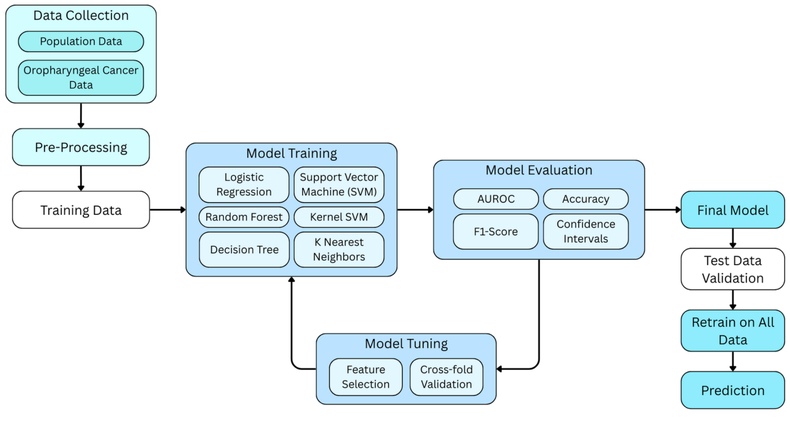
The underlying mechanisms likely involve complex sociodemographic, behavioral, and environmental interactions.

To better understand these contributors, this study employed **machine learning (ML)** techniques to identify **sex-specific community-level risk factors associated with oropharyngeal cancer incidence** in Massachusetts between 2001 and 2022.

Methods

We performed a **retrospective analysis** of Massachusetts public health data from 2001 to 2022.

- Population data across 352 towns and oropharyngeal cancer incidences were sourced from the Massachusetts Health Data and Environmental Public Health Tracking Tools.
- Six ML models were trained, evaluated, and tuned (dimensionality reduction, GridSearchCV) on training data only, then validated on test data.
- The most predictive models were used to identify key risk factors stratified by sex.



Results

Across all Massachusetts data, the K Nearest Neighbors model performed best for males:

- AUROC: 0.666 (CI: 0.589–0.738)
- Accuracy: 0.663 (CI: 0.596–0.725)
- F1-Score: 0.504 (CI: 0.403–0.600)

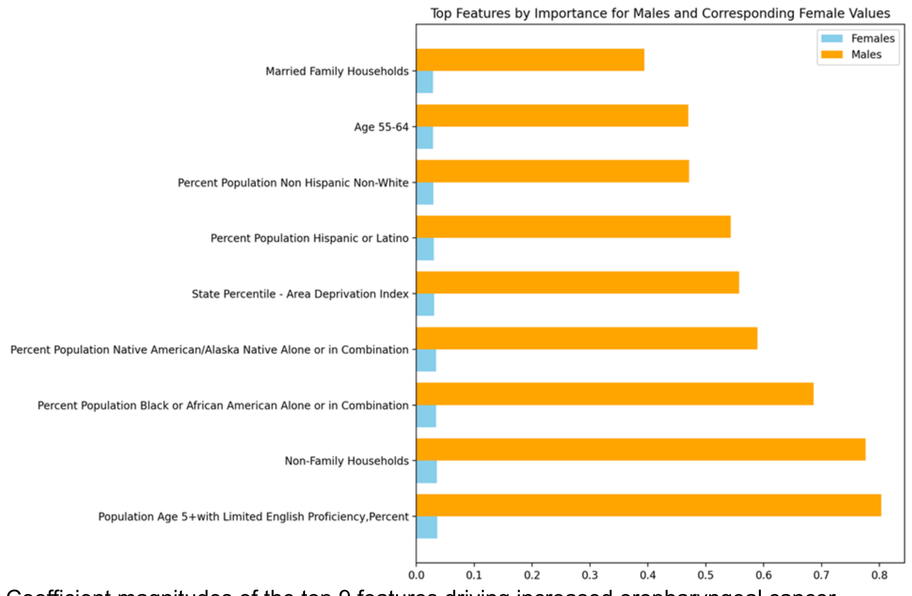
For males, key risk factors included:

- Limited English proficiency
- Low educational attainment
- Poverty
- Physical Inactivity
- Being of non-white race
- Residing in non-family households

For females, the Random Forest model showed the best performance:

- AUROC: 0.636 (CI: 0.559–0.709)
- Accuracy of 0.636 (CI: 0.554–0.683)
- F1-score of 0.503 (CI: 0.411–0.584)

However, no significant predictors were identified for females, highlighting distinct epidemiological pathways.



Coefficient magnitudes of the top 9 features driving increased oropharyngeal cancer incidence in males, alongside their corresponding coefficient magnitudes for females using the same variable set, demonstrating distinct epidemiological pathways of malignancy.

Key takeaway: Strong community-level predictors of oropharyngeal cancer incidence were identified in males, but associations remain unclear for females, pointing towards a potential sex disparity in how environmental and social factors influence cancer risk.

Discussion

The risk factors identified in this study reflect **broader social determinants** that influence behaviors like tobacco and alcohol use, which are known contributors to HPV-related oropharyngeal cancer.[2]

While it is well-known that HPV-associated oropharyngeal cancers predominantly affect males, this **sex disparity remains unclear** and cannot be solely explained by behavioral patterns.[3] In Massachusetts, this disparity in HPV-associated cancers is especially pronounced, with rising male oropharyngeal cancer rates surpassing national trends, while cervical cancer rates among females continue to decline.[4]

Future Directions

1. Further Investigation: Underlying Drivers of Sex Disparities
   - Explore the complex, sex-specific interactions among HPV exposure, behavioral risk factors, and access to preventive care.
2. Public Health Efforts: HPV Vaccination Coverage and Preventative Services
   - Emphasize the vaccine's role in preventing both cervical and oropharyngeal cancers and promote vaccine completion, particularly among racial and ethnic minorities.

Acknowledgements

We thank the for the American Academy of Otolaryngology–Head and Neck Surgery Foundation (AAO-HNSF) for the opportunity to present our research.

For any further questions, please feel free to contact Austin Coppinger (austin_coppinger@brown.edu).

References:

1. [Figure adapted from] Lim, Y.X., D'Silva, N.J. HPV-associated oropharyngeal cancer: in search of surrogate biomarkers for early lesions. *Oncogene* 43, 543–554 (2024). https://doi.org/10.1038/s41388-023-02927-9
2. Zumsteg ZS, et al. Global epidemiologic patterns of oropharyngeal cancer incidence trends. *J Natl Cancer Inst.* 2023 Dec 6;115(12):1544-1554. doi: 10.1093/jnci/djad169.
3. Windon MJ, et al. Sex differences in HPV immunity among adults without cancer. *Hum Vaccin Immunother*. 2019;15(7-8):1935-1941. doi: 10.1080/21645515.2019.1568157.
4. Cook EE, et al. Trends of two HPV-associated cancers in Massachusetts: cervical and oropharyngeal cancer. *Cancer Causes Control.* 2018 May;29(4-5):435-443. doi: 10.1007/s10552-018-1016-1.