

Association for Information Systems

## AIS Electronic Library (AISeL)

---

NEAIS 2023 Proceedings

New England Chapter of Association for  
Information Systems

---

10-28-2023

# Predictive Models of Chronic Obstructive Pulmonary Disease in the United States

Arnold Kamis

*Brandeis University*, [akamis@brandeis.edu](mailto:akamis@brandeis.edu)

Follow this and additional works at: <https://aisel.aisnet.org/neais2023>

---

### Recommended Citation

Kamis, Arnold, "Predictive Models of Chronic Obstructive Pulmonary Disease in the United States" (2023).  
*NEAIS 2023 Proceedings*. 14.

<https://aisel.aisnet.org/neais2023/14>

This material is brought to you by the New England Chapter of Association for Information Systems at AIS Electronic Library (AISeL). It has been accepted for inclusion in NEAIS 2023 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# **Predictive Models of Chronic Obstructive Pulmonary Disease in the United States**

## ***Completed Research Paper***

Arnold Kamis  
Brandeis University  
akamis@brandeis.edu

### **ABSTRACT**

In this paper, we investigate Chronic Obstructive Pulmonary Disease in the United States 2012-2017. We integrate data from multiple sources and use them to analyze COPD at the level of core-based statistical area. We include cigarette smoking and race / ethnicity categories because of well-known health disparities in the United States. We develop a baseline model with multiple linear regression and then attempt to improve upon it with machine learning methods, including Lasso Regression, Ridge Regression, Generalized Additive Model, Support Vector Machines, Artificial Neural Network, Random Forest, and Gradient Boosted Tree. The best machine learning model, a Support Vector Machine, captures an additional 6% variance explained in a strongly predictive model. Overall, cigarette smoking and household income are the strongest predictors. Future directions for research and practice are discussed.

### **Keywords**

Chronic Obstructive Pulmonary Disease, Multiple Linear Regression, Support Vector Machine, Cigarette Smoking, Household Income, Health Disparities, Non-Linear Models