Health Equity & Community Engagement

Assessment of multi-pollutant ambient air composition on type 2 diabetes mellitus using machine learning.

Naomi Oiwa Riches, Ramkiran Gouripeddi and Julio C. Facelli
University of Utah

ABSTRACT IMPACT: We explored the use of machine learning to explore how multi-pollutant air quality is related to type 2 diabetes, which is more representative than the single pollutant models often employed to assess this relationship. OBJECTIVES/GOALS: Single pollutant air pollution models have correlated air pollution components with type 2 diabetes mellitus (DM). However, air pollution is a complex mixture, therefore, we explored the relationship between multi-pollutant air quality and DM incidence using machine learning. METHODS/STUDY POPULATION: Annual diabetes incidence from the CDC for each US county was downloaded for the years 2007-2016. Daily air pollution concentrations for PM2.5, PM10, CO, SO2, NO2, and O3 were downloaded from the US EPA for the years 2006-2015. K-means clustering, an unsupervised machine learning method, was employed to partition all air pollution components, for each day and county monitored, into the optimal number of clusters. Change in DM incidence was matched to air pollution clusters by county, lagged by one year. Additionally, NASA satellite-derived air pollution data will be compared to EPA data to inspect as a potential source for future clustering analysis of counties that do not have an EPA monitor. RESULTS/ANTICIPATED RESULTS: The largest increase of annual DM incidence was associated with the cluster having the highest average PM10, PM2.5, and CO, and the second greatest average NO2 concentrations. Inversely, the most significant decrease of annual DM incidence was associated with the cluster having the lowest PM10, PM2.5, and CO. While average PM10, PM2.5, SO2, NO2, and CO showed a rising tendency with elevating change of DM incidence, ozone did not show any such trend. It is anticipated that the NASA satellite-derived air pollution data will approximate the EPA air quality data and will be usable in assessing the air pollution-DM relationship for areas currently not monitored by the EPA. DISCUSSION/SIGNIFICANCE OF FINDINGS: Using an unsupervised k-means algorithm, we showed multiple ambient air components were related to increased incidence of T2DM even when average concentrations were below the National Ambient Air Quality Standards. This work could help guide policy making regarding air quality standards in the future.

Precision Medicine

Characterizing Temporal Patterns in Glucose Dysregulation Following SARS-CoV-2 Infection

Sejal Mistry, Ramkiran Gouripeddi and Julio C. Facelli
Department of Biomedical Informatics University of Utah

ABSTRACT IMPACT: Understanding the longitudinal glucose changes following SARS-CoV-2 infection can inform point-of-care guidelines and elucidate the viral hypothesis of diabetes mellitus pathogenesis. OBJECTIVES/GOALS: Hyperglycemia has emerged as an important manifestation of SARS-CoV-2 infection in both diabetic and non-diabetic patients. Whether clinically-detectable glycemic changes persist following SARS-CoV-2 infection remain to be elucidated. This work aims to characterize temporal patterns in glucose dysregulation following SARS-CoV-2 infection. METHODS/STUDY POPULATION: Electronic health records of patients with a diagnosis of COVID-19, positive laboratory test for SARS-CoV-2, and negative history of Diabetes Mellitus prior to infection were extracted from the TriNetX database. 7,502 patients with at least one blood glucose value 2 years to 2 weeks before, 2 weeks before to 2 weeks after, and 2 weeks after to 1 year after COVID-19 diagnosis were used for analysis. Temporal patterns are characterized by training state-of-the-art clustering algorithms, including fuzzy short time-series clustering, k-means for longitudinal data, and spectral clustering. Clustering performance is evaluated using internal evaluation metrics of the Silhouette coefficient, Calinski-Harabasz score, and Davies Bouldin index. RESULTS/ANTICIPATED RESULTS: Based on the success of prior clustering methods with random blood glucose measurements, we anticipate that the proposed time-series clustering algorithms will appropriately characterize temporal patterns of glycemic dysregulation. The best performing algorithm based on interval evaluation metrics will be selected for further analysis. Associations between blood glucose values and cluster membership will be evaluated using Kruskal-Wallis one-way ANOVA and effect size will be calculated using unbiased Cohen's d. Clinical phenotypes for each cluster will be characterized in terms of current diagnoses, prior medication use, pertinent laboratory tests, and vital signs. DISCUSSION/SIGNIFICANCE OF FINDINGS: A clearer understanding of the longitudinal glucose changes following SARS-CoV-2 infection can elucidate clinically-detectable patterns of glycemic dysregulation, identify sub-phenotypes of patients who are more susceptive to glycemic dysregulation, and inform appropriate point-of-care guidelines.

A Phenomics Approach to the Categorization and Refinement of Heart Failure

Nosheen Reza, MD, William Bone, BA, Pankhuri Singhal, BS, Anurag Verma, MS, PhD, Ashwin C. Murthy, MD, Srinivas Denduluri, PhD, Srinath Adusumalli, MD, MSc, Marylyn D. Ritchie, PhD and Thomas P. Cappola, MD, ScM
Perelman School of Medicine at the University of Pennsylvania

ABSTRACT IMPACT: Measuring and analyzing qualitative and quantitative traits using phenomics approaches will yield previously unrecognized heart failure subphenotypes and has the potential to improve our knowledge of heart failure pathophysiology, identify novel biomarkers of disease, and guide the development of targeted therapeutics for heart failure. OBJECTIVES/GOALS: Current classification schemes fail to capture the broader pathophysiologic heterogeneity in heart failure. Phenomics offers a newer unbiased approach to identify subtypes of complex disease syndromes, like heart failure. The goal of this research is to use data-driven associations to redefine the classification of the heart failure syndrome. METHODS/STUDY POPULATION: We will identify < 10 subphenotypes of patients with heart failure using unsupervised machine learning approaches for dense multidimensional quantitative (i.e. demographics, comorbid conditions, physiologic measurements, clinical laboratory,