Health Equity & Community Engagement

73432

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

27337

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.

42855

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,

https://doi.org/10.1017/cts.2021.524 Published online by Cambridge University Press

imaging, and medication variables; disease diagnosis, procedure, and billing codes) and qualitative data extracted from an integrated health system electronic health record. The heart failure subphenotypes we identify from the integrated health system electronic health record will be replicated in other heart failure population datasets using unsupervised learning approaches. We will explore the potential to establish associations between identified subphenotypes and clinical outcomes (e.g. all-cause mortality, cardiovascular mortality). RESULTS/ANTICIPATED RESULTS: We expect to identify < 10 mutually exclusive phenogroups of patients with heart failure that have differential risk profiles and clinical trajectories. DISCUSSION/SIGNIFICANCE OF FINDINGS: We will attempt to derive and validate a data-driven unbiased approach to the categorization of novel phenogroups in heart failure. This 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.

44499

Heterogeneity of treatment effect among patients with type 2 diabetes and body mass index >=27kg/m^2 in the Jump Start Study

Elizabeth A. Kobe[1], Matthew J. Crowley, MD[2], Amy Jeffreys, MStat[3], William S. Yancy, Jr., MD[2], Jennifer Zervakis, PhD[3], David E. Edelman, MD[2] and Cynthia J. Coffman, PhD[2]
[1]Duke University School of Medicine, [2]Durham Veterans Affairs Center of Innovation to Accelerate Discovery and Practice Transformation (ADAPT), Duke University School of Medicine and [3]Durham Veterans Affairs Center of Innovation to Accelerate Discovery and Practice Transformation (ADAPT)

ABSTRACT IMPACT: This is the first study to use QUINT analyses to examine heterogeneity of treatment effect for group medical visits among individuals with type 2 diabetes. QUINT is a data driven method that assumes no a priori assumptions regarding effect moderators - an important step in the path towards personalized medicine. OBJECTIVES/GOALS: To examine heterogeneity of treatment effect (HTE) in Jump Start, a trial that compared the effectiveness of group medical visits (GMVs) focused on medication management only versus the addition of intensive weight management (WM) on glycemic control for patients with type 2 diabetes and body mass index >=27kg/m^2. METHODS/STUDY POPULATION: Jump Start patients (n=263) were randomized to a GMV-based medication management plus low carbohydrate diet-focused WM program (WM/GMV; n = 127) or GMV-based medication management only (GMV; n = 136) for diabetes control. We used QUalitative INteraction Trees (QUINT), a tree-based clustering method, to determine if there were subgroups of patients who derived greater benefit from either WM/GMV or GMV. Subgroup predictors included 32 baseline demographic, clinical, and psychosocial factors. Outcome was hemoglobin A1c (HbA1c). We conducted internal validation via bootstrap resampling to estimate bias in the range of mean outcome differences among arms. RESULTS/ANTICIPATED RESULTS: QUINT analyses indicated that for patients who had not previously attempted weight loss, WM/GMV resulted in better glycemic control than GMV alone (mean difference in HbA1c improvement = 1.48%). For patients who had previously attempted weight loss and had lower cholesterol and blood urea nitrogen levels, GMV alone was better than WM/GMV (mean difference in HbA1c improvement = 1.51%). Internal validation resulted in moderate corrections in the mean HbA1c differences between arms; however, differences remained in the clinically significant range. DISCUSSION/SIGNIFICANCE OF FINDINGS: Among patients with diabetes and BMI>=27kg/m^2, a low-carbohydrate, weight loss focus may better improve HbA1c in those who have never attempted weight loss. A medication management focus may be better in those who have attempted weight loss and have lower cholesterol and blood urea nitrogen.

89976

ASSESSING PROTEIN BIOMARKERS ROLE IN CVD RISK PREDICTION IN PERSONS LIVING WITH HIV (PWH)

Sandra Safo[1], Lillian Haine[1], Jason Baker[2], Cavan Reilly[3], Daniel Duprez[1], Jim Neaton[1], Jiuzhou Wang[1], Mamta K. Jain[4], Alejandro Arenas Pinto[5], Therese Staub[6] and Mark Polizzotto[7]
[1]University of Minnesota, [2]Hennepin Healthcare, [3]University of Minnesota School of Public Health, Biostatistics, [4]UT Southwestern Medical Center, [5]University College London, [6]Centre Hospitalier de Luxembourg and [7]The Kirby Institute, University of New South Wales Sydney

ABSTRACT IMPACT: Our findings could potentially identify CVD at-risk persons living with HIV who might benefit from aggressive risk-reduction. OBJECTIVES/GOALS: PWH have higher rates of CVD than the general population yet CVD risk prediction models rely on traditional risk factors and fail to capture the heterogeneity of CVD risk in PWH. Here we identify protein biomarkers that are able to discriminate between CVD cases and controls in PWH, and we assess their added benefit beyond traditional risk factors. METHODS/STUDY POPULATION: We analyzed 459 baseline protein expression levels from five OLINK panels in a matched CVD (MI, coronary revascularization, stroke, CVD death) case-control study with 390 PWH from INSIGHT trials (131 cases, 259 controls). We formed 200 datasets via bootstrap. For each bootstrap set, a two-component partial least squares discriminant model (PLSDA) was fit. The importance of each variable in the discrimination of cases and controls in the PLSDA projection was assessed by the variable importance in projection (VIP) score. Proteins with average VIP scores > 1 were used in penalized logistic regression models with elastic net penalty, and proteins were ranked based on the number of times the protein had a nonzero coefficient. Proteins in the top 25th percentile were considered to have high discrimination. RESULTS/ANTICIPATED RESULTS: Participants had mean age 47 years, 13% were females, 4.9% had CVD at baseline and 69% were on ART at baseline. Eight proteins including the hepatocyte growth factor and interleukin-6 were identified as able to distinguish between CVD cases and controls within PWH. A protein score (PS) of the top-ranked proteins was developed using the bootstrap (for weights) and the entire data. The PS was found to be predictive of CVD independent of established CVD and HIV factors (Odds ratio: 2.17 CI: 1.58-2.99). A model with the PS and traditional risk factors had a 5.9% improvement in AUC over the baseline model (AUC=0.731 vs 0.69), which is an increase in model predictive power of 18%. Individuals with a PS above the median score were 3.1 (CI: 1.83- 5.41) times more likely to develop

https://doi.org/10.1017/cts.2021.524 Published online by Cambridge University Press