GNS HEALTHCARE

ISPOR 19TH ANNUAL
INTERNATIONAL MEETING

ISPOR

MAY 31-JUNE 4, 2014
Palais des Congrès de Montréal
Montreal, QC, Canada

# IDENTIFICATION OF DETERMINANTS OF PROGRESSION TO TYPE 2 DIABETES USING ELECTRONIC HEALTH RECORDS AND 'BIG DATA' ANALYTICS

Anderson JP[1], Parikh JR[1], Shenfeld DK[1], Church BW[1], Laramie JM[1], Piper BA[2], Willke RJ[2], Mardekian J[2], Rublee DA[2]

[1]GNS Healthcare, Cambridge, MA, USA

[2]Pfizer, New York, NY, USA

OBJECTIVES: Type 2 diabetes (T2D) is strongly associated with morbidity and mortality, and carries a heavy financial burden. An improved understanding of the factors that predict progression to prediabetes and T2D is warranted. Thus our objective was to identify these determinants, applying 'big data' analytic methods to electronic healthcare records (EHR) data.

METHODS: Using GNS Healthcare's analytical platform (REFS$^{TM}$), we built unbiased predictive models for progression to T2D, computationally exploring trillions of potential models, in Humedica data. Humedica's data acquisition model aggregates de-identified EHR from providers across the continuum of care. Our study population consisted of 24,331 adults without history of prediabetes or T2D, from 2007-2012. Prediabetes was defined using World Health Organization criteria; incident T2D was identified using ICD-9 codes. Accuracy of prediction models was assessed using an area under the curve (AUC) statistic. We validated resulting prediction models in an independent dataset.

RESULTS: Our baseline model accurately predicted progression to T2D from normoglycemia (AUC = 0.76). We validated this model with an independent dataset where the AUC increased to 0.78. When the model was extended to include time-varying covariates the AUC increased to 0.87. Our model of progression from normoglycemia to T2D consisted of established risk factors (blood glucose measures, hypertension, income, race, triglycerides, lipid disorders, and blood pressure), whereas predictors of progression to prediabetes included novel factors such as high-density lipoprotein, alanine aminotransferase, C-reactive protein, and core body temperature (AUC = 0.70).

CONCLUSIONS: Using an extensive EHR database, we built accurate prediction models of diabetes risk using a hypothesis-free, machine-learning Bayesian approach, and validated them with an independent dataset. In doing so, we identified novel factors representing emerging areas of diabetes research and potential new targets for clinical management. Our ability to make individual T2D risk predictions has valuable applications to personalized medicine and clinical trial recruitment.