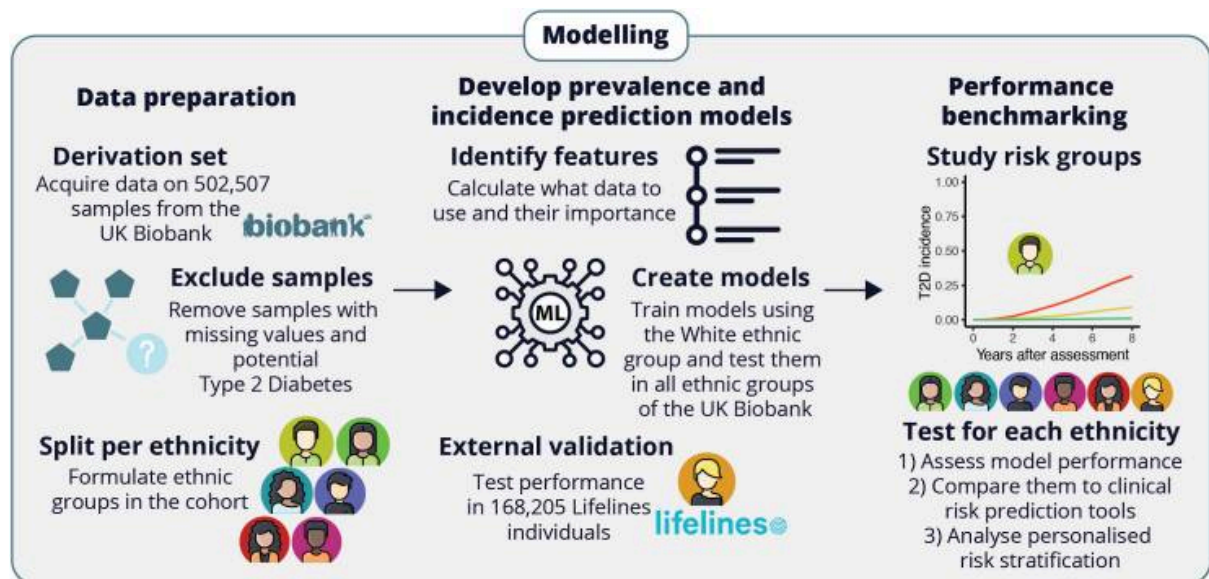


A Retrospective Analysis of Type-2 Diabetes Risk 10 Years Later

Type-2 Diabetes (T2D) poses a significant global health challenge, characterized by its complex etiology and prolonged asymptomatic phase. Early identification of individuals at risk is crucial for timely intervention and preventive measures. The landscape of T2D prediction has evolved, and predictive models have been developed to assess an individual's likelihood of developing the condition.



Workflow showing the steps taken to prepare the data and to create questionnaire-based prediction models for prevalent and incident type 2 diabetes.

The existing predictive models (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10626169/>) have shown promise in discerning individuals with an elevated risk of T2D. However, to enhance the practical utility of these models, it is imperative to investigate their performance over an extended timeframe. T2D is known to manifest itself years after the initial risk assessment, emphasizing the need for a comprehensive evaluation of these predictive models' efficacy in anticipating long-term outcomes.

The main objective of this project is to assess the effectiveness of current T2D predictive models in identifying individuals who would develop the condition a decade later. This involves creating a user-friendly tool that incorporates a machine learning algorithm into a R Shiny application, allowing users to easily generate logistic regression models.

Research question: "How well does the existing Type-2 Diabetes risk prediction model identify individuals who develop T2D a decade after the initial assessment?"

Analysis Plan

- **Data Collection:** Utilize the updated dataset with a larger T2D population for model development and validation. Study data are available from UK Biobank and Lifelines.

Data are, however, available from the authors upon reasonable request and when granted permission by the UK Biobank and Lifelines.

- Tool Development: Implementing a user-friendly R Shiny application that incorporates a machine learning algorithm for building logistic regression models. This tool aims to make it easy for users to generate and interact with the predictive models.
- Model Evaluation: Assess the performance of the previously published predictive model in detecting T2D risk 10 years prior to its manifestation. Utilizing algorithms like LASSO (Least Absolute Shrinkage and Selection Operator) to identify the most predictive variables. Employing techniques such as Ridge regression to prevent overfitting, ensuring model generalizability. Applying k-fold cross-validation to assess model performance and stability across different subsets of the dataset.

Significance

This project aims to advance public health outcomes by enhancing T2D prediction accuracy, enabling early detection, and facilitating preventative interventions. The approach merges an extensive dataset with cutting-edge machine learning strategies, setting a new standard for disease risk prediction.

Organisation

Ancora Health Bv
Lübeckweg 2, 9723 HE Groningen
Dr. Sipko van Dam
sipko@ancora.health
+31 (0) 6 383 245 59