Prediction of treatment efficacy in patients with OSA and T2D

Obstructive sleep apnoea (OSA) and type 2 diabetes (T2D) are commonly co-occurring conditions that significantly increase the risk of cardiovascular consequences such as myocardial infarction, heart failure, and/or death. Despite this, many patients with T2D remain underdiagnosed for OSA, leading to missed treatment opportunities. This is partly due to uncertainty about the long-term health effects of OSA treatment, especially in patients with T2D. Since these questions are challenging to address through randomized clinical trials, where the effects take a long time to materialize, alternative methods are required.

This project aims to address these knowledge gaps using longitudinal registry data, also known as "real-life data". Analysis of these data is challenging due to their size covering millions of observations and tens of gigabytes of data. Due to this, the Big Data approach will be taken in this project assuming using Apache Spark and MLLib.

Description of Dataset

The data for the project is derived from Swedish national health registries, which offer extensive longitudinal information on diagnoses, treatments and outcomes. The main datasets include:

- **National Diabetes Register (NDR):** Covers the majority of patients with T2D in Sweden and contains detailed information on metabolic control, treatment, and outcomes, with multiple data points per individual over time.
- Swedish Sleep Apnoea Registry (SESAR): Provides information on OSA patients, including diagnostic parameters and treatment decisions, but has limited longitudinal follow-up.
- National Prescription Drug Register (NPDR): Provides detailed information on prescribed and dispensed medications, allowing for analyses of pharmacological treatments and their impact on outcomes.
- **Cause of Death Register:** Nearly complete coverage of causes of death, though naturally lacking longitudinal follow-up.

Research Questions

- Which method is most appropriate for modelling survival probabilities in the Big Data context?
- How does the performance of this method compare with the performance of the default MLLib survival analysis model for prediction of treatment effects in OSA patients with T2D, particularly related to CPAP treatment?
- How does the outcome of these models differ for CPAP treated and untreated groups?
- How do concurrent pharmacological treatments, such as antihypertensive and lipid-lowering medications, affect the efficacy of CPAP treatment in reducing cardiovascular risk?

Prerequisites

Good knowledge of statistics, machine learning and programming. Experience with Big Data analytics in Apache Spark.

Contacts

- The main supervisor for the project is Oleg Sysoev <u>oleg.sysoev@liu.se</u>
- The medical expertise is provided by Fredrik Iredahl, Associate Professor and General Practitioner, and Jonas Agholme, PhD student and Specialist in Internal Medicine.