

Using National Electronic Health Care Registries to Analyse and Predict Alcoholic Liver Disease

Dhouha Grissa*, Ditlev Nytoft Rasmussen**
Aleksander Krag**, Søren Brunak*, Lars Juhl Jensen*

*Novo Nordisk Foundation Center for Protein Research,
University of Copenhagen, Denmark.

{dgrissa, soeren.brunak, lars.juhl.jensen}@cpr.ku.dk

**Department of Gastroenterology and Hepatology, Odense University Hospital.
{Ditlev.Nytoft.Rasmussen2, Aleksander.Krag}@rsyd.dk

1 Analysis of health registry data

Alcoholic liver disease (ALD) is a common chronic liver disease worldwide Mathurin and Bataller (2015). It progresses from fatty liver to alcoholic liver fibrosis (ALF) then to cirrhosis (ALC). Unfortunately, the clear majority of patients with ALD are diagnosed so late that they already have irreversible damage to their livers. It is then important to find an effective way to detect ALD at an early stage, by identifying alcohol over-using patients at risk of developing ALD. To examine this problem, we use the Danish National Patient Registry (NPR) comprising diagnoses covering the entire population of Denmark during ~19 years (1996–2014). The data set covers 6.6 million patients with a total of 153 million clinical encounters, and contains ~101 million unique assignments of primary and secondary diagnoses coded in the International Classification of Diseases 10th Revision. These diagnoses are combined with data on sex and birth/death dates Jensen et al. (2014). Accordingly, we use machine-learning techniques to discover relevant upstream diagnoses helping to give a new avenue of understanding how ALD develops and progresses in heavy drinker patients.

The analysis of NPR data lead to the identification of a cohort of 33,391 patients with ALD, from which we derive groups of patients with ALC and ALF. From each group, we extract 2-year upstream time window of the first discharge date of ALC/ALF diagnosis, and obtain three sets of patients: (i) 10,831 of patients with ALC; (ii) 12 patients with ALF but not ALC; and (iii) 23 patients with ALF at least 6 months before ALC. These numbers underline the main motivation for this work, namely that the clear majority of ALD patients are discovered only after ALF has developed into irreversible ALC.

2 Learn from ALC and test on ALF

Based on the group of 10,831 patients with ALC, we perform a matched case-control study where we randomly select for each case, 5 matched controls not suffering from ALD if available. We then transform their NPR data into a binary dataset of the form 52,926 patients