

**Autumn VRREG Meeting 2020: “Collaborative research in VRREG”
October 1, 2020**

Schedule

Introduction

10:00 - Opening remarks from Erin Gabriel, Jesper Lagergren and Anna Schandl (moderator).

15 Minute presentations by Nodes about research

10:05 - Upper Gastro-Intestinal Research (UGIR)

Survival after surgery vs medication for gastro-oesophageal reflux disease in a Nordic registry-based study - Manar Yanes, KI.

Abstract:

Background: Whether antireflux surgery influences survival in patients with gastro-oesophageal reflux disease (GORD) is unclear. We aimed to examine the hypothesis that antireflux surgery improves long-term survival compared to antireflux medication in patients with severe GORD.

Methods: Population-based cohort study of adults with severe and objectively determined GORD, i.e. reflux esophagitis or Barrett’s oesophagus, in 1980-2014 in Denmark, Finland, Iceland, or Sweden. Multivariable Cox regression provided hazard ratios (HR) with 95% confidence intervals (CI) of all-cause mortality and cause-specific mortality, adjusted for sex, age, calendar period, country, and comorbidity.

Results: Among 240,226 cohort patients with severe GORD, 33,904 (14.1%) underwent antireflux surgery. Compared to antireflux medication, the overall HR of all-cause mortality was decreased after antireflux surgery (HR 0.61, 95% CI 0.58-0.63), and the HR was lower after laparoscopic (HR 0.56, 95% CI 0.52-0.60) than open surgery (HR 0.80, 95% CI 0.70-0.91). The overall HRs were decreased for mortality from cardiovascular disease (HR 0.58, 95% CI 0.55-0.61), respiratory disease (HR 0.62, 95% CI 0.57-0.66), laryngeal or pharyngeal cancer (HR 0.35, 95% CI 0.19-0.65), and lung cancer (HR 0.67, 95% CI 0.58-0.80), but not from oesophageal cancer (HR 1.05, 95% CI 0.87-1.28), comparing antireflux surgery with medication. The risk reductions remained decreased over time after antireflux surgery.

Conclusions: In patients with severe GORD, antireflux surgery may decrease the risk of mortality from all causes, cardiovascular disease, respiratory disease, laryngeal or pharyngeal cancer, and lung cancer, but not from oesophageal cancer, compared to antireflux medication.

10:20 - Statistical Methods for RBR

Methods for survival outcomes with competing risks in register research - Michael Sachs, KI

Abstract:

In the early aughts, pseudo-observations were introduced as a statistical technique to estimate relatively simple regression models for complex event history outcomes, such as recurrence-free survival. The procedure involves perturbing a marginal, usually nonparametric estimator to transform the outcome into a single continuous variable. It can be applied to specify regression models for the restricted mean survival, survival probabilities, and in settings with competing risks. Prediction models involving machine learning algorithms can be easily adapted to event history outcomes using this technique. In this talk, I will describe a project from our node that developed pseudo-observation based ensemble prediction modeling for survival outcome data with competing risk, with an application in Crohn's disease.

10:35 – Register based machine learning

Defining neighborhoods using satellite imagery and register data- Rolf Lyneborg Lund, Aalborg University

Abstract:

Considering neighborhoods and especially the social aspects of neighborhoods is a core component in sociology. Does it matter where we grow up? Does neighborhood composition affect later life outcomes? Does moving away from a neighborhood change individual action? All of these questions lead to the same root question; what defines a neighborhood? By using the physical landscape and satellite imagery we try to encompass how the physical neighborhood affect the social neighborhood; can we see social segregation in the physical world and what does it look like? Using a combination of register data and image recognition on neighborhood level satellite data might paint a very different picture of what deprivation or prosperity really is.

10:50 – Social Science Genomics - Uppsala University

The genetics of life course outcomes – Leveraging new methods to advance social-science genomics - **Aysu Okbay, Vrije Universiteit Amsterdam**

Abstract:

We conducted a genome-wide association study (GWAS) of educational attainment (EA) on ~3 million individuals of European ancestry. We estimated both additive and dominance effects for SNPs on EA. We found 3744 loci with genome-wide significant additive effects. Our analysis shows that dominance effects play only a limited role in the effects of common genetic variants on EA. We constructed a polygenic score (PGS) from the additive effect estimates, and found that it explained 13% of the variation in educational attainment.

11:05 - Coffee break 10 min

11:15 – AIR Lund

AIR Lund Chest Pain Substudy - Explainable clinical decision support based on health care register data - **Ulf Ekelund and Stefan Larsson, Lund University**

Abstract:

The AIR Lund Chest Pain Substudy aims to develop decision support tools based on machine learning for the management of emergency department (ED) chest pain patients. For this purpose, we exploit the ESC-TROP register which contains extensive data from approximately 30 000 consecutive chest pain patients at the Skåne EDs in 2017-2018. As a first step, we have created an artificial neural network (ANN) to rule out acute myocardial infarction based on patient age, gender and two serial blood samples for troponin T. Preliminary results show that the ANN can increase the safe and early discharge of low risk chest pain patients from the ED compared with currently used decision rules. Future development steps include adding information to the ANN from e.g. the ECG and previous medical history.

The AIR Lund project also includes ethical and legal concerns with data-driven predictive tools. A challenge lies in making complex predictions explainable in the sense that they are sufficiently transparent for auditing and accountability as well as for being trustworthy in the context they are used in. The decision support tools we develop - particularly those providing high-stakes recommendations based on less intuitive models such as the above - will therefore also be assessed for explainability in relation to trust and accountability, which includes the relationship between the tool, its professional user and the patients.

11:30 - STELLAR, Gothenburg and Luleå

Deep learning-based phenotyping of airway diseases in adults - **Rani Basna, Krefting Research Centre, University of Gothenburg**

Abstract:

A typical challenge in clustering heterogeneous data lies in handling both numerical and categorical variables at the same time. Many commonly used approaches are limited in reflecting the correct distance that is needed to perform a good clustering.

Airway diseases in adults, such as asthma and COPD, are very heterogeneous diseases, proposed to be embedded with varying underlying disease phenotypes. Machine-learning techniques are nowadays helping to uncover the latent phenotypes, but dealing with the

heterogeneous landscape of variables that contribute to deriving the phenotypes remains an ongoing computational challenge. We demonstrate the use of mutual information (MI)-based Unsupervised Feature Transformation algorithm to convert our categorical variables to numerical variables, thereby preserving the information under consideration for onward phenotype discovery. The conventional Principal Component Analysis usually applied as a dimensionality reduction method has the limitation of only able to uncover linear correlation between the variable and produce a latent space that is faithful to that linearity. We proceeded with performing a developed version of the Deep Embedding Clustering (DEC) called a manifold Deep Embedding Clustering (MDEC) method to learn non-linear latent space that preserves properties of the data both locally and globally. Once this is done, we then followed the approach of finding a consensus clustering of different initializations of the MDEC as a robust approach for getting a final stable clustering result to represent the disease phenotypes. Finally, we validated the stability of the derived disease phenotypes based on replication analysis by implementing a random forest classifier. The random forest classifier also tests the reproducibility of the phenotypes and the validity of the resulted structure. The above pipeline of methods was applied to the population- representative cohorts of the West Sweden Asthma Study and the Obstructive Lung Diseases in Northern Sweden.

11:45 – **GOCARTS, Gothenburg**

How can the decreasing CHD mortality in Sweden 2002-2016 be explained? Design of a modelling study using registers and study cohort data - **Lena Björck, senior lecturer, Sahlgrenska Academy, University of Gothenburg, and Sahlgrenska University Hospital/Östra Hospital, Gothenburg**

Abstract:

Background: Coronary heart disease (CHD) mortality rates have been falling in Sweden by more than 80% since the 1980s. We have previously used the validated IMPACT CHD model to explain how much of the mortality decrease between 1986 and 2002 was due to medical and surgical treatments and how much to changes in risk factors. We found that improvement in risk factors, mainly a marked decrease in serum cholesterol, explained more than half of the CHD mortality decrease. Since 2002, both medical treatments and interventions has improved while physical inactivity, obesity and intake of saturated fat have increased. We have also seen an increasing prevalence of diabetes type 2. During the same time preventative medical treatment has improved in individuals without prior cardiovascular disease (CVD) in the population, as well as clinical management of CVD. Therefore, we aim to update the model and examine how much of the changes in CHD mortality between 2002 and 2016 that can be attributed to improved medical treatments and how much to changes in risk factors.

Methods: We use the validated IMPACT mortality model to combine data on uptake and effectiveness of cardiological treatments and interventions and risk factors. The main data sources are official statistics, national quality of care registers, published trials and meta-analyses, and national population surveys. The model calculate the number of deaths expected in 2016 if the CHD mortality rates in 2002 had persisted by multiplying the age-specific mortality rates for 1986 in the population for each age group in 2016. All data are analyzed separately for men and women, aged 25-84 years, divided six age groups (25-34; 35-44; 45-54; 55-64; 65-74; 75-84 years).

Data sources: Total population and age distribution, deaths and CHD mortality from 2002 to 2016 are obtained from the National Board of Health and Welfare and information on number of patients admitted because of acute myocardial infarction, unstable angina, heart failure, and cardiac procedures (CABG,PCI) from the National Hospital Discharge register. Information on medical treatment and/or interventions is collected from national Quality of Care registries: SWEDEHEART: acute myocardial infarction (AMI), unstable angina, CABG and PCI; SEPHIA (secondary prevention after AMI, CABG, PCI); RiksSvikt (heart failure); the Swedish Cardiac Arrest Registry (CPR in community)

Population risk factors are obtained from population studies and national surveys

- smoking, physical inactivity, BMI: ULF the Official Statistics of Sweden;
- cholesterol, blood pressure: population studies (MONICA, AdIn2, PURE, SCAPIS)
- diabetes and hypertension in community: The VGR Administrative Healthcare Database

Proposed results of the study: The IMPACT mortality model will combine and analyse data on uptake and effectiveness of treatments and risk factor trends in Sweden. Number of deaths prevented or postponed by treatments in individuals and by population risk factor changes will be calculated. The results will be used to inform the scientific community and stakeholders about the quantitatively most important steps to reduce coronary mortality in the community.

12:00 - SINGS – **Anita Berglund, Karolinska Institutet**

Description:

Swedish registers constitute a unique resource for research, and thus contribute to better public health and welfare. Individual-level data, so-called microdata, includes vital events, health aspects, and demographic and socioeconomic indicators for the entire population over decades. These features make these data an indispensable and powerful resource for answering a multitude of research questions, in a time- and cost-effective manner. However, when using these properties for research purposes, ethical and legal considerations must be encompassed to the full. The Swedish Interdisciplinary Graduate School in Register-Based Research (SINGS) covers a variety of quantitative research disciplines, such as epidemiology, public health, sociology, demography, psychology, statistics, health economics, and other medical and social sciences. The school is coordinated by Karolinska Institutet, and six other Swedish higher education institutions take active part. SINGS intends to develop deeper knowledge, skills, and scientific and ethical approaches regarding how different sources of data can and should be utilised in research. The target group is doctoral students involved in register-based research. The school runs in 2-year cycles, and students are admitted biannually. Courses represent the major part and include core courses and elective courses that are offered on different topics and at different levels in order to enable the tailoring of individualised study plans.

12:15 –13:00 Lunch

Randomly assigned small groups for brainstorming about collaboration

13:00 – All participants will be randomly assigned to different break-out rooms on Zoom, and given 20 minutes in each room to introduce themselves (1 min/person) and discuss potential collaborations (concrete examples, i.e. specific projects, which we can do better together).

13:05 - Randomly assigned Break-out groups

13:25 - Randomly assigned Break-out groups

13:45 - Randomly assigned Break-out groups

14:05 Short break

Assignment to a longer discussion group

14:10 – 14.20

**Based on participants' selection sent by chat or by hand-raising during this period we will assign all participants to one of the longer discussion break-out rooms.*

Discussions based on morning sessions hosted by each of the Nodes

14:20-15:00 Running in parallel from

Room 1 - **Upper Gastro-Intestinal Research (UGIR)** – Hosts: Manar Yanes, Giola Santoni and Jesper Lagergren

Room 2 - **Statistical Methods for RBR** - Hosts: Michael Sachs and Erin Gabriel

Room 3 - **Register based machine learning** – Hosts: Rolf Lyneborg Lund and Maria Brandén

Room 4 - **Social Science Genomics, Uppsala University** – Host: Sven Oskarsson

Room 5 - **AIR Lund** – Hosts: Jonas Björk, Ulf Ekelund, and Stefan Larsson

Room 6 - **STELLAR, Gothenburg and Luleå** – Hosts: Rani Basna and Helena Backman

Room 7 - **GOCARTS** – Hosts: Annika Rosengren and Lena Björck

Room 8 - **Information and discussion about SINGS** - Host: Anita Berglund

Summary and Wrap-up

15:00 – 15:15 Erin Gabriel, Jesper Lagergren and Anna Schandl