

The Future of Statistical Modeling in Medical Data

17 – 21 September 2018 @ Snellius

Description and aims of the workshop

The validity and practical utility of medical research depend critically on study design, data quality, and statistical methods. The inherent complexity of the processes involved in disease occurrence, progression and treatment, together with the increasing volume and complexity of data collected in observational studies (“Big Data”), pose important conceptual and analytical challenges. Advanced statistical modelling is increasingly applied in this research area. In this workshop, we aimed to link model building challenges in the areas of prediction, explanation and causal effects.

The core of this workshop was formed by three topic groups (TG) of the STRATOS (STRengthening Analytical Thinking for Observational Studies) initiative (<http://www.stratos-initiative.org/>), TG 2: Selection of variables and functional forms in multivariable analysis; TG 6: Evaluating diagnostic tests and prediction models; and TG7: Causal inference. The workshop intended to deepen the work in each of three of the topic groups, and to explore overlapping topics together. The workshop also aimed to introduce young statisticians (PhD and post docs) to the work of STRATOS to enlarge their research networks. Comparing competing strategies and derive evidence supported guidance is one of the main aims of STRATOS. As well designed simulation studies are the key instrument for this task, we invited an expert and dedicated a half day plenary session to it.

Results

The workshop has been very fruitful:

Ongoing work on manuscripts has been updated and (nearly) finalized. This included a tutorial paper on causal inference and a paper entitled ‘Overview of methods for the selection of variables and functional forms in multivariable analysis’ where key issues of TG2 are discussed.

A set up for a series of subsequent papers has been made. For TG2: papers on (a) spline based approaches, (b) overview of and recommendations for variable selection procedures (c) overview for researchers with low statistical knowledge. For TG6, four initiatives were taken for scientific papers on prediction: performance of survival models, choosing decision thresholds, measures for calibration, and a review on sample size for prediction models.

Furthermore, several of these papers will be joint work of different topic groups. 1. A paper for applied researchers on simulation studies (TG2, TG6, and TG7), and a paper on variable selection in causal inference (TG2 and TG7).

The session about simulation studies started with an invited guest talk from Tim Morris (London). It stimulated interesting discussions. Motivated by this we decided to work on two papers (a) background of simulations studies (for people without any knowledge and statistical training) and (b) relevant issues to design simulation studies (for researchers without deeper training and knowledge).

New collaborations have been formed: members of TG 6 and TG 7 will work together on topics arching prediction and causal inference. Members of TG 7 responded to work of TG 2 members in variable selection. Results of many issues discussed and progress from work in the coming months will be presented in June 2019 at the 2nd general meeting of the STRATOS initiative at the Banff International Research Station (<https://www.birs.ca/>).

Organization/Format:

We organized the workshop as a mixture of plenary sessions in the morning where overlapping topics were discussed (simulation studies, modelling strategies and variable and function selection, prediction versus causality and machine learning). Further elaboration and discussion was done in smaller workgroup sessions in the afternoons. This was well appreciated by the participants and the workshop location with smaller and larger rooms was very suited for this set-up. The boat trip with dinner organized by the Lorentz center was a great success. Even though some participants could not attend the full 5 days we look back to a very successful meeting, which has given an impulse to the field of statistical modeling in medical data and will result in substantial concrete research output.

Organizers

Saskia Le Cessie (Leiden, The Netherlands)

Willi Sauerbrei (Freiburg, Germany)

Ewout Steyerberg (Leiden, The Netherlands)