Basel Biometrics Society Seminar
Basel, 1st November 2019

Second announcement - BBS Seminar: Predictive modelling, machine learning and causality

Date: Friday, November 1st, 2019, 08:30-16:45
Venue: Auditorium Roche Builing 683, Viaduktstrasse 31-35, Basel

The BBS is pleased to host a full-day seminar on predictive modelling, machine learning, and causality with several eminent speakers. The talks will present recent methodological advances and challenges as well as case studies from the pharmaceutical industry and academia. We welcome all quantitative scientists to this event which will be a great opportunity to meet with colleagues and exchange ideas on this emerging and vibrant field.

The seminar is free of charge but registration is mandatory for organizational reasons.
Please register via email to fred.sorenson@xcenda.com by Friday, October 18, 2019, the latest.

Program:

08:30 – 9:00 **All**
Registration

09:00 – 9:15 **Uli Burger, BBS President**
Welcome and introduction

9:15 – 10:00 **Ewout Steyerberg, Leiden University Medical Center and Erasmus MC**
Clinical prediction models in the age of artificial intelligence and big data

10:00 – 10:30 **Coffee break**

10:30 – 11:15 **Willi Sauerbrei, University of Freiburg**
The EQUATOR network and guidelines for prediction models

11:15 – 12:00 **Torsten Hothorn, University of Zurich**
Score-based transformation learning

12:00 – 12:55 Lunch

12:55 – 13:40 **Peter Bühlmann, ETH Zürich**
Causal regularization for distributional robustness and replicability

13:40 – 14:00 **Case study 1: Giusi Moffa, University of Basel**
Predicting putative intervention effects after causal structure learning from survey data

14:00 – 14:20 **Case study 2: Andrew Shattock, Swiss TPH**
Using machine learning and disease models to evaluate target product profiles of novel interventions

14:20 – 14:50 **Coffee break**

14:50 – 15:10 **Case study 3: Federico Mattiello, Roche**
Identifying high-risk patients in follicular lymphoma by building a prognostic score

15:10 – 15:40 **Mark Baillie, Novartis**
Novartis benchmarking initiative: making sense of AI

15:40 – 16:10 **Chris Harbron, Roche**
Experiences from running internal prediction challenges within a pharmaceutical company

16:10 – 16:40 **Panel discussion with all the speakers**

16:40 – 16:45 **Uli Burger, BBS president**
Closure

We look forward to your participation!

Abstracts

Clinical prediction models in the age of artificial intelligence and big data

Ewout Steyerberg, Leiden University Medical Center and Erasmus MC

Clinical prediction models hold the potential to provide individualized estimates of risks of diagnostic and prognostic outcomes, which may support medical decision making and improve clinical outcomes. This potential may better be fulfilled by using more data and better methods. In this talk I will aim to address strengths and weaknesses of Big Data initiatives, and consider links between classical statistical approaches, machine learning, and artificial intelligence (AI). An internet search shows that there is considerable confusion on terminology, where some may call any somewhat larger data set an example of Big Data, and any method that aims to learn from data, including linear regression, an example of AI. Several examples will be given for illustration, including limitations in data quality and dubious improvement in analysis methods. Furthermore, validation and updating approaches are essential if we agree on the purpose of prediction models as providing reliable predictions for a specific context.

The EQUATOR network and guidelines for prediction models

Willi Sauerbrei, University of Freiburg

For many years the quality of research in the health sciences has been heavily criticized. It is argued that serious improvement would be possible if biomedical research is better chosen, designed, executed, analyzed, regulated, managed, disseminated, and reported. Serious improvements are far from being simple for many of the issues mentioned, but suitable guidelines have been developed to improve on the reporting of research. Severe weaknesses in this area are unnecessary and can be avoided.

Concerning issues in reporting of health science the EQUATOR (Enhancing the QUAlity and Transparency Of health Research, https://www.equator-network.org/) network acts as an umbrella organization. The REMARK guidelines have been proposed for Tumor Marker Prognostic Studies (McShane et al 2005) and TRIPOD for the reporting of multivariable prediction models for individual prognosis or diagnosis (Collins et al 2015). For both reporting guidelines more detailed 'explanation and elaboration' papers were published (Altman et al 2012, Moons et al 2015), which also include several parts about statistical analyses. The TRIPOD checklist distinguished between model development and validation.

Despite of guidelines for reporting, many reviews of publications have clearly shown that the quality of reporting is still bad (Sekula et al 2017, Heus et al 2018). For (nearly) all relevant diseases many prediction models have been published, but they are often far from being clinically applicable (Kleinrouweler et al 2016). For many prognostic and prediction studies even basic items of the study population and relevant details of statistical analyses are often not provided. The REMARK profile has been proposed to improve completeness and transparency of the reporting of statistical analyses (Altman et al 2012, Winzer et al 2017).

Statistical methodology has seen substantial development in recent times. Unfortunately, many of these methodological developments are ignored in practice. Consequently, design and analysis of observational studies often exhibit serious weaknesses. These observations led to the initiation of the STRATOS (STRengthening Analytical Thinking for Observational Studies, http://www.stratos-initiative.org/) initiative, a large collaboration of experts in many different areas of biostatistical research (Sauerbrei et al 2014). With an emphasize on topic groups TG2 'Selection of variables and their functional forms in multivariable analysis' and TG6 'Evaluating diagnostic tests and prediction models' we want to discuss some of the issues when developing a prediction model.

References

Altman DG, McShane L, Sauerbrei W, Taube SE (2012): Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. In: PLoS Med 9(5)

Collins GS, Reitsma JB, Altman DG, Moons KG (2015): Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. Ann Intern Med. 162:55-63

Heus P, Damen JAAG, Pajouheshnia R, Scholten RJPM, Reitsma JB, Collins GS, Altman DG, Moons KGM, Hooft L (2018): Poor reporting of multivariable prediction model studies: towards a targeted implementation strategy of the TRIPOD statement. BMC Medicine. 16:120. doi:10.1186/s12916-018-1099-2

Kleinrouweler CE, Cheong-See FM, Collins GS, Kwee A, Thangaratinam S, Khan KS, Mol BW, Pajkrt E, Moons KG, Schuit E (2016): Prognostic models in obstetrics: available, but far from applicable. Am J Obstet Gynecol. 214(1):79-90.e36. doi: 10.1016/j.ajog.2015.06.013

McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM for the Statistics Subcommittee of the NCI-EORTC Working on Cancer Diagnostics (2005): REporting recommendations for tumor MARKer prognostic studies (REMARK). Journal of the National Cancer Institute. 97: 1180-1184

Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS (2015): Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. Ann Intern Med.162(1):W1-W73

Sauerbrei W, Abrahamowicz M, Altman DG, le Cessie S and Carpenter J on behalf of the STRATOS initiative. (2014): STRengthening Analytical Thinking for Observational Studies: the STRATOS initiative. Statistics in Medicine., 33: 5413-5432

Sekula P, Mallett S, Altman DG, Sauerbrei W (2017): Did the reporting of prognostic studies of tumour markers improve since the introduction of REMARK guideline? A comparison of reporting in published articles. PLoS ONE. 12(6):e0178531. doi:10.1371/journal.pone.017853

Simera I, Moher D, Hoey J, Schulz KF, Altman DG (2009): The EQUATOR Network and reporting guidelines: Helping to achieve high standards in reporting health research studies. Maturitas. 63 (1): 4–6. doi: 10.1016/j.maturitas.2009.03.011

Winzer KJ, Buchholz A, Schumacher M, Sauerbrei W (2016): The prognostic ability of the Nottingham Prognostic Index in breast cancer: better use of standard information and clinical relevance. PLoS ONE, 11(3):e0149977

Score-based Transformation Learning

Torsten Hothorn, University of Zurich

Many statistical learning algorithms can be understood as iterative procedures for explaining variation in scores, that is, in the gradient vector of some target function. The statistical interpretation of boosting as functional gradient descent is maybe the most prominent representative, but also model-based trees and forests have been discussed from this point of view.

While these algorithms are agnostic with respect to the target function, we specifically discuss scores obtained from the likelihood of fully parameterised transformation models. This model class is sufficiently large and interesting while at the same time allows for a unified theoretical and computational treatment.

In this line of thinking, we can understand and implement classical procedures, such as the Wilcoxon-Mann-Whitney-Rank-Sum test, the log-rank test, maximally selected rank statistics, or regression trees and contemporary statistical learning procedures, most importantly random

forests and boosting, as extremes in a continuum of increasingly complex models featuring directly interpretable parameters.

We discuss prognostic and predictive models of increasing complexity as transformation models for conditional distributions. The estimation of heterogeneous treatment effects from experimental and observational data is presented as one application currently receiving much interest in various disciplines.

Causal regularization for distributional robustness and replicability

Peter Bühlmann, ETH Zürich

The common notion of replicability of statistical discoveries deals with generalization from a data set to a new unobserved population from the same data-generating distribution (and is typically quantified by some statistical uncertainties).

We discuss the problem when the new population comes from a different distribution than the one generating the observed data. We will present a principled way to achieve replicability in such settings: it builds on distributional robustness and borrows ideas from causality. So-called anchor regression with a simple, yet effective ``causal regularization'' provides a novel methodology for predictive robustness and replicability. We highlight the potential and limitations of the approach and provide some illustrations on bio-medical data.

The presentation is based on joint work with Dominik Rothenhäusler, Nicolai Meinshausen and Jonas Peters.

Predicting putative intervention effects after causal structure learning from survey data

Giusi Moffa, University Hospital Basel

Directed acyclic graphs (DAGs) have been traditionally used in forward causation to estimate the effects of causes given a postulated causal structure informed through domain experts. Thanks to computational progress in structure learning for Bayesian networks, DAGs have now also gained popularity in reverse causation. Inferring the DAG structure from observational data may give us insights about putative causal mechanisms, though only under very strict assumptions. By going through networks compatible with the observed data we can derive putative intervention effects. Advances in sampling Bayesing networks from their posterior distribution given the data further allow us to follow a fully Bayesian approach to derive a posterior distribution of putative causal effects which account for the uncertainty in the estimation of the DAG structure. Here we focus specifically on binary variables with a case study in Psychosis.

Using machine learning and disease models to evaluate target product profiles of novel interventions

Andrew Shattock, Swiss TPH

Target Product Profiles (TPPs) define the minimum and preferred product characteristics (PPCs) of a proposed medical intervention, therapeutic drug or other product, in order to make decisions about whether the new intervention should be developed or used. However, TPPs and their respective PPCs are often set by expert opinion and consensus, without quantitatively considering the complex dynamics of a disease and intervention, nor quantitative predictions of likely impact. Mathematical models of disease can be used to bridge this gap, as they can quantitatively estimate

the impact of arbitrary interventions, and can include considerable detail of disease progression and transmission, individual immune system, and health system dynamics and their interaction with new interventions. However, searching all possible interventions for a disease in conjunction with all possible scenarios leads to a combinatorial explosion in parameter space. In addition to the extreme computational cost in simulating all combinations, the considerable model and parameter complexity obscures the relationships between intervention parameters and health outcomes or public health impact.

Here we present a framework for quantitatively defining TPPS or PPCs, based on a machine learning approach that generates computationally light emulators of detailed mathematical models of disease dynamics. These emulators capture not just the mean tendency of the models, but also explicitly represent the variance inherent in stochastic models. This approach allows us to efficiently perform sensitivity analyses of public health impacts on intervention and health scenario parameters, while limiting computational cost. This framework provides a method for defining TPPs by efficiently searching highly complex parameter spaces of mathematical disease models, and by quantitatively identifying the determinants of desired public health impact.

We illustrate our framework by application to malaria disease, defining the required profiles or key performance parameters in TPPs of new putative interventions to reach desired public health goals such as elimination or prevalence reduction. In doing so we take into account operational and other constraints that for example limit coverage or efficacy. We used an individual-based model of malaria transmission dynamics to quantitatively examine all proposed interventions from drugs, vaccines to novel vector control from the last ten years.

The presentation is based on joint work with Guojing Yang, Monica Golumbeanu, Flavia Camponovo, Nakul Chitnis, Ewan Cameron, and Melissa A. Penny.

Identifying high-risk patients in follicular lymphoma by building a prognostic score

Federico Mattiello, Roche

Follicular Lymphoma is an indolent haematological cancer where available therapies already provide very good outcomes for most patients. However, a non-negligible subset of the population does not respond to therapy and is at high-risk of progression/death.

In this study, we sought to identify these high-risk patients at baseline by combining in-house data sources, and to build a clinically-actionable prognostic score suitable for use in both routine clinical decision making and recruitment of patients to new clinical trials.

In this talk, we will show the models we proposed and some surprises we found along the way.

Novartis benchmarking initiative: making sense of AI

Mark Baillie, Novartis

There is a big hope that AI will be transformative in drug development. Progress is fast moving with many AI, machine learning and advance analytics companies regularly approaching Novartis with capabilities that may have the potential to advance our programs. But how do we systematically evaluate and identify partners who will add value? At Novartis, we are developing a standard process for benchmarking that: (i) scientifically evaluates and compares vendors on real tasks, (ii) de-risks engagement by removing the need to provide sensitive data during evaluation, (iii) and reduces internal resource required to engage vendors through this screening. In this presentation we will provide an overview of the initiative. We will also describe an application to a case study on applying machine learning and AI to two phase III studies, presenting our learnings so far.

Experiences from running internal prediction challenges within a pharmaceutical company

Chris Harbron, Roche

Roche has a strategy of encouraging the internal development of Advanced Analytic skills such as predictive modelling and machine learning, ensuring knowledge of these methods are available in a critical mass of individuals and across all relevant departments within the company. One important driver of this strategy has been the running of company-wide prediction challenges which has been highly successful in driving engagement, discussion and collaboration across the company. These have also provided valuable people-development introducing people to new data types and sources and prediction methods as well as increasing the profile of Advanced Analytics within the company.

In this talk I'll outline how the challenges were set up and what the results across the competing teams looked like. I will also share some insights from comparing the predictions from teams using different approaches and also discuss the potential value of concensus scoring, combining predictions from different teams.