2nd year:	**Module 5**
Module Title:	**Biostatistics & Advanced Epidemiology**
Number of Credits:	**9 EC**

Aim of the module
This module introduces some multivariable analysis techniques using the generalized linear models (GLM) framework. Multivariable logistic modeling will be used to demonstrate model building to predict outcomes. Linear mixed models for analysis of repeated measures will be introduced, and Cox proportional hazard regression will be used to demonstrate multiple regression analysis as a means to adjust for confounding factors in the context of causal reasoning. Propensity score methods will be used as as an upcoming addition to multivariable adjustment techniques. Students will expand knowledge on more complex techniques for clinical research on estimation of effects of a determinant or effectiveness of treatments. Also, the potential impact of various types of missing data will be discussed, and the use multiple imputation techniques to deal with missing data will be demonstrated.

Overall synopsis
The student will be guided through a range of multivariable statistical techniques. In the course of seven lectures and computer-aided tutorials, introducing (multivariable) lineair regression, logistic modeling, analysis of repeated measurements, analysis of survival data using Cox's proportional hazard regression, causal modelling and propensity score matching. In addition, missing data mechanisms and their handling with the appropropriate imputation method are included

Learning outcomes as a whole
The student is able to analyse data from clinical studies using multivariable regression analysis, the student is aware of threads to the internal and external validity (including missing data) of studies, the student is able to circumvent these issues when designing and analyzing a study, and the student knows how and when to correct for confounding during analysis of the data and interpret the results.

Teaching and learning strategies
Each session consists of a lecture and a tutorial. Students are encouraged to read textbooksfor deeper understanding of a topic. Papers on methodology and medical studies, as well as text books are used to improve analysis skills and critical thinking.

Assessment strategies
The assessment will be a written exam with open ended and multiple choice questions.
Session 1: Lecture – Recapitulation on elementary statistical analysis & multivariable linear and Cox regression

The lecture starts with a recapitulation of module 3 on elementary statistical analysis. Then the lecture will proceed with introducing multivariable linear regression analysis and Cox proportial hazard analysis to cope with confounding and investigate effect modification.

Learning Objectives
The student has an understanding of multivariable linear and Cox regression. The student can operate R-studio to perform a multivariable regression analysis, can model data to select independent determinants, study interactions, or to control for confounding and can interpret the result.

Session 2: Lecture – Missingness and data imputation methods

Missing data in clinical research are a rule rather than an exception. Their potential to undermine the validity of research results has often been overlooked in the medical literature. Several strategies are available to handle missing data, each of which has advantages and disadvantages. The amount of, and reason for data missing determines which strategy should be used.

Learning objectives
The student has knowledge of consequences of missing data, has knowledge of missing data mechanisms and imputation methods, including multiple imputation by Chained Equations (using the mice package in R), and can choose and apply appropriate imputation techniques to handle missing data.

Session 3: Lecture – Analysis of repeated measurements

The analysis of data which has been measured on multiple occasions in the same individuals during a given time span is problematic, as measurements further in time are dependent on initial measurements. A lineair mixed-effects model regression analysis ("multi-level linear model") can deal with this kind of data.

Learning Objectives
The student has knowledge and understanding of the problems associated with analysis of dependent data, can perform a lineair mixed-effects analysis using R to model a continuous outcome that was measured repeatedly, also adjusted for other covariates, and can interpret the results.

Session 4 & Session 5: Lectures – Prediction model building

The lectures focus on how to build and apply a multivariable model to predict the occurrence of an event (How to develop a parsimonious model that fits the data well and can be used in new patients (evaluation of *overfitting*).

Learning Objectives
The student has at an intermediate level understanding of multivariable modeling techniques to predict an event, their assumptions, and methods to assess the calibration and accuracy of the model. The student can operate R-studio to perform a multivariable logistic regression analysis, understands the use of non-linear terms (splines), can carry out diagnostics to identify numerical instability, assess collinearity, linearity in the logit, issues around the identification of independent predictors (variable selection). Student can apply the methods to determine internal and external (cross-) validity of the developed prediction model: Bootstrapping, narrow and broad validation, and shrinkage of parameters to obtain a more accurate prediction in new patients.

Session 6: Lecture – modelling a causal association

Causality always implies at least some relationship of dependency between the cause and the effect. Deeming something a cause may imply that, all other things being equal, if the cause occurs, the effect does too, or at least that the probability of the effect occurring increases. Also a cause chronologically precedes the effect. Causal modeling is related to, but not the same as, a variety of other analytical methods such as multiple regression. More specifically, causal modeling guides regression adjustments and stipulates rules whether of not to correct for a potential confounding variable.

Learning objectives
The student knows the principles of causal modeling in an observational design setting. Student has a basic understanding of causal pathways or Directed Acyclyc Graphs (DAGs)

Session 7: Lecture - Introduction to propensity score matching.

In absence of randomization, propensity score matching is a method to provide unbiased estimation of treatment or exposure effects. Lacking randomization, observational studies are particularly prone to confounding by indication and frequently provide biased estimates of treatment effects. A propensity score is a balancing score that is obtained by calculating the probability of exposure / treatment as a function of the measured baseline characteristic of participants, usually by logistic regression. Subsequent matching of subjects on their propensity score attempts to reduce the confounding effects of covariates in one go, without having to meet complex modelling assumptions and limitations of multiple variable adjustments for the unbiased estimation of treatment / exposer effects.

Learning objectives
The student knows the theory and assumptions of propensity scores outlined by Rosenbaum & Rubin (1983) and computational methods to obtain propensity scores and methods of balancing groups on propensity scores (individual matching and subclassification on PS), as well as additional multivariable correction using the computed propensity score.

Session 8: – Putting it all together.

Synthesis of Lectures 1 to 7. The Tutorial is a catch up of tutorial assignments and training for the final exam using a example exam that was handed out during Lecture 7.