1st year:	**Module 3**
Module Title:	**Biostatistics - elementary analysis**
Number of Credits:	**9 EC**
Module leader:	**M.M. Stuiver, PhD**
Lecturers:	**S. Perry, MSc / S. de Wolf, MSc**

Aim of this module:
This module introduces students into Biostatistics and the use of statistical software. Students will be able to select appropriate, common statistical test and estimation procedures in relation to the research designs, to check the requirements before applying these procedures, to perform these procedures using a statistical software program, and to interpret the results. Moreover, students will learn about the boundaries of their competence with respect to analysis.

Overall synopsis:
In the course of seven sessions, students will be familiarized with a range of basic statistical procedures to estimate an effect size, to test a predefined hypothesis or to assess common variation between measurements. In addition, students will be familiarized with the more advanced statistical techniques for modeling linear relationships (linear regression, logistic regression & Cox's proportional hazard regression)
All sessions will comprise a theoretical lecture in combination with a practical workshop. During the workshop, students will practice their skills using a statistical software program (RStudio), based upon computer assignments for analyzing an existing dataset from a randomized clinical trial.

Learning objectives as a whole:

- The student is able to identify the type of measurement scale for any given variable, and to select the appropriate descriptors for the whole population.
- The student is able to compute an estimate with confidence interval and interpret the results, both by hand as well as with the use of a statistical computer program.
- The student is familiar with the array of basic statistical procedures to test predefined hypotheses regarding differences between different groups, to check all requirements for any given procedure, to select the appropriate statistical procedure in accordance with the research question, to execute said procedure using a statistical software program, to interpret and to report the results.
- The student is familiar and experienced with the statistical procedures to compute and interpret correlations between different variables, and the techniques for linear, logistic and Cox's proportional hazard regression.
- The student is able to perform above mentioned procedures using an appropriate statistical software program (RStudio)
- The student is able to critically reflect upon used statistical procedures commonly applied in the analysis of randomised clinical trials.

Assessment strategies:

Knowledge of biostatistics and analytical skills will be tested by a written exam with open-ended theoretical questions.

Notes:

Final mark for this module will only be granted when the module assignment (analysing a given dataset and reporting on the results) has been turned in.

Session 1 – Descriptive statistics
Student can explain differences between descriptive statistics and inferential statistics; Student can name characteristics of a Normal distribution; Student can distinguish between categorical (nominal, ordinal) data and numerical (discrete, continuous) data; Student can name and explain characteristics of these measurement levels of data; Student can recognize and name the measurement level of variables; Students understands the role of descriptive statistics in data checking and reporting; Student can describe variables with appropriate measures (or graphs) of point location and spread.

Session 2 – Analysing numeric data from two groups
Student knows the distinction between dependent and independent variable; Student knows the difference between estimation and testing; Student can use features of z-distribution to describe aspects of a normalized and normal distribution; Student can calculate 95% CI around sample mean by hand; Student can describe difference between paired and unpaired testing ; Student can decide when to use parametric or non-parametric analyses; Student can describe the relationship between critical ratio and p-value (reiteration).

Session 3 – Analysing numeric date from more than two groups
Student can apply a systematic approach towards parametric and non-parametric testing considering: study question at hand; measurement level of variables; number of groups (2 or more); (in)dependency of data; test assumptions.

Session 4 – Correlation and lineair regression
Student can use parametric and non-parametric techniques to describe a linear relationship between two continuous variables; Student can make an informed choice between these techniques; Student can define a linear regression to inter- and extrapolate points of this relationship; Students can describe the function and use of both regression coefficients; Student can evaluate all assumptions that are associated with (univariate) linear regression.

Session 5 – Contigency tables & logistic regression
Students can analyse data presented in contingency tables; Students can carry out a regression analysis to assess the effect of a categorical or a continuous (numerical) independent variable, on a binary dependent variable (= logistic regression); Students can describe the function and use of both the β and Exp(β) regression coefficients.

Session 6 – Survival analysis & Cox's proportional hazard regression
The student knows the principles of the time-to-event data and can perform a survival analysis and can interpret the results of Kaplan-Meier analysis and Cox's proportional hazard regression.

Session 7 – Multiple Regression & recap
Student knows rationale and aims of multiple regression. This lecture will conclude with a comprehensive recap of the module. Emphasis will be put on reporting the results of statistical analyses.