

## Biostatistics & Epidemiological Data Analysis using R -- winter semester 2021-2022

**Lecturer:** Dr. Stefan Konigorski ([stefan.konigorski@hpi.de](mailto:stefan.konigorski@hpi.de))

This course teaches (i) basic epidemiological concepts and (ii) biostatistical methods and their application for data analysis of large epidemiological datasets using the statistical software R ([www.r-project.org](http://www.r-project.org)) and the graphical interface RStudio ([www.rstudio.com](http://www.rstudio.com)). To this aim, the class starts with an introduction to R and RStudio. R Markdown will be used as a tool for documentation and reporting of the analysis results. Next, the class covers data processing steps and introduces epidemiological study designs as well as theoretical and practical aspects of basic and more advanced biostatistical methods. In addition to classical biostatistical approaches such as linear and linear mixed models, newer methods how to deal with missing values, how to perform meta analyses, and for causal inference will be discussed and applied.

### Content:

- Introduction to R, RStudio
- Documentation and report writing using R Markdown
- Data setup: create, import, export datasets in R
- Format datasets in R: transform variables and manipulate datasets
- Descriptive statistics
- Tables and graphics to visualize data and results
- Epidemiological study designs and study planning
- Introduction to statistical parameter estimation and hypothesis testing
- Statistical methods for dealing with missing values
- Linear and logistic regression models
- Linear mixed models for the analysis of clustered and longitudinal data
- Meta analysis
- Survival analysis
- Statistical methods for causal inference

### Learning goals:

At the end of the course, the students will be able to

- understand the main concepts of basic and more advanced biostatistical methods and select appropriate methods for data analysis of epidemiological studies
- import and manipulate datasets in R for statistical analysis
- perform a data analysis in R considering measurement error and missing values
- document the analysis and report the results using R Markdown.

<u>Course format:</u>	Lectures and tutorials with practical exercises
<u>SWS:</u>	4+2
<u>Credits:</u>	6 ECTS
<u>Prerequisites:</u>	Laptop with R and RStudio installation. While the class is self-contained, any previous exposure to programming, data analysis, and statistics is helpful.
<u>Condition for admission to final exam:</u>	Hand in solutions to 10 of the 12 weekly assignments
<u>Final grade:</u>	Open book take home final exam