

Research Programmes Board, FUN

## **Approaches to the Handling of Missing Data MEHSD01**

**1.5 credits**

**Third cycle**

### **General information**

The course provides a background to the issue of missing data and to the consequences of simple ad hoc methods to address the issue. The advantages and shortcomings of different methods will be discussed. The method in focus on the course is multiplied imputation (MI), which participants will have the opportunity to test in the laboratory components. The course is primarily intended for doctoral students at the Faculty of Medicine who handle sets of data including missing data in their projects. Secondly and space permitting, other applicants, such as postdocs and senior researchers who wish to learn more about how to handle the problem, will be admitted. Participants are required to have prior knowledge equivalent to Applied Statistics I and II and solid experience of statistics software able to implement multiple imputation (e.g. R, Stata or SPSS). Furthermore, participants must have theoretical and practical knowledge of linear and logistic regression models.

### **Language of instruction**

English

### **Objective**

The aim of the course is to make participants aware of the consequences of incorrect handling of missing data in medical research in general and to provide them with tools for correct handling of missing data in their own research.

### **Learning outcomes**

On completion of the course, the participants shall be able to

- identify and discuss different types of missing data (mechanisms)
- implement multiple imputation (MI) using the chained equations technique
- analyse data sets imputed via MI
- interpret and report results from MI analyses
- account for the limitations of MI

## Content

1. Introduction to missing data
  - Identifying missing data
  - Potential consequences of missing data
  - Mechanisms for the generation of missing data
  - Brief overview of methods for handling missing data
2. Multiple imputation
  - Brief theoretical background to MI
  - The chained equations method
  - Constructing an imputation model
  - Analysing imputed data
  - Diagnosis of the MI model (model validation)
3. Reporting MI results and the limitations of the method
  - Guidelines for reporting analyses of MI-generated data
  - Limitations of the MI method

This continuation course introduces the concept of missing data and the potential consequences of this problem, and provides a brief overview of methods to handle missing data. The focus is on the method of multiple imputation. The idea and theory behind the method is explained briefly – the emphasis is placed on practical applications. How do you construct an imputation model? How do you summarise and interpret the results? What sort of diagnosis is required? The course is concluded with practise in reporting statistical analyses of MI data. The concrete examples primarily consist of regression models for cohort studies and randomised controlled trials.

## Design

The course comprises one week of full-time work. Three full days consist of timetabled lectures, group exercises and computer labs and two days are intended for independent work, home assignments and an exam. In the computer labs, the participants perform analyses using statistics software (R, Stata or SPSS) which they are to be familiar with from previous courses.

## Assessment

The assessment is based on a take-home exam including tasks that are to be solved using statistics software. For a Pass on the course, participants must have passed the exam and participated actively in the timetabled course days.

## Grades

The grades awarded are Pass or Fail.

## Admission requirements

To be admitted to the course, applicants must have prior knowledge equivalent to Applied Statistics I and II. In particular, they are to be familiar with the theory of

linear and logistic regression models and be able to adapt these models and interpret corresponding output from their statistics software. The choice of software is optional, but the participants are required to be familiar with the program they choose and that it has the method multiple imputation via chained equations implemented (e.g. R, Stata or SPSS). The participants are responsible for ensuring that they have a suitable and working statistics program installed on their computer before the start of the course.

## Required reading

Information on articles and other study resources will be provided before and during the course.