STRATOS Workshop

Presenter: Lara Lusa
Title: **STRATOS initiative: An overview**

Abstract: The STRATOS initiative (STRengthening Analytical Thinking for Observational Studies) is a large collaboration of experts in many different areas of biostatistical research. Its main goal is to improve the current practice in design and statistical analyses of observational studies in practice by closing the gap between available statistical methodology and methods applied in real-life though guidance for researchers with different levels of statistical expertise. As of 2024, the STRATOS initiative includes more than 100 members from 21 countries and 5 continents that work in 9 topic groups and 12 panels. In this short introduction I will introduce the goals and current activities of the STRATOS initiative.

Presenter: Marianne Huebner
Title: **Content of a Statistical Analysis Plan**

Abstract: Developing a statistical analysis plan (SAP) is a critical component of designing and conducting research studies. It influences collection and analysis of data, and interpretation of results. A well written SAP prevents research waste, can be published on its own, allows replication of the study by independent researchers, can help when writing the manuscript. A systematic process of initial data analysis (IDA) is often ignored in SAPs but is necessary for transparency and reproducibility. The focus of this presentation is on SAPs for observational studies and reporting the effort of an international group of statisticians of the STRATOS Initiative to develop a template for such an SAP with IDA.

Presenter: Anne-Laure Boulesteix
Title: **Towards reliable empirical evidence in methodological biostatistical research: Recent developments and remaining challenges**

Abstract: Statisticians are often keen to analyze the statistical aspects of the so-called "replication crisis in science". They condemn fishing expeditions and publication bias across empirical scientific fields applying statistical methods, such as health sciences. But what about good practice issues in their own - methodological - research, i.e. research considering statistical (or more generally, computational) methods as research objects? When developing and evaluating new statistical methods and data analysis tools, do statisticians and data scientists adhere to the good practice principles they promote in fields which apply statistics and data science? I argue that methodological researchers should make substantial efforts to address what may be called the replication crisis in the context of methodological research in statistics and data science, in particular by trying to avoid bias in comparison studies based on simulated or real data. I discuss topics such as publication bias, cherry-picking, and the design and necessity of neutral comparison studies, and review recent positive developments towards more reliable empirical evidence in the context of methodological biostatistical research.

Presenter: Federico Ambrogi
Title: **Analysis of High Dimensional Biomedical Data**

Abstract: In high-dimensional data (HDD) the number of measured variables greatly exceed the number of participants included in the study. In biomedical research, omics HDD data measure characteristics of the genome, proteome or metabolome; also electronic health records can provide HDD due to the large number of variables being recorded. The aim of this talk is to provide an overview of the characteristics of HDD, of the possible aims of the experiments that use them, and of the most commonly used approached used to prepare data and analyze them.

Presenter: Willi Sauerbrei

Title: **Multivariable model building - on some issues in selection of variables and functional forms**

Abstract: In most observational studies, many variables are measured. For a multivariable analysis, the selection of variables with a potential association with the outcome is necessary, and determination of the functional form for continuous covariates is required. There is general consensus in the scientific community that subject matter knowledge should play a key role in determining suitable models. However, is there sufficient knowledge to determine a model without important data-dependent decisions? What would constitute a 'state-of-the-art' analysis?
In this talk, we will provide an overview of important issues addressed in topic group 2 'Selection of variables and functional forms in multivariable analysis' of the STRATOS initiative (Sauerbrei et al, 2020) and discuss key issues of the multivariable fractional polynomials approach (Sauerbrei et al, 2022; https://mfp.imbi.uni-freiburg.de/).

Sauerbrei, W., Kipruto, E., Balmford, J. (2022). Effects of Influential Points and Sample Size on the Selection and Replicability of Multivariable Fractional Polynomial Models. Diagnostic and Prognostic Research, 7(1):7

Sauerbrei W, Perperoglou A, Schmid M, Abrahamowicz M, Becher H, Binder H, Dunkler D, Harrell Jr. FE, Royston P, Heinze G for TG2 of the STRATOS initiative (2020). State of the art in selection of variables and functional forms in multivariable analysis - outstanding issues. Diagnostic and Prognostic Research, 4:3, 1-18