Project A:

Order-restricted response-adaptive randomisation

A well-conducted clinical study needs to ensure that treatment groups are comparable among patients, which can be achieved by randomising patients to treatment arms. This randomisation is implemented using allocation procedures that can remain unchanged or be adapted during the trial.

Response-adaptive randomisation allows the allocation probabilities to treatments in a clinical trial to change based on the accumulating data, in order to achieve experimental objectives such as assigning more patients to the best-performing treatments. In this internship project, we consider a clinical trial setting where multiple treatment arms are considered concurrently, and an order between them can be assumed (i.e., if multiple doses of a treatment are simultaneously tested, then higher doses can lead to higher efficacy compared to lower doses). It is not currently clear how to incorporate this ordering into response-adaptive randomisation procedures. The aim of this project is to explore how to do this and assess the potential benefits through simulation.

Supervisors: Dr Alessandra Serra and Dr David Robertson (Efficient Study Design Theme)

Project B:

Modelling time-to-onset of suspected adverse drug reactions using pharmacovigilance data

In the recent years, the public interests on adverse reactions of drugs and vaccines has rapidly grown. In pharmacovigilance, spontaneous reports of suspected adverse drug reactions are used to characterise the safety profile of drugs. In many studies the focus is on the frequency of adverse events for a specific drug, while other information of clinical relevance is overlooked. In particular, the time-to-onset, defined as the time difference between the first drug administration and the first manifestation of the adverse event, plays a fundamental role which is not fully explored. The analysis of the time-to-onset has numerous implications from a biological perspective, as it leads to the understanding of novel pharmacological mechanisms of actions of the drug, as well as a direct clinical impact on patients, who benefit from an early detection and timely management of adverse events.
The objective of the project is to review and develop statistical models for time-to-onset analysis in pharmacovigilance data from the FDA Adverse Event Reporting System (FAERS) official database. The intern will produce an exploratory analysis on a case study, then apply and compare different regression models (in the class of accelerated failure time models), using packages available in R, to study the association between time-to-onset and predictors of interest such as age, sex and reporter type. For this project, a basic knowledge of generalised linear model (GLM) is expected. Familiarity with time-to-event analysis is desirable but not essential.

Supervisor: Dr Marco Palma (Precision Medicine Theme)

Project C:

A comparison between global and multiple testing procedures

Clinical trials often consider several dependent outcomes. In the study motivating this project, the outcomes of interest are the reduction in both blood pressure and cholesterol levels between treatment and control groups for patients with cardiovascular disease. This work will consider methods allowing gains in efficiency by considering the dependence between the target quantities. Such gains means that fewer patients are exposed to futile drugs and that effective treatments are delivered to the broader population sooner, with studies requiring smaller sample sizes.

Conducting multiple statistical tests within the same study can lead to an increased risk of obtaining false positive results. This is where a significant treatment effect is detected when it does not truly exist. The aim of the internship will be to make a formal comparison between methods which control Type 1 error rates when considering multiple outcomes. The first option for comparison is to test each outcome separately and apply a multiplicity adjustment. The comparator method is a novel testing procedure which identifies a global summary statistic to summarise the overall effect of treatment given all outcomes combined.

The first key milestone will be to implement the most common methods which consider outcomes separately and create a framework for evaluating efficiency measures. A subtle task here is to define the efficiency measures and translate these to patient benefit outcomes. Depending on the interests of the intern, there are theoretical and simulation-based approaches to consider. In the second month, the intern will make a formal comparison between the approaches, changing factors such as the number of outcomes analysed, orderings of importance and the given statistical model.

Supervisors: Dr Abigail Burdon and Dr Dominique-Laurent Couturier (Efficient Study Design Theme)

Project D:

The internship will consist in improving the handling of missing values' imputation in differential quantitative proteomics data. Quantitative proteomics aims to identify and quantify proteins in a given biological system at a given time and under given conditions. In the more specific context of differential proteomics, the goal is to identify a subset of proteins differentially expressed between two conditions (e.g., sick vs healthy patients), which serve as potential biomarkers.

Current practice consists of performing statistical tests, namely moderated t-tests (Smith, G. 2014), to compare the mean intensities of each protein between both conditions. However, quantitative proteomics data usually contains 5-15% missing values, which are imputed, *i.e.,* replaced by a user-defined value. Yet, these imputed data are less reliable than truly observed values, meaning that they add uncertainty to the final statistical output. That uncertainty is usually not taken into account downstream of the state-of-the-art statistical methods. Our previous work (Chion *et al.* 2021) led to a moderated t-test that accounts for the multiple imputation-based variability. It adds a projection step of the covariance matrix derived from the multiple imputation process before injecting it into the moderated t-test statistic. While this work provided better results than state-of-the-art methods, some perspective work could refine the testing strategy. This includes switching

the projection and the moderation step and eventually removing the projection step to benefit from all the information in the covariance matrix. The purpose of the internship is to implement and evaluate both ideas. Basic knowledge of statistical inference and testing procedures will be appreciated.

Supervisor: Dr Marie Chion