

Project A:

To be approved for use in humans, a new therapy must first be investigated through a series of clinical trials. While the hope is to identify a dosage of the therapy which is effective, it is paramount that it be safe for patients as well. Dose-escalation trials seek to identify a maximally tolerated dose (MTD), defined as the largest dose that satisfies all safety requirements. These trials characterize the dose-toxicity curve, which estimates the probability of toxicity at each dose level. A dose-escalation trial first evaluates a cohort of patients at the lowest dose, and then sequentially assigns new cohorts to successively higher doses until a MTD is identified. While the literature has explored the impact of many components of dose-escalation designs on their performance, the mathematical transformation applied to the specific dose levels used in the trial has received comparatively less attention. This dose transformation changes the shape of the dose-toxicity curve, and some transformations may result in curves where the MTD is easier to identify than in others. The aim of this internship project is to investigate the impact of different dose transformations on the performance of dose-escalation designs. Via regular interaction with pharmaceutical industry collaborators, the intern will implement a realistic Bayesian dose-escalation design. They will then explore the research question in depth by comparing existing dose transformations and developing a novel transformation that offers more robust performance. At the end of the internship, the student will write a short report summarizing their work and will present their findings to researchers at the BSU.

Supervisor: Dr James Willard (Efficient Study Design Theme)

Project B:

This internship project focuses on the development of new response-adaptive designs for clinical trials for rare diseases, where such trials may be the only opportunity for patients to access promising therapies. While traditional clinical trials assign patients to different treatments using fixed allocation probabilities, response-adaptive designs dynamically adjust these probabilities based on accumulating data, enhancing trial efficiency and patient benefit. This project will specifically explore applications in oncology trials with survival endpoints. The intern will learn foundational concepts in clinical trials and engage with the literature on response-adaptive designs and statistical techniques for handling survival data. The ultimate goal will be to develop an efficient response-adaptive design for survival endpoints and conduct simulation studies to evaluate its strengths and limitations across various scenarios.

Supervisor: Dr Gianmarco Caruso (Efficient Study Design Theme)

Project C:

Integration of high-dimensional features for cancer detection and monitoring using cell-free DNA

This research project offers the opportunity to investigate circulating cell-free DNA (cfDNA) from blood samples as a non-invasive biomarker for cancer detection and monitoring [1, 2].

Liquid biopsy techniques have revolutionized oncology by enabling the analysis of circulating tumour DNA - small DNA fragments shed by tumours into the bloodstream. By studying changes in copy number aberrations (CNAs) and fragmentation patterns [3], we can gain insights into cancer detection [4] and treatment response [2].

However, most studies utilize one-dimensional summary statistics indicative of global changes. A complementary strategy involves examining local genome-wide enrichment metrics.

This project will focus on:

1. Investigating genomic abnormalities in cfDNA using local genome-wide enrichment metrics.
2. Applying statistical and computational approaches to generate high-dimensional features.
3. Exploring ways to integrate these features into machine learning models for improved cancer detection.

This is an exciting opportunity to develop skills in bioinformatics, statistical modelling, and cancer genomics while contributing to innovative research in liquid biopsy analysis.

For more details, contact solon.karapanagiotis@mrc-bsu.cam.ac.uk

References

- [1] <https://pubmed.ncbi.nlm.nih.gov/30380390/>
- [2] <https://www.biorxiv.org/content/10.1101/2023.03.03.530936v1>
- [3] <https://doi.org/10.1093/noajnl/vdac066>
- [4] <https://doi.org/10.1038/s41586-019-1272-6>

Supervisor: Dr Solon Karapanagiotis (Precision Medicine Theme)

Project D:

In a randomized clinical trial where the effect of an experimental treatment is compared against a placebo on an outcome variable, e.g. the effect of some new medication on blood pressure, the standard procedure is to randomly allocate half of the patients to the experimental treatment and half of the patients to the placebo. Hopefully, the result after the clinical trial is finished show that the studied treatment is more effective than the placebo so that future patients with high blood pressure issues receive good treatment. However, half of the patients in the clinical trial received the placebo, meaning that they were not treated for their condition. To circumvent this issue, an alternative approach is to use a response-adaptive randomization procedure, i.e. a design where the allocation probabilities to the different treatments are updated throughout the trial based on the outcomes on previous patients. This is intended to produce trials where more patients are allocated to the best treatment. A problem with this type of design is that traditional inference methods does not work as the assumption of iid observations is violated. Thus, alternative methods for estimation and hypothesis testing are needed. In this project, the intern, together with the supervisor, will investigate a new type of hypothesis test that uses the allocation probabilities to the different treatment instead of the outcome variable in the two-arm setting (treatment vs placebo). The intern will also study how these tests are affected by design choices for the trial and the underlying data model. Test properties that are of interest are, for example, test power, type I error, critical values and rejection

regions, null and alternative distributions of the test statistics. The intern will investigate how these properties change with, for example, the sample size of the clinical trial, the underlying data models (e.g. success rates for different treatments and trends), and the treatment allocation procedure. The investigations will be done through simulations, and the details and setup of the simulation study will be part of the intern's responsibilities in agreement with the supervisor.

Supervisor: Dr Stina Zetterstrom (Efficient Study Design Theme)

Project E:

Gene expression data analysis is one of the most exciting topics in bioinformatics/biostatistics. Its aim is the understanding of complex diseases via the discovery of patterns of gene activity across patients. Among existing tools, traditional clustering techniques offer a way of identifying these patterns by revealing relevant groups of samples. However, they are normally based on the entire set of genes (i.e., features), which may miss clusters related to a key subset of disease-relevant genes.

Biclustering addresses this by grouping subsets of samples and genes simultaneously, uncovering patterns across certain genes that differ between samples, helping define subtypes of diseases. Unfortunately, this is still a challenging task, particularly in commonly found large gene expression datasets, which often include tens of thousands of genes but typically only hundreds of samples. To improve this, we developed an outcome-guided approach, which integrates disease outcomes to better inform the biclustering task.

The main objective of this internship is to explore other alternatives for incorporating disease outcomes. Our goal is for the intern to compare and test existing machine/deep learning classifiers that may capture disease-related profiles not detected by our proposed approach. The intern will evaluate these classifiers based on computational cost, accuracy, and performance. Finally, the intern will recommend whether our approach might be improved or not, and provide a detailed comparison of the tested machine/deep learning classifiers to support their decision.

Supervisor: Dr Luis Vargas-Mieles (Causal Mechanisms Theme / CITIID)

Project F:

Recent innovation in trial design to improve study efficiency has led to the development of basket trials in which a single therapeutic treatment is tested on several patient populations, each of which forms a 'basket' where patients across all baskets will share a common genetic marker/trait that the treatment is targeted towards. To date only a handful of basket trials have implemented a control arm. Several challenges exist for defining a control arm in this setting. The goal of the project is to explore the issues of incorporating control arms into the basket trial design. The primary goal is to extend existing Bayesian methodology (which is implemented to improve statistical power in the case of small sample sizes) to the randomised setting and to gain an insight on how this will impact inference. This will primarily involve conducting thorough simulation studies implemented using R.

Supervisor: Dr Libby Daniells (Efficient Study Design Theme)

Project G:

The *randomized controlled trial* (RCT) is a well-established method for generating reliable evidence to advance medical knowledge. Although many variants exist, the most common RCT design randomly assigns participants to one of the treatments under evaluation (typically with equal probability), where one of the treatments is termed the control (placebo or the standard of care). After some time, outcome variables are collected from the participants and compared across treatment groups to determine whether any of the treatments showed a significant benefit over the control.

While providing advantages in implementation and credibility, the above design has been criticized for its rigidity: in settings where outcomes are controversial (e.g., survival or early pre-term birth) or in rare diseases, there is an incentive to minimize the number of trial participants assigned to inferior treatment(s). An alternative is to use a *response-adaptive* (RA) design, where the proportion of participants allocated to a treatment changes based on interim data analyses.

Most of the literature on RA designs focuses on trials with a single outcome, however an analysis of multiple outcomes might be more suitable in certain settings: in disease settings such as cardiology, the health status of a patient cannot be adequately represented by a single primary endpoint. The use of multiple outcomes can furthermore increase inferential precision and yields a possibility for participants to elicit preferences in the outcomes.

The goal of this internship is to design, evaluate, and compare response-adaptive designs based on the method of *generalized pairwise comparisons* (GPC). The method of GPC determines one (benefit) statistic from multiple outcomes by comparing realisations for each outcome across treatment groups and combining the results of each comparison.

After an initial period where the intern can get acquainted with RA designs, GPC, and initial ideas/methods to combine the two, resulting in a literature summary, the intern will compare operating characteristics (e.g., power or expected outcomes) of several GPC-based RA designs (preferably including their own developed designs) using simulation studies. Interns with a mathematical interest could furthermore aim to derive asymptotic properties of tests or estimators under GPC-based RA designs.

Supervisor: Dr Stef Baas (Efficient Study Design Theme)

Project H:

Simulation-Based Evaluation of Extended Blood Matching Across Patient Groups

Transfusion can be lifesaving for patients with cancer or inherited red blood cell (RBC) disorders. However, if the donated blood is not well matched to the patient, the patient could form alloantibodies to foreign antigens from the donor RBCs. When this occurs, further exposures may lead to severe illness or even death in the patient.

Providing well matched blood is often not straightforward for several reasons, especially when the patient group has a different antigen profile to the donor base.

This project focuses on the problem of allocating blood for transfusion. It extends previous research by utilising the assessment of NHS and NHSBT clinical experts regarding the likelihood and severity of transfusion reactions in regularly transfused patient groups (e.g., patients with sickle cell disorder, or myelodysplastic syndromes).

The patient group preferences given by the clinical experts will be incorporated into a model of the blood supply chain in England, with which various blood matching and allocation scenarios can be simulated to investigate the impact of these preferences on patient outcomes. This work will contribute to the ongoing computational implementation and evaluation of blood matching models for use in the NHS.

Supervisor: Dr Folarin Oyebolu (Precision Medicine Theme)
