Project Details	
Project Code	PHS27Ex McKinley
Title	Advancing Bayesian machine learning methods for precision medicine applications in Type 2 diabetes
Research Theme	Population Health Sciences
Project Type	Dry lab
Summary	It is well-known that the efficacy of treatments for a specific disease or condition can vary across individuals, so that in settings where multiple treatment options are available, different individuals may require different treatments to obtain the best outcome. Precision medicine methods leverage individual-level characteristics to help optimise treatment choices for individuals. This project will leverage recent advances in Bayesian statistical and machine learning methodology to help deal with key challenges in developing such models in large-scale observational electronic healthcare record data. These models will be applied to important real-world applications in diabetes treatment selection.
Description	Type 2 diabetes (T2D) is a chronic health condition characterised by high blood sugar levels, potentially leading to long-term microvascular and macrovascular complications (such as heart attacks, strokes, blindness and kidney disease). Glucose levels are regulated by insulin; a hormone that allows cells to absorb glucose and use this for energy. T2D can be caused by either a reduction in insulin-production, or by the body becoming insulin-resistant. The treatment goal is to reduce and stabilise blood sugar levels. Once the first-line Metformin therapy is deemed insufficient to control glucose and potentially mitigate other cardiometabolic risks, a range of additional second-line treatments are considered. Current treatment guidelines do not support targeting treatment based on glucose-lowering response, although there are a range of recently developed treatment selection models that show great utility in this area (Dennis et al., 2025). These are sophisticated statistical and machine-learning models that utilise individual-level characteristics to help optimise treatment choices for individuals. A key challenge with building such models is that there is a paucity of direct head-to-head clinical trial data for comparing glucose response across different competing drug treatments. Instead these models must be developed on observational (usually large-scale electronic healthcare record) data. These datasets potentially suffer from key biases, such as indication bias, caused by the presence of unobserved confounding variables (i.e. ones that affect both the glucose-lowering response and also the probability of an individual being given a treatment in the first place). Many treatment selection methods ignore these potential biases, because they are hard to deal with. This project will leverage recent advances in Bayesian methodology to help deal with the challenge of building these models and dealing with unobserved confounding, which arises because the treatment a patient receives isn't randomised and may depend on a multitude of factors, not all of which will be appropriately measured, meaning that standard regression adjustment methods could provide erroneous conclusions. To address this, we will build joint models between the key outcomes-of-

	interest (e.g. blood glucose, or longer-term cardio-vascular or renal outcomes), and the treatment propensity score (the probability that an individual is given the treatment). This framework allows for unobserved latent confounding variables to influence treatment choice via flexible and structured joint residual processes, using recent advances such as Dirichlet Process Mixture Models. This PhD will explore the utility of such approaches applied to different models for optimising treatment selection for T2D. This includes how to build such models utilising recent advances in Bayesian variable selection techniques, or flexible and powerful machine learning methods, such as Bayesian additive regression trees. Furthermore, there are associated challenges in dealing with non-ignorable missing data, which can be conceptualised in a similar way utilising latent variables and joint models between the outcome and the probability-of-missingness, so the project could be extended into this direction as well depending on the student's preference. The successful student will gain an opportunity to learn state-of-the-art computational methods for developing precision medicine models, build experience working with, visualising and modelling complex electronic healthcare record data applied to important real-world problems. Some initial objectives are given below, but the student will have opportunities to steer the project depending on which aspects of the work they feel are the most interesting and potentially fruitful. Objective 1: Perform a literature review to identify recent advances in Bayesian methods that could be applied to these problems. Also examine different software options for implementing these ideas, for example: Stan/Nimble/PyMC-BART. This will give the student a good grounding in these methods and ideas, and their implementation. Objective 2: Develop methodological ideas and perform a simulation study to examine the performance of the proposed methods under a range of different scenarios, comparing the results to existing approaches such as marginal treatment effect, and person-centred treatment effect estimation. These simulations will be used to understand the strengths and weaknesses of the proposed approaches, and how the results compare to existing methods. Objective 3: Apply the methods developed in Objective 2 to treatment selection models for Type 2 diabetes. There are various directions this could take depending on the interests of the student and the results of Objective 2. A first step would be to apply these ideas to different two-drug comparisons and compare the results to models that have already been developed in Exeter. A subsequent step would be to explore extending these ideas to a much more complex five-drug comparison, or to extend to different types of models, such as binary or survival outcome models. Objective 4: Expand recent advances in Bayesian additive regression tree (BART) models for modelling two-drug comparisons into the five-drug case. This is novel both methodologically and will also require a bespoke code base to be developed. One option here is the new PyMC-BART package, which allows BART priors to be implemented within customised models.

Can be completed part time?	Yes
Supervisory Team	
Lead Supervisor	
Name	Dr Trevelyan McKinley
Affiliation	Exeter
College/Faculty	Faculty of Health and Life Sciences
Department/School	Clinical and Biomedical Sciences
Email Address	t.mckinley@exeter.ac.uk
Co-Supervisor 1	
Name	Professor Rhian Daniel
Affiliation	Cardiff
College/Faculty	Biomedical and Life Sciences
Department/School	Population Medicine, School of Medicine
Co-Supervisor 2	
Name	Professor Jack Bowden
Affiliation	Exeter
College/Faculty	Faculty of Health and Life Sciences
Department/School	Clinical and Biomedical Sciences
Co-Supervisor 3	
Name	Dr John Dennis
Affiliation	Exeter
College/Faculty	Faculty of Health and Life Sciences
Department/School	Clinical and Biomedical Sciences