Project Details	
Project Code	PHS27Ex Young
Title	Data-Driven Approaches to Understanding Heterogeneity in Type 1 Diabetes: Towards Better Prediction and Personalised Care
Research Theme	Cardiometabolic Disease
Project Type	Dry lab
Summary	Over 400,000 people in the UK live with type 1 diabetes. The condition is highly heterogeneous, with substantial variation in age at diagnosis, disease progression, treatment response, and risk of long-term complications. As disease-modifying therapies emerge and precision medicine become increasingly achievable, understanding this heterogeneity is critical. This PhD will use large-scale routinely-collected healthcare data and cutting-edge statistical approaches to characterise disease trajectories, develop risk prediction models, and identify factors associated with differential treatment outcomes. The findings will improve understanding of long-term outcomes in type 1 diabetes and provide evidence to inform risk stratification and personalised approaches to clinical care.
Description	Background Type 1 diabetes affects approximately 400,000 people in the UK(1) and is characterised by marked heterogeneity in age at diagnosis, disease trajectory, treatment response, and long-term complications. Despite universal insulin therapy, only one-third of adults achieve recommended glycaemic targets(2), and people with type 1 diabetes remain at substantially increased risk of renal, cardiovascular and other serious complications. There is growing interest in better understanding differences in risk to support more personalised approaches to care. Adjunctive glucose-lowering therapies used in type 2 diabetes are increasing used in type 1 diabetes, but evidence from real-world populations remains limited and it is unclear which individuals are most likely to benefit. Large-scale routinely collected healthcare data now provide an opportunity to address this gap. CPRD Aurum includes over 150,000 individuals with type 1 diabetes with linked hospital and mortality data, enabling long-term follow-up in routine care. UK Biobank additionally provides biochemical measures such as C-peptide in a subset of individuals, enabling exploration of residual insulin secretion as a potential modifier of disease progression and treatment response. These data, combined with advances in statistical learning and machine learning, enable improved modelling of disease trajectories and risk prediction, but have not yet been fully applied in UK type 1 diabetes populations. (1) Chan et al. Lancet. 2021;396:2019–2082. (2) NHS England. National Diabetes Audit 2021–22. Aim To characterise heterogeneity in type 1 diabetes progression, treatment response and long-term outcomes using large-scale routinely collected healthcare data. Objectives 1. To describe variation in demographic and clinical characteristics at diagnosis of type 1 diabetes in UK routine data.

	2. To characterise longitudinal disease trajectories, including glycaemic control, treatment progression, and development of microvascular and macrovascular complications. 3. To examine heterogeneity in treatment outcomes, including response to adjunctive glucose-lowering therapies where used in routine care. 4. To develop and internally validate risk prediction models for key long-term outcomes, including complications and treatment response. Study population The primary cohort will consist of individuals with type 1 diabetes identified in CPRD Aurum (~150,000 individuals), linked to inpatient hospital and ONS mortality data. CPRD provides longitudinal primary care records enabling assessment of diagnosis, treatment, and outcomes across time. A secondary cohort will use UK Biobank participants with type 1 diabetes and available C-peptide measurements (~27,000 individuals), enabling investigation of residual endogenous insulin secretion as a biological modifier of disease progression and treatment response. A comparator population without diabetes may be used where appropriate for estimation of excess risk and contextualisation of absolute risks. Patient features The following routinely-available features will be assessed as predictors of outcomes and for modelling heterogeneity in risk: • Age at diagnosis and current age • Sex, ethnicity, and socioeconomic deprivation • Glycaemic control (HbA1c and longitudinal trajectories) • Diabetes duration and treatment history (including insulin regimens and adjunctive therapies) • Comorbidities • Prior complications and healthcare utilisation • Clinical measurements and laboratory tests • Use of diabetes technologies (e.g. continuous glucose monitoring, closed-loop systems) • Diabetes-related prescribing patterns and care processes • C-peptide (UK Biobank subset) as a marker of endogenous insulin secretion Clinical outcomes will include: • Diabetic ketoacidosis and hypoglycaemia requiring emergency care • Microvascular complications (retinopathy, nephropathy, neuropathy) • Cardiovascular complications (e.g. myocardial infarction, stroke, heart failure) and kidney disease progression • Glycaemic response to adjunctive therapies used in routine care Methodology To model longitudinal disease trajectories, treatment response, and clinical outcomes in type 1 diabetes, the student will apply advanced epidemiological and statistical methods, including survival analysis and longitudinal modelling of glycaemic control, treatment progression, and complication risk over time.

	There will also be scope for the application of artificial intelligence and machine learning methods to better capture complex, non-linear patterns in disease progression and improve prediction of long-term outcomes. Prediction model development will follow TRIPOD+AI guidance. Initial models will use regression approaches, with comparison to machine learning methods to assess potential improvements in predictive performance. Model performance will be evaluated using discrimination, calibration, and clinical utility measures. Internal validation will use temporal and regional splits within CPRD. Student ownership The student will be supported to refine specific research questions and modelling strategies within the overall framework. There is flexibility to focus on particular outcomes (e.g. complications, treatment response) or methodological approaches, including advanced machine learning methods for longitudinal prediction and causal inference frameworks for treatment effects.
Can be completed part time?	Yes
Supervisory Team	
Lead Supervisor	
Name	Dr Katherine Young
Affiliation	Exeter
College/Faculty	Faculty of Health and Life Sciences
Department/School	Clinical and Biomedical Sciences
Email Address	k.young3@exeter.ac.uk
Co-Supervisor 1	
Name	Professor Angus Jones
Affiliation	Exeter
College/Faculty	Faculty of Health and Life Sciences
Department/School	Clinical and Biomedical Sciences
Co-Supervisor 2	
Name	Professor John Dennis
Affiliation	Exeter
College/Faculty	Faculty of Health and Life Sciences
Department/School	Clinical and Biomedical Sciences
Co-Supervisor 3	
Name	Dr Charlotte James
Affiliation	Bristol
College/Faculty	Bristol Medical School
Department/School	Population Health Sciences