YSM 2026 BOOK OF ABSTRACTS

Cambridge, UK

These are the abstracts presented at the Young Statisticians' Meeting 2026 held at the University of Cambridge 29th July-31st July

YSM Organising Committee
youngstatsmmeeting@gmail.com

Table of Contents

Oral presentation abstracts *4*

Genomics **4**

Adaptive Bayesian computation for efficient biobank-scale genomic inference 4

Detecting Rare Immune Profile through Interpretable Factor Analysis of Single-Cell Flow Cytometry Data 5

scFLAME: joint dimensionality reduction, clustering and marker gene identification for single-cell RNA-seq data 6

Finance **7**

A DSL-Based Approach for Reproducible National Accounting 7

Meta-Analyses and Metropolis **8**

A scaling limit for the Random Walk Metropolis with heavy-tailed proposals 8

Back-Transformations in Random-Effects Meta-Analysis - Impact and Interpretation 9

An R Package for Data Extraction in Systematic Reviews and Meta-Analyses Using Large Language Models 10

Environmental **11**

Accounting for Human Movement to Improve Exposure-Health Models 11

Bayesian statistical inference of locations and magnitudes of historical earthquake from archive data 12

Designing temporal structure in Bayesian disaggregation regression for infectious disease mapping 13

Large-graph approximations of coloured random graphs 14

Population Data Analysis **15**

A Novel Nested Case–Control sampling strategy for time-to-event analysis of large-Scale cohort data 15

All-cause mortality assessment at the point of diabetic eye screening: a streamlined approach to individualised preventive care 16

Improving Nested Exposure Case-Control Designs for Rare Time-Dependent Events via Counter-Matching 18

Quantifying the role of stage at diagnosis in racial disparities in US cancer survival: a population-based SEER study for 8 common cancers 19

Clustering **20**

Outcome-guided clustering within a generalised Bayes framework 20

A stability-driven clustering framework for clinical data: application of paediatric Hypertrophic Cardiomyopathy identifies patient groups with distinct outcomes 21

MultiCOCor: identifying multiple clustering structures in high-dimensional data 22

Bayesian **23**

A Bayesian Grid-Search Approach for Flagging Potentially Atypical Variants in Cancer Susceptibility Genes 23

Time-Dependent Bayesian Borrowing from Historical Controls in Randomized Clinical Trials with Temporal Drift 24

Using Bayesian K Regression Modelling and Bayesian Model Averaging to Model Frailty Indices and Reduce Risk of Misdiagnosis 25

Bayesian Simultaneous Credible Bands for Polynomial Regression 26

Longitudinal data **27**

Hierarchical Clustering for Sparse Longitudinal Data: Extensions for Continuous and Ordinal Outcomes 27

Comparing joint recurrent event modelling frameworks for dynamic prediction in chronic conditions: an empirical evaluation 28
Modelling the effect of Longitudinal Markers on Left-Truncated Time-to-Event Outcomes in Twin Studies 29
Investigating student dropout and performance using three-part mixed-effects quantile regression models for multivariate longitudinal data 30

Prediction Modelling 31

Who Benefits from Intensive Treatment in Psoriatic Arthritis? Deriving Personalised Treatment Rules using Causal Forests and Policy Trees 31
Quantifying Individual-level Uncertainty in the BOADICEA Breast Cancer Risk Prediction Model 33
A Statistical Framework for HPA Axis Adjacent Hormone Trajectory Analysis 34
Comparison of propensity score methods for a continuous exposure: application to causal inference in two Parkinson's disease cohorts 35
Poore, Mia, NHS Blood and Transplant/University of Bristol; Nodehi, Anahita, University of Bristol; Lawton, Michael, University of Bristol; Ben-Shlomo, Yoav, University of Bristol 35

Lightning talks 36

Deterministic vs Agent-Based Models of Influenza Spread: The Role of Stochastic Extinction in Small Populations 36
Improving 10-year cardiovascular risk prediction in patients with established cardiovascular disease: flexible addition of risk predictors on top of the SMART2 risk score 37
How to Evaluate Imputation Methods for Missing Data in Medical Research: A Comparative Simulation Study 38
Methods for clinical risk prediction under interventions: a comparative simulation study 39
Randomised Comparison of Ponatinib Versus Imatinib in Combination with Chemotherapy in Patients 55 Years of Age and Older with Newly Diagnosed Ph+ ALL: Primary Outcome Analysis of the EWALL PH03 Study 40
Joint modelling of longitudinal biomarkers and time to healing in venous leg ulcers 41
Joint Bayesian modelling of antibody and PCR test data 42
The cost of electronic bedside transfusion checks and its impact on the whole hospital transfusion pathway in England 43
Automated marking of Very Short Answer Questions in an applied health setting 44

Posters 45

Model selection tests for truncated vine copulas under nested hypotheses 45
Augmenting Randomized Trials With Real-World Controls for Negative Binomial Endpoints Using Bayesian Mixture Priors 46
Exploration and Exploitation under Uncertainty: A Comparative Simulation Study of the Thompson Sampling Family in Multi-Arm Trials 47
From Hybrid to Fully Bayesian: A Simulation Study Evaluating Operating Characteristics and Efficiency in an Adaptive Platform Trial 48
Recurrent Hamiltonian Directional Echo Learning (RHDEL): A Time-Reversible Framework for Memory-Efficient EEG Classification 49
Climate-driven modelling of dengue incidence using statistical and machine learning approaches 50
Machine Learning-Based Prediction of Paediatric Kidney Graft Survival: A Multi-Model, Multi-Horizon Analysis of the UNOS Registry 51
A joint meta-analysis framework for the accuracy of two diagnostic tests accounting for varying study designs 52
Modelling Multiplayer FPS Player Performance Using Hidden Markov Models 53
Revising Use of F-Tests in Weak Instrument Practice 54

Collaborative Targeted Maximum Likelihood Estimation for High-Dimensional Confounding: A Simulation Study and Julia Implementation 55
Quantifying the Long-Term Quality-of-Life Impact of ABO-Incompatible Transfusion Events Using SHOT Data and Clinical Evidence 56
Decomposing the Uncertainty of Dynamic Risk Predictions using Value of Information 57
A note on diffusive/random-walk behaviour in Metropolis-Hastings algorithms 58
Modeling Prediction Markets as Defaultable Bonds 59
Re-randomisation tests in clinical trials: a simulation case study and practical guidance 60
Methods for Informative Priors via Predictives using Predictively Oriented and Generalised Bayes 61
Divide and conquer sequential Monte Carlo for chained Markov melding 62
Equitable cardiovascular risk stratification in Type 2 Diabetes: development of a population-specific prediction model in a large, ethnically diverse London cohort 63
Predicting MACE and Mortality in Type 2 Diabetes: A DeepHit Competing Risks Model in North West London EHR Data 64
Using individual participant data meta-analyses to explore the effectiveness of weight loss interventions for people with knee OA 65
Progression of Multiple Long-Term Conditions in Type 1 and Type 2 Diabetes: Insights from A CPRD Matched Cohort Study 66

Oral presentation abstracts

Genomics

Adaptive Bayesian computation for efficient biobank-scale genomic inference

Yiran Li, MRC Biostatistics Unit, University of Cambridge; John Whittaker, MRC Biostatistics Unit, University of Cambridge; Sylvia Richardson, MRC Biostatistics Unit, University of Cambridge; Hélène Ruffieux, MRC Biostatistics Unit, University of Cambridge

Background: Modern biobanks, with unprecedented sample sizes and phenotypic diversity, have become foundational resources for genomic studies, enabling powerful cross-phenotype and population-scale analyses. As studies grow in complexity, Bayesian hierarchical models offer a principled framework for jointly modelling multiple biological dimensions or units', such as cells, traits and experimental conditions, increasing statistical power through information sharing. However, adoption of such models in biobank-scale studies remains limited due to computational inefficiencies, particularly in posterior inference over high-dimensional parameter spaces. Deterministic approximations such as variational inference provide scalable alternatives to Markov Chain Monte Carlo, yet current implementations do not fully exploit the structure of genome-wide multi-unit modelling, especially when biological effects are sparse, i.e., concentrated in a small number of units.

Results: We present AF-CAVI, an adaptive focus (AF) scheme within a block coordinate ascent variational inference (CAVI) framework that selectively updates subsets of parameters at each iteration, corresponding to units deemed relevant based on current estimates. We also develop a genome-wide multi-response pQTL mapping pipeline based on AF-CAVI–accelerated joint Bayesian modelling and demonstrate its performance on simulated data and selected regions of UK Biobank chromosome 1. AF-CAVI achieves a ≈50% runtime reduction compared to state-of-the-art CAVI with maintained statistical performance and improved power over univariate mapping.

Conclusions: AF-CAVI makes joint Bayesian modelling practical for biobank-scale multi-trait inference by controlling computational cost while retaining statistical performance. More broadly, it provides a flexible computational framework for high-dimensional multi-unit models with sparse signals, supporting scalable genomic discovery across traits and conditions.

Detecting Rare Immune Profile through Interpretable Factor Analysis of Single-Cell Flow Cytometry Data

Zhang, Xinzhi, MRC Biostatistics Unit, University of Cambridge; Kwong, Kristy, Department of Medicine, University of Cambridge; Cook, Matthew, Department of Medicine, University of Cambridge; INTREPID; Wallace, Chris, MRC Biostatistics Unit, University of Cambridge; Ruffieux, Hélène, MRC Biostatistics Unit, University of Cambridge

Background: Flow cytometry data is a powerful tool for studying the immune system, as it enables the simultaneous characterization of heterogeneous immune cell populations. Understanding the sources of heterogeneity is essential for gaining insight into immune disease. Despite this advantage, flow cytometry data are often collected across multiple experimental batches (e.g., different acquisition days), which can introduce systematic measurement bias and technical variation. Distinguishing technical variation from true biological signals is therefore a central challenge for reliable downstream analysis.

Methods: In this study, we analyzed Peripheral Blood Mononuclear Cells (PBMCs) from a heterogeneous cohort, including healthy controls and patients with various immune diseases. Protein expression levels were measured at the single-cell level, resulting in a large-scale dataset. Our analysis consisted of several steps: (1) density-based downsampling to reduce sample size while preserving rare cell populations; (2) we adopted a batch-specific sparse factor analysis framework in which some factors can be shrunk in certain batches while remaining active in others, allowing latent factors to capture either batch-related effects or biological signals; and (3) downstream analyses to identify outlier samples associated with distinct immune pathways.

Results: We found that batch-specific factor analysis performs better than standard approaches on unnormalized data where batch effects remain uncorrected. In this framework, some factors primarily reflect technical variation, while others correspond to immune pathways, as interpreted from protein loadings. Importantly, rare immune profiles were identified not only based on cell-level factor scores but also through shifts in the mean and the variance of factor scores within each sample. A shifted mean or elevated variance in biologically interpretable factors suggests dysregulation of the corresponding immune pathways. In addition, we are investigating methods to capture non-linear relationships among biological factors (e.g., using t-SNE) to improve the grouping and characterization of rare samples.

Conclusions: Sparse factor analysis provides an interpretable framework for disentangling technical and biological variation in flow cytometry data and enables the detection of biologically meaningful rare immune profiles.

scFLAME: joint dimensionality reduction, clustering and marker gene identification for single-cell RNA-seq data

Rao, Jackie, MRC Biostatistics Unit, University of Cambridge; Kirk, Paul D.W., MRC Biostatistics Unit, University of Cambridge

Background: Single-cell RNA sequencing enables exploration of cellular transcriptional heterogeneity at unprecedented resolution. In particular, cluster analysis can reveal cell heterogeneity and allow for distinct cell populations to be identified. However, clustering single-cell data is often a statistical challenge due to high dimensionality, the count-based nature of the data, and the presence of biological and technical noise. Many computational approaches first perform dimensionality reduction to mitigate these challenges, and clusters are often inferred using heuristic methods, limiting interpretability.

Methods: We introduce scFLAME (single-cell Factor Latent Analysis with Mixture Embeddings), a probabilistic generative model that jointly performs dimension reduction and clustering of the data through a Gaussian mixture prior over a low-dimensional latent space. By learning a latent space directly associated with cluster structure, scFLAME provides interpretable embeddings while enabling identification of cluster-specific marker genes via a linear decoder that maps latent factors to gene expression counts through a negative binomial likelihood, without the need for post-hoc differential expression analysis. scFLAME also employs a post-clustering merging strategy to reveal hierarchical structure, allowing users to identify both broad cell types and finer subtypes within complex single-cell datasets.

Results/Conclusions: Experiments on simulated and real scRNA-seq datasets show that scFLAME achieves competitive clustering performance compared to existing methods while yielding biologically interpretable latent spaces and coherent gene programs. The model provides a scalable and unified approach for analysing transcriptional heterogeneity in single-cell data.

Finance

A DSL-Based Approach for Reproducible National Accounting

Taleb, Iliass., Direction de la Comptabilité Nationale, Haut-Commissariat au Plan (HCP), Morocco

Background: Supply–Use Tables (SUTs) are a central component of national accounting systems used to ensure consistency in key macroeconomic indicators such as Gross Domestic Product (GDP). They reconcile production, consumption, imports, exports, and other uses across all sectors of the economy. In practice, compiling these tables is a complex and iterative process that often relies on graphical interfaces or spreadsheets, requiring repeated manual adjustments. Changes in assumptions frequently force analysts to redo large parts of the balancing process, with limited reuse of previous work. In addition, detailed economic reports must often be produced alongside the balancing under strict deadlines.

Methods: We propose AccScript, a domain-specific language (DSL) designed to formalize the compilation and balancing of SUTs within a declarative framework. Instead of manual adjustments, users express economic relationships, assumptions, and balancing rules as structured code, capturing the analytical logic applied by national accountants. The language reflects the System of National Accounts (SNA 2008) and uses intuitive, domain-specific syntax. AccScript is implemented through a pipeline including parsing, validation, and execution, and incorporates reusable primitives for operations such as volume–value transformations, price index application, tax and margin allocation, and balancing procedures.

Results: Applied to realistic scenarios, the framework reduces manual intervention and enables reuse of analytical logic across iterations. It ensures reproducibility, improves auditability through structured error detection, and allows efficient testing of alternative assumptions without restarting the full process. Overall, the approach leads to significant reductions in time and manual effort.

Conclusions: AccScript demonstrates how statistical computing and formal language design can modernize SUT compilation. It provides a reproducible, auditable, and efficient alternative to current practices, while supporting knowledge transfer and standardization across institutions.

Meta-Analyses and Metropolis

A scaling limit for the Random Walk Metropolis with heavy-tailed proposals

Liu, Yuxin, UCL; Livingstone, Samuel, UCL; Mijatovic, Aleks, Warwick

Background: Random Walk Metropolis (RWM) is a widely used MCMC sampling method, but its efficiency depends critically on the scaling of the proposal distribution. If the scale is too small, each proposed move is tiny and convergence is slow; if too large, proposals are frequently rejected, again slowing convergence. It is already known that when the proposal is Gaussian and the dimension tends to infinity, the scaling limit of RWM is a Langevin diffusion multiplied by a speed function. This result provides both a practical guideline, choose the scale that maximises the speed function and theoretical insight into the dynamics of the algorithm. However, Gaussian proposals perform poorly when the target distribution is heavy-tailed, since the light-tailed proposals fail to explore the tails efficiently. This motivates the use of heavy-tailed proposals, such as the increments of a general Lévy process proposal, and raises the natural question of what the corresponding scaling limit is in this setting.

Methods: We derive the generator of RWM with a heavy-tailed distribution with dimension going to infinity. We identify the speed function and the core generator, determine the optimal scaling that maximises the speed function, and characterise the limiting process by finding a stochastic process with the same generator, thereby gaining insight into the dynamics.

Results: As the dimension tends to infinity, the generator of RWM with increments of a general Lévy process as proposal is related to a hidden Lévy process. The limiting process decomposes into three components: a continuous part driven by a Brownian motion, a drift part, and a jump part driven by a Poisson point process.

Conclusions: We characterise the limiting process explicitly and develop a deeper understanding of the RWM dynamics in the heavy-tailed setting.

Back-Transformations in Random-Effects Meta-Analysis - Impact and Interpretation

Igelmann, Jan-Bernd, TU Dortmund University; Pauly, Markus, TU Dortmund University and Research Center Trustworthy Data Science and Security UA Ruhr; Viechtbauer, Wolfgang, Maastricht University

Background: In meta-analyses, effect size measures with bounded or non-normal sampling distributions are commonly analyzed on a transformed scale to justify normality assumptions. While point estimates and confidence intervals (CIs) are routinely back-transformed to the original scale for interpretation, this step is nontrivial in random-effects models. In practice, simple inverse transformation functions are often applied and the resulting estimates are presented as the mean effect sizes, although they are not means in general.

Methods: We highlight that, due to Jensen's inequality, such back-transformations generally yield the median rather than the mean effect size. To address this issue, we formulate a general integral-based back-transformation that recovers mean effect sizes on the original scale, along with corresponding confidence intervals. Our framework is applicable across a wide range of transformation functions and we further discuss back-transformation approaches for transformations that do not map onto the entire real line.

Results: Through analytical investigations and empirical examples across different effect size measures and transformations, we demonstrate that, although integral back-transformations are required to obtain correct mean estimates, they must be used with caution. In particular, they are sensitive to heterogeneity estimation and may become unstable for unbounded transformations. Moreover, certain asymmetric transformations can lead to inconsistent inference results.

Conclusions: Overall, we emphasize that the choice of back-transformation should be guided by the analyst's objective. The standard inverse back-transformation remains well suited for descriptive purposes and is often preferable in practice, provided it is correctly interpreted as a back-transformed median and not mean effect size. Integral back-transformations should be used when a mean effect size is explicitly required or when for example downstream analyses rely on mean-based optimality properties. These considerations are highly relevant in meta-analysis practice and extend beyond this setting to any context where transformations and back-transformations are used, but are frequently overlooked in applied statistical work.

An R Package for Data Extraction in Systematic Reviews and Meta-Analyses Using Large Language Models

Dai, Danyang, The University of Queensland and King's College of London; Tanaka, Emi, Australian National University; Pole, Jason, The University of Queensland

Background: Systematic reviews and meta-analyses are central to evidence-based medicine, yet data extraction remains labor-intensive and prone to error. Although recent studies have shown that Large Language Models (LLMs) can assist with data extraction, their use has largely been confined to proof-of-concept applications, with substantial barriers to broader adoption among researchers.

Methods: We developed metaextractoR, an open-source R package designed to streamline data extraction for systematic reviews and meta-analyses by integrating open-source LLMs through the Ollama framework. The package consists of three modular Shiny applications implementing a human-in-the-loop workflow: (1) manual abstract-level data extraction, (2) iterative prompt engineering and model selection, and (3) validation of LLM-generated outputs. Each abstract undergoes double extraction—once by a human and once by an LLM—to mirror Cochrane Handbook recommendations. The package runs entirely on local machines without requiring APIs or external data transfer, and comprehensive logging ensures transparency and reproducibility. Due to token limitations, the current version is restricted to abstract-level extraction.

Results: In an illustrative example, seven data elements were extracted from 100 abstracts and compared with manual extraction. Using the LLaMA 3.1 model, the overall extraction accuracy was 69.3%, with the highest performance observed for identifying sample size (84.6% accuracy).

Conclusion: metaextractoR provides a principled, user-friendly, and privacy-preserving framework that lowers technical barriers to incorporating LLMs into evidence synthesis. By enabling structured, reproducible, and human-validated extraction workflows aligned with Cochrane guidelines, the package has the potential to improve the efficiency and transparency of systematic reviews. Future development will focus on supporting closed-source LLMs such as ChatGPT, which may further enhance extraction accuracy.

Environmental

Accounting for Human Movement to Improve Exposure-Health Models

Tahir, Hadiqa, University of Leicester; Smart, Simon, University of Leicester; Cai Samuel, University of Leicester; Ng, André, University of Leicester; Vande Hey, Joshua, University of Leicester; Lucas, Tim, University of Leicester

Background: Environmental exposure-health studies typically average exposures over an arbitrary circular buffer around residential locations, treating individuals as stationary and failing to account for human movement. This can lead to exposure misclassification and biased exposure-response estimates, potentially distorting our understanding of the true health effects of environmental conditions. We developed statistical models that explicitly account for human movement when linking environmental exposures to health outcomes.

Methods: We linked pixel-level environmental exposures to individual health outcomes by weighting each pixel according to distance from the home location, using distance as a proxy for human movement. Weights were either predefined or estimated through a latent distance-decay parameter inferred from the data. Model performance was assessed across multiple sample sizes (N = 1,114; 50,000; and 100,000). A simulation study examined parameter recovery using bias, empirical standard error, and credible interval coverage. As a case study, Demographic and Health Surveys data from Albania were used to link acute respiratory infection outcomes among children under five to pixel-level NDVI within a 3 km buffer around DHS cluster centroids, and the proposed models were applied to these data.

Results: Parameter recovery improved with increasing sample size. At N = 1114, estimates were biased and imprecise, with incorrect effect direction for exposure-response parameters (e.g., learned-weight model B1 bias = -0.84; EmpSE = 2.61; coverage = 0.86). In contrast, the models accurately recovered parameters at larger sample sizes, including the latent distance-decay parameter (bias = −0.02; EmpSE = 0.15; coverage = 0.95 at N = 100,000), demonstrating its ability to reliably learn movement-based exposure weights when sufficient data were available.

Conclusion: Instead of relying on arbitrarily-sized buffers, this framework provides a novel method to studying exposure-health relationships while accounting for human movement. With sufficiently large sample sizes, it can accurately estimate the influence of disaggregated environmental exposures on individual-level health and help address exposure misclassification arising from residential-only metrics. This methodology remains scalable, interpretable, and adaptable to other exposures and outcomes, offering a foundation for future work that integrates richer mobility-informed exposure-health research.

Bayesian statistical inference of locations and magnitudes of historical earthquake from archive data

Wenjin Lu; Prof Ian Main; Dr Mark Naylor; Prof Finn Lindgren and Dr Brian Baptie

Historical earthquake records in the United Kingdom rely on macroseismic intensity observations, which are sparse, irregular, and often incomplete, leading to substantial uncertainty in estimating epicentres and felt areas. Existing approaches are typically deterministic or assume radial symmetry, limiting their ability to capture spatial variability and uncertainty, particularly for offshore or poorly constrained events. In this study, we develop a Bayesian framework to reconstruct intensity fields using data from the British Geological Survey. Intensity observations are treated as ordered categorical outcomes within a cumulative logit model, with spatial variation represented through an anisotropic elliptical distance function parameterised by epicentre location, scale, orientation, and axis ratio. To improve parameter identifiability and computational efficiency, the model is reparameterised in terms of a global scale parameter and a transformed vector encoding ellipse shape and orientation. Informative priors are specified for key geometric parameters, particularly the axis ratio, to prevent unrealistic elongated solutions in data-sparse regions. The model is fitted using Markov Chain Monte Carlo implemented in JAGS, with initial values derived from maximum likelihood estimation to enhance convergence. The proposed framework produces stable and interpretable posterior estimates of epicentral location and intensity-specific felt areas, with coherent uncertainty quantification. Results indicate that the scale parameter is well constrained by the data, while the axis ratio remains close to unity for most events, supporting near-radial attenuation patterns in the UK. The approach performs particularly well for onshore events with moderate observational density, while informative priors improve robustness for coastal and offshore cases. This framework provides a flexible and physically interpretable approach for modelling historical earthquake intensity fields and offers a principled way to integrate prior knowledge with sparse observational data.

Designing temporal structure in Bayesian disaggregation regression for infectious disease mapping

Bano, Samana, NIHR/University of Leicester ; Dr Lucas, Tim, University of Leicester

Background: Routine infectious disease surveillance systems usually report annual case counts aggregated to administrative regions, while environmental covariates are available on fine spatial grids. This resolution mismatch complicates the use of rich gridded covariates in disease risk mapping. Bayesian disaggregation regression links polygon-level counts to latent pixel-level risk surfaces via an aggregation constraint, but most implementations are effectively cross-sectional and offer limited guidance on how to incorporate time without harming calibration.

Methods: Using West Nile Virus surveillance data for eastern USA counties (2003–2019) and 1 km environmental covariates, I extend a Bayesian disaggregation framework to multiple non-overlapping temporal windows. Within an INLA+SPDE model I compare four temporal specifications: (i) no time effect, (ii) a linear trend in window index, (iii) a penalised spline smooth (f(t)), and (iv) a smooth (f(t)) plus an independent window-level random effect. I repeat this comparison under several non-overlapping window schemes (including 2-, 5- and 10-year windows). Performance is assessed using WAIC, effective degrees of freedom for (f(t)), and empirical coverage and width of posterior credible intervals for both fitted county counts and held-out areas.

Results: At the time of submission, model fitting and comparison are ongoing. I will compare the four temporal specifications across the different window schemes using WAIC, effective degrees of freedom for (f(t)), and empirical coverage and width of posterior credible intervals for fitted and held-out county counts. I will also examine how posterior uncertainty and the stability of spatial patterns change as additional years are incorporated and as the temporal structure is made more flexible. These analyses will be completed before the meeting, allowing me to report which temporal formulations are supported by the data and how sensitive conclusions are to the choice of temporal aggregation.

Conclusions: This work provides a case study in temporal model design for disaggregation regression, highlighting how choices about temporal aggregation and random-effect structure impact calibration and interpretability. The framework suggests practical guidelines for extending spatial disaggregation models to multi-period settings and is designed to be incorporated into existing R tools for disease mapping.

Large-graph approximations of coloured random graphs

Dutta, Kushankur, University of Nottingham; Khuda Bukhsh, Wasiur, University of Nottingham

Background: We study large-population stochastic models motivated by applications such as epidemic spread (SIR models) and opinion dynamics, where a key quantity of interest is the proportion of individuals in each state. A common approach models the population as an interacting particle system on a random graph, such as the Erdős–Rényi or configuration model. Building on the effective degree framework of Ball and Neal (2008), we formulate a jump Markov process evolving on a dynamic network that represents the proportion of individuals of a given type over time.

Methods: We analyse the limiting behaviour of this process as the population size tends to infinity. Using tools from functional analysis, and probability theory, we establish a functional law of large numbers, showing that the stochastic process converges to a deterministic limit. We prove that the limiting process is a continuous function that satisfies an ordinary differential equation. In addition, by applying martingale techniques, we prove a propagation of chaos phenomenon, demonstrating that individual-level interactions vanish asymptotically.

Results: Our results show that, for sufficiently large populations, the effect of individual interactions averages out, and the system can be accurately approximated by a mean-field model. For the SIR model, we assume an unbounded degree distribution for the network under suitable moment conditions, which improves upon the previous bounded degree assumption.

Conclusions: The proposed general framework applies broadly to epidemic models, voter models, rumour spreading processes, and related systems in statistical physics.

Population Data Analysis

A Novel Nested Case–Control sampling strategy for time-to-event analysis of large-Scale cohort data

Amine, Ines, Environment & Health Modelling (EHM) Lab, London School of Hygiene and Tropical Medicine, London, United Kingdom; Gasparrini, Antonio, Environment & Health Modelling (EHM) Lab, London School of Hygiene and Tropical Medicine, London, United Kingdom

Background: Time-to-event analyses based on Cox proportional hazards models encounter substantial computational challenges in very large cohort datasets, especially when time-varying predictors are involved, due to the need for data expansion. Nested case–control (NCC) sampling offers a promising solution to reduce data size and computational burden, but traditional schemes can introduce bias in large datasets due to ties (i.e., simultaneous events).

Methods: We developed and compared NCC sampling strategies for large cohort data with ties, using both theoretical and empirical approaches. Two main strategies were evaluated: (1) traditional sampling of controls by event time, combined with methods for handling ties (exact, Efron, or Breslow), and (2) a novel approach in which ties are split, and controls are sampled separately for each event. We first derived the likelihoods and statistical properties of the Cox estimators under each strategy. We then conducted simulation studies to assess bias and loss of precision relative to full cohort analyses across different sampling schemes. Finally, we illustrated the method by estimating the effects of time-invariant and time-varying exposures on mortality risk in a large cohort of 502,360 participants from the UK Biobank, focusing on smoking and air pollution.

Results: Specific schemes of the novel event-based sampling strategy appeared to be unbiased both theoretically and empirically, with minimal loss in precision compared to the full cohort analysis, and more efficient than traditional NCC schemes. In the UK Biobank, we observed similar estimates using the novel NCC schemes compared to the full-cohort analysis, with negligible increase in standard errors and a noticeable reduction in computation time, inversely proportional to the number of controls per event.

Conclusions: We developed a novel sampling method based on the nested case–control design that allows to efficiently analyse large cohort datasets without introducing bias in the presence of ties or significantly reducing precision. This approach is especially well suited for datasets comprising millions of individuals, where traditional methods for time-to-event analysis are often infeasible.

All-cause mortality assessment at the point of diabetic eye screening: a streamlined approach to individualised preventive care

Adhikari, Nayani G., St George's School of Health and Medical Sciences, City St George's, University of London, London, UK; Olvera-Barrios, Abraham, NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; Warwick, Alasdair, NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; Shakespeare, Royce, St George's School of Health and Medical Sciences, City St George's, University of London, London, UK; Chambers, Ryan, Diabetes and Endocrinology, Homerton Healthcare NHS Foundation Trust, London, UK; Bolter, Louis, Diabetes and Endocrinology, Homerton Healthcare NHS Foundation Trust, London, UK; Anderson, John, Diabetes and Endocrinology, Homerton Healthcare NHS Foundation Trust, London, UK; Fajtl, Jiri, School of Computer Science and Mathematics, Kingston University, London, UK; Welikala, Roshan, School of Computer Science and Mathematics, Kingston University, London, UK; Barman, Sarah, School of Computer Science and Mathematics, Kingston University, London, UK; Pitt, Matilda, St George's School of Health and Medical Sciences, City St George's, University of London, London, UK; Egan, Catherine, NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; Tufail, Adnan, NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; Morris, Joan K., St George's School of Health and Medical Sciences, City St George's, University of London, London, UK; Owen, Christopher G., St George's School of Health and Medical Sciences, City St George's, University of London, London, UK; Rudnicka, Alicja R., St George's School of Health and Medical Sciences, City St George's, University of London, London, UK

Background: Diabetic retinopathy (DR) is a major complication of diabetes and a leading cause of vision loss in working-age adults. Beyond ocular disease, DR reflects systemic vascular damage and is associated with premature mortality. Integrating mortality risk into eye screening could improve risk stratification and enable targeted prevention. Examining heterogeneity by age, DR severity, and ethnicity is important to assess equity and understand risks beyond the dominant effect of age.

Methods: Data were obtained from the North-East London Diabetic Eye Screening Programme (NEL DESP), including 198,921 individuals with type 2 diabetes screened between Jan 3, 2012, and Dec 31, 2024. Cox proportional hazards models adjusted for sociodemographic and clinical factors were used to estimate mortality risk, including interaction terms. The proportional hazards assumption was assessed using Schoenfeld residuals. DR progression from mild non-proliferative DR (R1) to pre-proliferative/proliferative DR (R2/R3) was classified as fast (<3 years) or slow (>5 years). Sensitivity analyses were conducted in those with available electronic health records (EHR).

Results: Over the study period, with a median (IQR) follow-up of 7.1 (3.0–11.4) years, 28,706 (14.4%) participants died. Mortality risk increased strongly with DR severity across all age groups, with higher overall risk at older ages but a steeper dose–response across DR grades in younger groups. Age × DR grade interactions significantly improved model fit (likelihood ratio test $x^2(20)=166.05$, $p<0.001$). Ethnic differences were modest, with slightly lower hazards in Black and South Asian participants versus White participants. Women had lower mortality than men (HR 0.77, 95% CI 0.75–0.79). Lower risk was also observed in less deprived groups and those with less severe contralateral DR. Among 2,057 individuals progressing from R1 to R2/R3, slow progressors had substantially lower mortality than fast progressors (HR 0.43, 95% CI 0.27–0.69). Sensitivity analyses incorporating additional EHR phenotypes did not materially alter the findings.

Conclusions: Baseline DR grade, age, and faster DR progression showed strong, graded associations with all-cause mortality in all ethnic groups, highlighting their value in risk stratification. Incorporating these factors within DR screening could inform targeted prevention with more intensive risk factor management alongside tighter diabetes control.

Improving Nested Exposure Case-Control Designs for Rare Time-Dependent Events via Counter-Matching

Wiedemann, Stefanie, Department of Mathematical Statistics and Artificial Intelligence in Medicine, University of Augsburg, Augsburg, Germany; Beyersmann, Jan, Institute of Statistics, Ulm University, Ulm, Germany; Friedrich-Welz, Sarah, Department of Mathematical Statistics and Artificial Intelligence in Medicine, University of Augsburg, Augsburg, Germany; Feifel, Jan, Merck Healthcare KGaA, Darmstadt, Germany

Background: In observational studies, the influence of intermediate events, such as hospital-acquired infections, on outcomes like hospital discharge is often a subject of interest. However, these intermediate events may occur rarely, which makes it difficult, and sometimes impossible, to collect sufficient information. This complicates the estimation of hazard ratios and related estimators. In situations where the occurrence of outcomes is rare, the nested case-control design has been proposed, where controls are selected using sampling strategies such as simple random sampling or counter-matching. In particular, incorporating additional covariate information into the sampling process through counter-matching leads to enhanced efficiency, achieved through the generation of more informative risk sets. The nested exposure-case-control (NECC) design extends this design to settings with a rare intermediate event by integrating the occurrence of the time-dependent event into the sampling process. To date, the selection of controls in NECC primarily relies on simple random sampling, which may result in risk sets that contain limited information on the intermediate event, thereby hindering the estimation of hazard ratios.

Method: We propose the use of counter-matching within the NECC framework to obtain more informative risk sets. We integrate this into a counting process framework, allowing for the development of estimation procedures for relative and absolute risks that are comparable to those used in full cohort analyses. We investigate possible adaptations of the partial likelihood to derive estimators for the regression parameters and related quantities.

Results: The performance of the proposed approach is assessed through simulation studies, comparing it to simple random sampling designs and full cohort analyses. In addition, the method is applied to data on hospital-acquired infections and length of stay to explore its practical implications. Preliminary investigations suggest that the proposed design provides estimates with favorable statistical properties.

Conclusion: Adapting established models to better reflect real-world data challenges remains an important area of research. Extending nested case-control designs through counter-matching within the NECC framework offers a useful direction for handling rare intermediate events and provides a resource-efficient alternative to the standard Cox regression.

Quantifying the role of stage at diagnosis in racial disparities in US cancer survival: a population-based SEER study for 8 common cancers

Vayani, Aisha, Biostatistics Research Group, Division of Public Health & Epidemiology, University of Leicester, Leicester, UK; Rutherford Mark Biostatistics Research Group, Division of Public Health & Epidemiology, University of Leicester, Leicester, UK; Booth Sarah Biostatistics Research Group, Division of Public Health & Epidemiology, University of Leicester, Leicester, UK

Background: Black-White disparities in cancer survival remain substantial in the US. Late stage-at-diagnosis is a plausible and potentially modifiable contributor, but its impact likely varies by cancer site. We quantified, across eight common cancers, how much of the racial gap in five-year all-cause survival (ACS) is mediated by stage-at-diagnosis and translated this into postponable deaths under a hypothetical intervention shifting the stage distribution of Black patients to that of White.

Methods: We analysed SEER 21 data for Black and White patients aged 15-89 years diagnosed with malignant solid tumours from 2004-2022 (follow-up to 31 December 2022), covering 41.9% of the US population. Cancers included breast (female), lung and bronchus, colon (excluding appendix), uterine corpus, pancreas, oral cavity and pharynx, rectum and ovary. Stage-at-diagnosis was classified as localised, regional, or distant. A complete-case analysis was used (stage completeness >90%). Mediation was implemented within the relative survival framework using flexible parametric excess-hazard models stratified by sex, applying a period window (2021-2022) to estimate ACS. We partitioned the total causal effect of race on survival into the indirect (stage-mediated) and direct effects, and report the proportion mediated by stage across age groups. We additionally report deaths postponed among Black patients diagnosed in 2022 (including those with missing stage).

Results: The cohort included 3,054,225 females and 1,389,767 males. Black patients were diagnosed at similar/younger ages and more often with distant-stage disease than White patients, except for rectal cancer in females <75 years. At 5 years, stage-at-diagnosis accounted for roughly half of the disparity for breast (49.87%), colon (48.08% in males; 50.69% in females) and corpus uteri (45.33%), and a larger share for lung (70.22% in males; 59.85% in females) and oral cavity/pharynx (59.68% in females). Mediation was limited for ovary (18.17%), pancreas (15.41% in males; 3.02% in females), oral cavity/pharynx and rectum in males (30.84% and 23.03% respectively). Postponable deaths were time-dependent and greatest where both the indirect effect and case burden were large.

Conclusion: Stage-at-diagnosis explains a substantial but heterogeneous share of Black-White survival disparities. Policies promoting earlier diagnosis among Black populations could reduce short-term cancer mortality for several major cancers.

Clustering

Outcome-guided clustering within a generalised Bayes framework

Chen, Sida, University of Cambridge; Richardson, Sylvia, University of Cambridge; Tom, Brian, University of Cambridge

Background: In molecular precision medicine, there is a growing interest in identifying disease subtypes from high-dimensional molecular-level covariate data that are associated with clinically relevant outcomes. A motivating example arises in proteomic studies of knee osteoarthritis, where the aim is to identify patient subgroups whose proteomic signatures may support better disease stratification and prediction, and eventually more targeted treatment or prognostic assessment. Bayesian profile regression is an established method for such outcome-guided clustering. However, it may suffer from the outcome being weakly relevant when informing the clustering structure and from poor out-of-sample outcome prediction in realistic settings; potentially limiting its practical utility.

Methods: We propose a generalised Bayes (GB) framework for outcome-guided clustering by reformulating the problem as a multi-task decision problem with separate loss components (tasks) for clustering and prediction. The framework offers great flexibility and admits standard Bayesian profile regression as a special case. To improve the interpretability of the weighting of different loss components, we propose a principled standardisation that places each task risk on a common scale. Selection of the learning rate and task weights can be guided by an estimate of the expected out-of-sample risk under a user-chosen evaluation loss, which can be conveniently obtained from Markov chain Monte Carlo output.

Results: Preliminary simulation studies indicate that, with a suitably chosen evaluation loss and tuning parameters, the proposed GB approach achieves satisfactory cluster recovery and out-of-sample outcome prediction in challenging examples where standard profile regression fails. Results from an application to proteomic knee osteoarthritis data are also anticipated.

Conclusions: Some pitfalls of standard Bayesian profile regression arise from inherent limitations of the likelihood-based specification. GB provides a decision-theoretic framework that offers the flexibility to encode explicit priorities across tasks and to deliver more goal-aligned performance, with potential for broader modelling and application settings.

A stability-driven clustering framework for clinical data: application of paediatric Hypertrophic Cardiomyopathy identifies patient groups with distinct outcomes

Taylor, Cameron Skye, Centre for Inherited Cardiovascular Diseases at Great Ormond Street Hospital and Institute for Cardiovascular Sciences at UCL; Field, Ella, Centre for Inherited Cardiovascular Diseases at Great Ormond Street Hospital and Institute for Cardiovascular Sciences at UCL; MacPherson, Nina, Centre for Inherited Cardiovascular Diseases at Great Ormond Street Hospital and Institute for Cardiovascular Sciences at UCL; Cotina-Borja, Mario, Institute of Child Health at UCL; Lai, Alvina, Institute of Health Informatics at UCL; Kaski, Juan Pablo, Centre for Inherited Cardiovascular Diseases at Great Ormond Street Hospital and Institute for Cardiovascular Sciences at UCL; Norrish, Gabrielle, Centre for Inherited Cardiovascular Diseases at Great Ormond Street Hospital and Institute for Cardiovascular Sciences at UCL

Introduction: Heterogeneity in clinical populations limits efficient patient stratification. Clustering can identify at-risk patients, but it is often unstable and not reproducible. Especially in rare paediatric diseases, such as childhood Hypertrophic Cardiomyopathy (HCM), where research is limited. A stability-driven clustering framework is used to derive robust patient subgroups with distinct disease trajectories.

Methods: Baseline clinical and genetic data were analysed from a paediatric HCM cohort at Great Ormond Street Hospital (GOSH). A stability-driven optimisation framework tested multiple clustering methods (k-means, k-medoids, spectral, hierarchical) across many preprocesses, including imputation (random forest and MICE imputation) and feature reduction (no feature reduction, FAMD, tSNE). The clustering method with the highest mean Adjusted Rand Index was selected, and consensus clustering created final clusters. The concordance of the clustering in the test dataset tested generalisability. Bootstrapping the consensus clustering on the full development cohort created the final clusters. Phenotypic clusters and their clinical outcomes were analysed. A decision tree trained on the development cohort predicted cluster labels in the validation cohort (International Paediatric Hypertrophic Cardiomyopathy Consortium (IPHCC)).

Results: K-means clustering of 3 clusters was most stable, with a mean of 0.386 compared to other methods <0.26. The training set clusters were phenotypically distinct and generalised well in the test set. Bootstrapping in the full development set was highly consistent, 93.2% of patients had stability <0.85. Cluster A was characterised by younger patients with a milder phenotype, cluster B were more severe and symptomatic, and cluster C had older patients with the mildest phenotype but higher symptom burden. The clusters were associated with distinct outcomes, for example, arrhythmias within 5 years in 20% and 4% of clusters B and A, respectively. A decision tree trained on the development set classified patients from the IPHCC cohort into clusters that remained characterised as those at GOSH, with distinct outcomes.

Conclusions: The stability-driven framework is reproducible in heterogeneous clinical datasets. It identified clinically relevant clusters in HCM, with distinct event-free survival in both the development and external validation sets. This highlights the value of stability-based approaches for patient stratification.

MultiCOCor: identifying multiple clustering structures in high-dimensional data

Hodgkinson, Jack, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK Kirk, Paul DW, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

Background: Clustering is an exploratory tool for molecular datasets, for example to discover disease subtypes. Such high-dimensional data often reflect multiple biomolecular processes and pathways, each associated with different groups of variables, and may therefore exhibit multiple clustering structures. Standard clustering methods that produce a single partition of the dataset are not suited to these settings, as they can conflate heterogeneous signals arising from distinct processes and obscure meaningful structure.

Methods: We propose a novel, two-stage algorithm known as MultiCOCor that recovers multiple clustering structures. MultiCOCor proceeds in two stages: first, features are partitioned into distinct groups ('views') based on their correlation structure; second, samples are clustered within each view to identify subtype structures. This enables the identification of meaningful feature sets, where each set captures distinct biomarkers that may provide molecular signals for distinct subtypes.

Results: We benchmark MultiCOCor using a simulation study that evaluates structure recovery performance across a wide range of scenarios. The method demonstrates robust performance and scalability across high-dimensional settings, establishing it as a competitive alternative to existing approaches. We further apply MultiCOCor to transcriptomic data from The Cancer Genome Atlas Network (PMID: 23000897), where sample clusters inferred within individual views show strong agreement with established clinical annotations, including PAM50 subtypes and hormone receptor status. The resulting views correspond to interpretable sets of co-regulated genes, highlighting distinct molecular signals associated with known disease subtypes.

Conclusion: Overall, MultiCOCor provides a scalable, flexible, and computationally efficient approach to multi-view clustering, enabling the discovery of multiple latent structures in heterogeneous high-dimensional data.

Bayesian

A Bayesian Grid-Search Approach for Flagging Potentially Atypical Variants in Cancer Susceptibility Genes

Michaelides, Damianos, Department of Biostatistics, The Cyprus Institute of Neurology and Genetics; Zanti, Maria, Department of Biostatistics, The Cyprus Institute of Neurology and Genetics; Dennis, Joe, Centre for Cancer Genetic Epidemiology University of Cambridge; Boddicker, Nicholas, Department of Laboratory Medicine and Pathology Mayo Clinic; Goldgar, David, Department of Dermatology University of Utah; Couch, Fergus, Department of Laboratory Medicine and Pathology Mayo Clinic; Easton, Douglas, Centre for Cancer Genetic Epidemiology University of Cambridge; Spurdle, Amanda, Public Health Program, QIMR Berghofer Medical Research Institute; Michailidou, Kyriaki, The Cyprus Institute of Neurology and Genetics

Background: Accurate interpretation of genetic variants is essential for effective clinical decision-making and for reducing uncertainty experienced by patients undergoing genetic testing. The case-control likelihood ratio (ccLR) method incorporates age-specific information from cases and controls, enabling a precise use of case-control data in VUS classification. In ccLR, all pathogenic variants within a gene confer the same level of risk. In practice, however, some variants may exhibit atypical behaviour, conferring lower or higher penetrance than the canonical gene-level risk. Such variants are difficult to detect under models that assume fixed penetrance, potentially leading to misclassification. In this study, we propose a method that flags variants with atypical risk. The method was applied to large scale sequencing data of BRCA1/2.

Methods: We developed the grid-search ccLR framework to allow dynamic variant-specific risk. A penetrance scaling parameter was introduced to represent different profiles of variant risk relative to the canonical gene-level penetrance. Likelihood ratios were computed across a grid of penetrance profiles and across three datasets (UK Biobank, BRIDGES, CARRIERS) comprising of 100K cases and 300K controls. Bayesian methods were then applied to integrate evidence across the penetrance continuum, producing a marginal Bayes factor (BF) for overall risk association and a deviation BF to evaluate whether observed data were inconsistent with canonical penetrance. These measures were used to flag variants with potential atypical behaviour.

Results: Across more than 3K rare BRCA1/2 variants evaluated, we identified variants with evidence consistent with reduced penetrance behaviour. Several candidate variants are known from literature, demonstrating that the approach can recover known signals from previous studies. Importantly, some variants evident as benign under the full-penetrance model showed signals consistent with elevated risk when penetrance was allowed to vary.

Conclusions: Our findings demonstrate that relaxing the fixed penetrance assumption can reveal variant-level heterogeneity in cancer risk that may be overlooked by traditional case-control analyses. By identifying variants with atypical risk profiles, the proposed grid-search Bayesian framework provides a tool for variant interpretation which will help reduce uncertainty by highlighting potentially atypical variants that warrant further investigation.

Time-Dependent Bayesian Borrowing from Historical Controls in Randomized Clinical Trials with Temporal Drift

Emmanuel Adewuyi; Jonathan Bartlett; Darren Scott; Alex Lewin

Introduction: Borrowing information from historical controls can improve efficiency in randomized clinical trials; however, standard borrowing approaches typically assume exchangeability across studies. This assumption may be inappropriate when standards of care improve over time, inducing temporal drift between historical and current control outcomes. We propose a Bayesian model with a time-dependent informative prior for borrowing from multiple historical studies to improve estimation of the treatment effect in the current trial.

Methods: We consider individual-level continuous outcomes from one current randomized trial and multiple historical control studies. We specify a prognostic score-adjusted outcome model in which prognostic covariate effects are shared across studies, while study-specific control intercepts evolve over calendar time. Temporal dependence among these intercepts is modeled through an autoregressive hierarchical prior, allowing borrowing to be stronger for more commensurate studies and weaker when temporal drift is present. Posterior draws are combined with g-computation to estimate the current-trial marginal mean under treatment, the marginal mean under control, and their difference. A simulation study compares the proposed method with no borrowing, full pooling, power priors, and exchangeable hierarchical priors under scenarios with no, mild, and strong drift.

Results: The simulation study is designed to assess bias, interval coverage, precision, and the degree of effective borrowing across drift scenarios. We expect the proposed approach to improve precision relative to no borrowing when drift is absent or mild, while remaining more robust than full pooling and exchangeable borrowing models when temporal drift is substantial. Final numerical results will quantify efficiency gains and protection against bias. In addition to the simulation study, a real randomized clinical trial dataset will be analyzed using both the proposed and existing approaches to compare their performance and assess consistency with the simulation findings.

Discussion: Time-dependent informative priors provide a principled framework for adaptive borrowing from historical controls when outcome processes evolve over time. By combining temporal modeling with covariate adjustment and marginal effect estimation, the proposed approach seeks to improve efficiency while preserving validity in contemporary clinical trials.

Using Bayesian K Regression Modelling and Bayesian Model Averaging to Model Frailty Indices and Reduce Risk of Misdiagnosis

Sun, Ziqian, University of Sheffield; Juarez, Miguel, University of Sheffield; Bellantuono, Ilaria, University of Sheffield

Background: Frailty is a heterogeneous and multifactorial clinical syndrome without an agreed definition. In practice, various frailty measures have been developed and used, but they are of different types in different scales, emphasising different aspects of frailty. This makes it expensive and time-consuming to test new interventions in clinical trials. Generalised Linear Models (GLMs) can be used to model various types of frailty measures. However, when modelling frailty indices, which are continuous bounded outcomes, the assumptions of GLM do not suit the characteristics of the data very well.

Methods: So, we propose to use Kumaraswamy Regression (KR) and Exponentiated Kumaraswamy Regression (EKR) models with a Bayesian approach to model frailty indices. The K and EK distributions can be used to model continuous scores with arbitrary boundaries. Furthermore, we apply Bayesian Model Averaging (BMA) to combine all useful models together by giving each a weight, reflecting how well the model is, to generate the averaged prediction of frailty.

Results: The KR and EKR models have been fitted with two median-dispersion reparameterisations (kappa and sigma, respectively), and they are compared with standard GLM models, including linear regression, Gamma-GLM, and beta regression in terms of information criteria (IC) and predictive errors. Results show that the model with kappa parameterisation exhibits some extent of identifiability issues, but with the usage of a hierarchical GaGa prior, it can provide better predictive results and faster model fit than the models with the sigma parameterisation and the other GLM models. Also, 110k+ models have been assessed with BMA in terms of posterior probability, and only averaging the top 10 can obviously provide more precise and accurate frailty predictions. A ranking of the predictors regarding their relative importance to frailty can be generated by accumulating the variable inclusion probabilities across models averaged in BMA. This information is useful in clinical trials to reduce the risk of misdiagnosis, given limited time and money.

Conclusions: KR and EKR models are suitable for modelling continuous bounded outcomes, e.g., frailty indices, in terms of both model assumptions and predictive ability. Using BMA can generate better frailty predictions because of its multifactorial nature, which can contribute to reducing misdiagnosis risk in clinical trials.

Bayesian Simultaneous Credible Bands for Polynomial Regression

Yang, Fei, The University of Manchester; Han, Yang, The University of Manchester; Liu, Wei, University of Southampton

Background: Assessing uncertainty in dose–response curves is crucial in early-phase drug development to support safe and reliable decision-making. An insufficient dose may lead to inadequate efficacy, while an excessive dose can result in safety issues. Although the frequentist simultaneous confidence band (FSCB) is widely used to quantify the plausible range of an estimated curve over relevant dose levels, practical methods for constructing Bayesian simultaneous credible bands (BSCBs) with exact coverage remain limited.

Methods: We propose a unified Bayesian framework for constructing two-sided hyperbolic BSCBs for polynomial regression curves over a finite covariate interval. An efficient simulation-based procedure is developed to determine the critical constant, reducing computation to a one-dimensional global optimization problem. The framework accommodates different levels of prior information through three approaches: the normal-gamma conjugate prior, the independent Jeffreys prior, and non-conjugate priors implemented via Hamiltonian Monte Carlo (HMC).

Results: We show that, under the independent Jeffreys prior, the proposed BSCB is identical to the FSCB in finite samples. More generally, under regularity conditions, the BSCB is asymptotically equivalent to the FSCB, thereby attaining the nominal frequentist coverage in large samples. Simulation studies demonstrate that the proposed method attains exact posterior simultaneous coverage across a wide range of settings while allowing prior information to improve precision relative to objective or frequentist approaches. In a Phase II dose–response application, the method provides a practical way to quantify efficacy uncertainty over the full dose range and to support minimum effective dose identification.

Conclusions: The proposed framework offers a flexible and reliable approach to Bayesian simultaneous inference for regression curves, combining exact posterior simultaneous coverage with strong frequentist justification. By accommodating different levels of prior information and remaining computationally practical, it provides a useful tool for dose–response uncertainty quantification and decision-making in early-phase clinical trials. An accompanying R package is provided for implementation.

Longitudinal data

Hierarchical Clustering for Sparse Longitudinal Data: Extensions for Continuous and Ordinal Outcomes

Han, Shawn, Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge; Barrett, Jessica, Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge; Winpenny, Eleanor, Mohn Centre for Children's Health and Wellbeing, School of Public Health, Imperial College London, London.

Introduction: Longitudinal behavioural outcomes in population health are often collected using questionnaires at a limited number of waves, resulting in sparse and unbalanced data. Standard longitudinal clustering approaches, such as latent class mixed models, can perform poorly in this setting because of unstable estimation, modelling assumptions and high computational burden. Hierarchical longitudinal clustering is a useful alternative, but early merging decisions are irreversible, making initial-stage stability critical. Existing formulations mainly focus on continuous outcomes, whereas many questionnaire-based measures are ordinal. We aimed to develop improved clustering methods for sparse ordinal and continuous longitudinal data.

Methods: We developed two methodological contributions targeting instability in sparse longitudinal clustering. For continuous outcomes, we introduce a roughness-penalised spline preprocessing strategy using a P-spline roughness penalty on spline coefficients. For ordinal outcomes, we propose a likelihood-based hierarchical clustering framework in which each cluster is represented by a proportional odds trajectory model and similarity is defined by likelihood loss under merging. To address non-identifiability in early stages, we introduce a surrogate likelihood based on global threshold estimates, allowing clusters to contribute level information without restrictive assumptions. A likelihood-based Calinski–Harabasz-type criterion is used to select the number of clusters, and model complexity is adapted dynamically.

Results: Simulation studies based on the structure of Understanding Society, a UK household survey, show that the proposed methods improve clustering stability and recovery of latent trajectory structures under extreme sparsity. The continuous extension improves robustness under both normal and gamma error distributions, while the ordinal framework yields stable and interpretable trajectory groupings. Both approaches are computationally efficient relative to latent class mixed models, achieving comparable or improved performance without restrictive distributional assumptions.

Conclusions: These findings highlight early-stage instability as a key challenge in clustering sparse longitudinal data and show that appropriate regularisation improves robustness and interpretability. The proposed framework provides a flexible and scalable approach for analysing questionnaire-based longitudinal data in life-course research.

Comparing joint recurrent event modelling frameworks for dynamic prediction in chronic conditions: an empirical evaluation

Thomas Spain, Laura J. Bonnett, Maria Sudell

Introduction: Joint models linking longitudinal predictors with recurrent event processes offer a principled framework for dynamic risk prediction in chronic conditions. Despite a growing methodological literature, the relative predictive performance of competing joint recurrent event frameworks under varying data conditions remains poorly understood. Practical barriers including convergence difficulties and computational burden further complicate framework selection, yet these are rarely evaluated systematically in applied settings.

Methods: We present an empirical comparison of joint modelling frameworks identified through a systematic review of existing applications in chronic condition research, with three distinct frameworks emerging from current practice: Andersen-Gill (AG), Prentice-Williams-Peterson (PWP), and shared frailty models. Each framework is fitted to multiple longitudinal clinical datasets from chronic condition settings, enabling evaluation across datasets with differing event rates and follow-up structures. A unified validation strategy is applied consistently across all frameworks, with predictive performance assessed using the C-statistic at a fixed time horizon. Convergence rates and computational time are additionally recorded to capture practical alongside statistical considerations.

Results: Results are not yet available at the time of submission. The analysis will characterise how predictive performance, convergence behaviour, and computational demands vary across frameworks and dataset characteristics, providing an evidence base for framework selection in applied recurrent event prediction.

Conclusions: This work addresses a practical gap for statisticians developing prediction models that incorporate longitudinal information alongside recurrent event outcomes. By evaluating frameworks across multiple real clinical datasets, this comparison moves beyond theoretical considerations to offer direct, data-driven guidance on when methodological complexity is justified and where simpler approaches remain adequate.

Modelling the effect of Longitudinal Markers on Left-Truncated Time-to-Event Outcomes in Twin Studies

Muli Annah, Department of Clinical Trials Research Unit, University of Leeds, United Kingdom Department of Statistics, University of Leeds, United Kingdom. Girondo Mar Rodríguez, Department of Biomedical Data sciences of the Leiden University Medical Centre, the Netherlands. Houwing-Duistermaat Jeanine, Department of Statistics, University of Leeds, United Kingdom Department of Applied Statistics, Eindhoven University of Technology, the Netherlands.

Background: The identification of biomarkers for disease onset in longitudinal studies necessitates precise estimation of the association between longitudinal markers and survival outcomes. Currently, methods for estimating these associations in the context of left-truncated and clustered survival outcomes are lacking.

Methods: In this study, we propose a novel model tailored to this scenario and develop several estimation methods: last observation carried forward, regression calibration, and a two-stage likelihood approach for joint modeling of longitudinal and survival processes.

Results: Simulation results indicate that the last observation carried forward method performs well only with a dense grid and no marker measurement error. For less dense grids and low measurement error, regression calibration approaches are preferred. Joint modeling approaches outperform ordinary calibration methods in the presence of measurement error, although they may suffer from numerical instability. In cases of numerical instability, calibration methods might be a good alternative. We applied these methodologies to the TwinsUK data to estimate the effect of bone mineral density (BMD) as a longitudinal marker on fracture incidence in 766 elderly females, approximately 138 of whom experienced a fracture during the study period. In the survival model, we included a shared gamma-distributed frailty to account for the additional correlation between the times to fracture of twin pairs. Estimates obtained using regression calibration and joint modeling approaches indicated a larger BMD effect compared to the last observation carried forward method, which can be attributed to the irregular BMD measurement process and minimal measurement error.

Conclusions: Overall, our methods offer valuable tools for modelling the effect of a longitudinal marker on survival outcomes in complex designs.

Investigating student dropout and performance using three-part mixed-effects quantile regression models for multivariate longitudinal data

Saiz, Maria, Sapienza University of Rome; Merlo, Luca, Link Campus University; Petrella, Lea, Sapienza University of Rome

Background: University success is a multidimensional construct, reflecting multiple interrelated factors, including student retention, exam participation, and academic performance. The joint analysis of these outcomes is essential for assessing university effectiveness and informing educational policies. Motivated by a longitudinal dataset of students enrolled in a bachelor program at the Sapienza University of Rome, this study investigates the impact of students' characteristics on both dropout risk and academic performance.

Methods: We propose a novel three-part mixed-effects multivariate quantile regression model. The model consists of: (i) a dropout component, (ii) a zero-inflated component capturing the probability of passing no exams in a given year, and (iii) a multivariate quantile regression for positive performance outcomes. Adopting a Maximum Likelihood approach, estimation is carried out through an Expectation–Maximization algorithm using the multivariate asymmetric Laplace distribution as working likelihood. This enables the joint modelling of the univariate conditional quantiles of the outcome distribution while also accounting the possible correlation between the responses. Unobserved heterogeneity and serial dependence are accommodated through discrete multivariate finite mixtures estimated from the data.

Results: The empirical application on Sapienza student data demonstrates the heterogeneity in the effect of individual characteristics across dropout risk and the univariate distributions of students' performance, measured with the number of annual gained credits and the weighted average of exam grade. Specifically, the significance and the magnitude of estimated coefficients vary across quantiles. Additionally, the model identifies latent sub-groups of students distinguishing clusters of students that exhibit a marked tendency toward strong academic performance, whereas others that display a higher likelihood of dropping out and lower performances.

Conclusions: With the aim of capturing the multidimensional nature of university success, this study proposes a three-part mixed-effects multivariate quantile regression. The main contribution lies in its joint modelling perspective, which moves beyond separate outcome-specific analyses and allows the identification of longitudinal patterns of academic achievement. The empirical findings demonstrates that the covariates' effect is not uniform across dimensions and different levels of performance.

Prediction Modelling

Who Benefits from Intensive Treatment in Psoriatic Arthritis? Deriving Personalised Treatment Rules using Causal Forests and Policy Trees

Lema Martinez, Jose Manuel, University of Oxford, NDORMS & Oxford Clinical Trials Research Unit, Oxford, UK; Jiang, Yunxuan, University of Oxford, NDORMS, Oxford, UK; Kang, Ji-Hyoun, University of Oxford, NDORMS, Oxford, UK; De Marco, Gabriele, University of Leeds, NIHR Leeds Biomedical Research Centre & Leeds Institute of Rheumatic and Musculoskeletal Medicine, Leeds, UK & Mid Yorkshire Teaching NHS Trust, Department of Rheumatology, Wakefield, UK; den Broeder, A.A., Sint Maartenskliniek, Department of Rheumatology, Nijmegen, Netherlands; van Herwaarden, Noortje, Sint Maartenskliniek, Department of Rheumatology, Nijmegen, Netherlands; Koc, Gonul Hazal, Erasmus MC, Department of Rheumatology, Rotterdam, Netherlands; Vis, Marijn, Erasmus MC, Department of Rheumatology, Rotterdam, Netherlands; Kok, Marc R., Maasstad Hospital, Department of Rheumatology and Clinical Immunology, Rotterdam, Netherlands; Luime, Jolanda J., Erasmus MC, Department of Rheumatology, Rotterdam, Netherlands; Marzo-Ortega, Helena, University of Leeds, NIHR Leeds Biomedical Research Centre & Leeds Institute of Rheumatic and Musculoskeletal Medicine, Leeds, UK; Mulder, Michelle L.M., Sint Maartenskliniek, Department of Rheumatology, Nijmegen, Netherlands; Tchetverikov, Ilja, Albert Schweitzerhospital, Department of Rheumatology, Dordrecht, Netherlands; Massa, Sofia, University of Oxford, NDORMS & Oxford Clinical Trials Research Unit, Oxford, UK; Coates, Laura C., University of Oxford, NDORMS, Oxford, UK

Background: Randomised trials of intensive treatment strategies in early, DMARD-naïve psoriatic arthritis (PsA) have produced conflicting results, suggesting clinically meaningful treatment effect heterogeneity obscured by reporting only the average treatment effect (ATE). Causal forests and policy trees allow estimation of individual-level conditional average treatment effects (CATEs) and translation into interpretable treatment rules. This is the first application of these methods to pooled individual participant data (IPD) from four PsA trials.

Methods: We conducted an IPD meta-analysis of four strategy trials (N=349 complete cases; N=379 with multiple imputation). Participants received intensive treatment or standard step-up care. The primary outcome was PASDAS at weeks 16–24 (lower scores indicate better disease control). Treatment heterogeneity was estimated using generalised random forests, with robustness assessed across >40 sensitivity specifications varying propensity score, honest splitting, covariate set, and pooling strategy. Personalised treatment rules were derived using policy trees trained on doubly robust pseudo-outcomes with leave-one-study-out cross-fitting. Depth-2 and depth-3 policies were evaluated against treat-all and treat-none benchmarks.

Results: The ATE consistently favoured intensive care across all specifications (range −0.27 to −0.62 PASDAS units). Approximately 27% of participants had CATEs crossing zero. Age and BMI were the main drivers of treatment effect heterogeneity. Depth-2 trees split first on age then BMI, identifying four subgroups; the primary depth-2 policy improved outcomes over treat-all intensively by 0.16 PASDAS units (range: 0.11–0.26). Depth-3 trees split first on baseline PASDAS with additional splits on sex and disease duration, yielding larger gains (primary estimate 0.27; range 0.24–0.42). Patients recommended intensive treatment had strongly negative mean effects (−1.37 to −0.73); those recommended standard care had positive estimates.

Conclusions: Substantial heterogeneity exists in response to intensive treatment in early PsA. Policy trees derived from causal forests provide interpretable decision rules outperforming

uniform strategies. Limitations include modest pooled sample size, heterogeneity in treatment approaches across trials, and absence of external validation.

Quantifying Individual-level Uncertainty in the BOADICEA Breast Cancer Risk Prediction Model

White, Bethan L., University of Cambridge; Ficorella, Lorenzo, University of Cambridge; Tischkowitz, Marc, University of Cambridge; Usher-Smith, Juliet A., University of Cambridge; Easton, Douglas F., University of Cambridge; Antoniou, Antonis C., University of Cambridge

Background: Breast cancer risk prediction models estimate an individual's probability of developing breast cancer and identify those who may benefit most from screening or preventive interventions. The BOADICEA breast cancer risk model, implemented in the CanRisk tool (www.canrisk.org), incorporates genetic, lifestyle, hormonal, family history, and anthropometric data to calculate individualised risk estimates. However, these predictions are subject to model uncertainty, which is rarely quantified at the individual level and can vary considerably depending on risk factor profiles. Quantifying individual-level uncertainty is crucial for evaluating risk prediction models, improving trust in risk estimates, and supporting effective risk communication.

Methods: We developed a framework to quantify individual-level uncertainty in breast cancer risk predictions from the BOADICEA breast cancer risk model. Monte Carlo simulation was used to sample from model parameter uncertainty distributions, and lifetime breast cancer risk was calculated for each sampled parameter set. This produced an uncertainty distribution of the individual's risk, from which summary measures such as uncertainty intervals and the spread across risk categories were calculated.

Results: Uncertainty varied considerably across individuals depending on risk factor values. For a woman with strong family history of breast cancer and median values for other risk factors, the BOADICEA point estimate for lifetime breast cancer risk was 22.3% (moderate risk) with a 95% uncertainty interval of 20.0% to 27.1%. For a woman with strong family history, a high-risk BRCA1 pathogenic variant, and median values for other risk factors, the lifetime risk point estimate was 85.0%, with a wider uncertainty interval of 74.2% to 95.8%.

Conclusions: There may be considerable uncertainty in breast cancer risk predictions for some individuals; this is particularly relevant when predicted risk lies near clinical decision thresholds. Because uncertainty varies between individuals, modelling it at the individual level is crucial. Communicating uncertainty alongside point estimates can improve transparency and support shared decision-making. This framework also helps identify parameters where improved precision could meaningfully reduce uncertainty in personalised breast cancer risk assessment.

A Statistical Framework for HPA Axis Adjacent Hormone Trajectory Analysis

Victor Applebaum, University of Exeter and University of Cambridge; Kyle Wedgwood, University of Exeter; Peter Challenor, University of Exeter; Thomas Upton, University of Bristol; Eder Zavala, University of Manchester; Krasimira Tsaneva-Atanasova, University of Exeter

Background: The Hypothalamic-pituitary-adrenal (HPA) axis governs the regulation of a number of stress response hormones including cortisol, cortisone, tetrahydrocortisol, allo-tetrahydrocortisol, 18-hydroxycortisol, corticosterone and aldosterone. These hormones and their daily trajectories are known to vary with a number of hormone dysregulation associated disorders, physiological and demographic variables, however, the mechanisms underlying these processes are little understood.

Methods: We present a semi-mechanistic framework for analysing the associations between physiological and demographic variables and changes in features of the trajectory using a dataset of 213 healthy individuals. Although we focus on six of these variables: gender, year of birth, smoking, systolic blood pressure, diastolic blood pressure, and BMI; the framework is designed to function with arbitrarily many variables, allowing future research to be conducted. The model itself is based on a multiplicative vector autoregressive (mVAR) / vector multiplicative error model (vMEM) structure. We construct the location argument of hierarchical Bayesian priors on interpretable parameters, such as production rates and decay rates, with linear sums of the physiological and demographic variables. This makes it possible to identify which processes are responsible for differences in hormone levels. Furthermore, the model can be used for forecasting, imputation and prediction of other hormone values when only one (e.g. cortisol) is available. We test our model on these tasks and find that the model can make well calibrated forecasts and imputations.

Results and Conclusions: The model allows us to view a large number of relationships between physiological variables and interpretable model parameters representing production rates, stochasticity, decay rates and more. The model performed well at forecasting and imputing, but generally overfitted when trying to predict other hormones using only cortisol.

Comparison of propensity score methods for a continuous exposure: application to causal inference in two Parkinson's disease cohorts

Poore, Mia, NHS Blood and Transplant/University of Bristol; Nodehi, Anahita, University of Bristol; Lawton, Michael, University of Bristol; Ben-Shlomo, Yoav, University of Bristol

Background: Propensity score methods offer an alternative to traditional covariate adjustment for minimising confounding bias in observational studies. However, their use in the continuous exposure setting remains underdeveloped. Uncertainty about which methods are most appropriate and how to choose between them represents an important gap in the causal inference literature.

Methods: We compared four propensity score methods - stratification, matching, inverse probability weighting, and use of the propensity score as a covariate in the outcome model – to traditional covariate adjustment. We later added G-estimation as an extra comparator. We estimated the causal effect of a continuous measure of postural hypotension on Parkinson's disease motor symptom severity in two early disease observational cohorts. We presented strengths and limitations of each propensity score method and used Love plots to assess the extent to which each method balanced observed confounders, in two independent samples.

Results: Inverse probability weighting balanced covariates well and, along with G-estimation, yielded effect estimates similar to traditional covariate adjustment. Stratification, matching, and use of the propensity score as a covariate in the outcome model performed poorly, although the latter improved with doubly robust implementation. Across methods, an increase in postural hypotension was associated with a small but potentially pertinent increase in motor symptom severity. Using traditional covariate adjustment, we estimated that a 10mmHg increase in systolic blood pressure postural drop was associated with a 0.540 (95% CI 0.091, 0.990) point increase in MDS-UPDRS-III motor severity score in one cohort, and an 0.307 (95% CI -0.103, 0.718) point increase in the other cohort. The causal direction of effect could not be determined due to cross-sectional analysis.

Conclusions: No method was universally superior; traditional covariate adjustment remains a viable option. Propensity score methods achieving good balance yielded effect estimates similar to traditional covariate adjustment. Careful balance assessment and context specific method choice are of upmost importance when considering the use of these methods. This work highlights both the promise and the limitations of PS methods for continuous exposures, reinforcing the importance of rigorous, context-appropriate approaches so observational studies continue to reliably complement RCT evidence.

Lightning talks

Please note some of these abstracts will also be presented as posters

Deterministic vs Agent-Based Models of Influenza Spread: The Role of Stochastic Extinction in Small Populations

Abdullahi, Aliyo, Aston University

Background: Mathematical modelling plays an important role in understanding infectious disease dynamics and informing public health decisions. Deterministic compartmental models are widely used due to their simplicity, but they rely on assumptions such as homogeneous mixing and cannot capture stochastic variability. This study investigates how these assumptions affect conclusions about disease spread, using influenza as a case study.

Methods: A deterministic susceptible-infectious-recovered (SIR) model was expanded to a susceptible–exposed–infectious–recovered–susceptible (SEIRS) model to reflect the biological properties of influenza. The SEIRS was developed using a system of ordinary differential equations to describe disease dynamics in a well-mixed population. The model was parameterised using literature-based estimates for influenza (H3N2). To assess stochastic effects and spatial structure, an agent-based SEIRS model was constructed on a two-dimensional square lattice, where individuals interact locally and move via a lazy random walk. Monte Carlo simulations and parameter sweeps over transmission and recovery rates were conducted.

Results: The deterministic model predicts a peak infected proportion of approximately 0.07 and convergence to an endemic equilibrium when $R_0 > 1$. In contrast, the agent-based model produces a mean peak infected proportion of 0.399, occurring significantly earlier, and consistently leads to disease extinction in a small population (N = 81), despite $R_0 > 1$. Sensitivity analyses highlight a threshold structure governed by $R_0 = 1$, while additional results show that population size and vaccination coverage strongly influence outbreak severity and persistence.

Conclusions: Deterministic models describe the average long-term disease dynamics. For influenza, R0>1 indicates disease persistence. However, deterministic models may give misleading predictions in small or spatially structured populations because of their simplifying assumptions. The agent-based model, on the other hand, shows that random effects can lead the infectious proportion down to zero, which leads to disease extinction even when R0>1. As the population size grows, the stochastic model converges to the deterministic solution, which backs up the mean-field approximation. But these results show that R0>1 doesn't mean persistence in small populations. In general, this shows the importance of choosing the right modelling framework when looking at how diseases spread.

Improving 10-year cardiovascular risk prediction in patients with established cardiovascular disease: flexible addition of risk predictors on top of the SMART2 risk score

Richards, Matthew, University of Cambridge; Holtrop, Joris, UMC Utrecht; Wood, Angela, University of Cambridge; Dorresteijn, Jannick, UMC Utrecht

Background: Clinical guidelines recommend the SMART2 model to estimate recurrent cardiovascular event risk in patients with established atherosclerotic cardiovascular disease (ASCVD) to support preventive treatment decisions. Guidelines further acknowledge that biomarkers, comorbidities, anthropometric, and social factors may refine these estimates, yet a systematic approach to integrating such add-on predictors within the SMART2 framework had not been rigorously evaluated.

Methods: Patients aged 40-80 with established ASCVD were included from 11 cohorts (n=179,382; 25,789 recurrent CV events). Candidate add-on predictors included biomarkers (troponin I, NT-proBNP, albuminuria), comorbidities (heart failure, atrial fibrillation, coronary multivessel disease), anthropometric measures (BMI, waist/hip circumference), social determinants (employment, education), and clinical history factors. Additional predictors were incorporated using Fine-Gray competing risks models, with SMART2 linear predictor scores entered as an offset to preserve calibration. Performance was assessed via cohort cross-validation using change in C-statistic, calibration, and net benefit through decision curve analysis.

Results: Subdistribution hazard ratios ranged from 0.77 (95% CI: 0.75-0.80) for current employment to 1.69 (1.63-1.76) for heart failure history. The largest discrimination gains were observed for NT-proBNP (ΔC: +0.0127; 0.0060-0.0193) and troponin I (ΔC: +0.0100; 0.0020-0.0181), with statistically significant but smaller gains for employment, heart failure, and atrial fibrillation. Calibration remained adequate throughout. Decision curve analysis supported net benefit beyond SMART2 for NT-proBNP, heart failure, atrial fibrillation, albuminuria, employment, coronary multivessel disease, and education across risk thresholds up to 40%.

Conclusions: A flexible add-on framework incorporating guideline-outlined clinical and social factors significantly enhances personalised CV risk estimation in patients with established ASCVD. NT-proBNP and key comorbidities offer the greatest incremental value, with decision curve analysis highlighting clinical utility across actionable risk thresholds. This scalable approach adapts established risk models to richer patient-level data without discarding validated predictor structures.

How to Evaluate Imputation Methods for Missing Data in Medical Research: A Comparative Simulation Study

Thurow, Maria, TU Dortmund University and UA Ruhr Research Center Trustworthy Data Science and Security; Veips, Inken, TU Dortmund University; Dumpert, Florian, Federal Statistical Office of Germany; Pauly, Markus, TU Dortmund University and UA Ruhr Research Center Trustworthy Data Science and Security

Background: Handling of missing data is a crucial aspect when preparing data sets for further statistical analyses in several research areas. Previous studies have shown that the choice of imputation method can strongly influence subsequent analyses, especially in medical research, where missing values often occur due to study design or data collection challenges.

Methods: We conduct a comparative simulation study of commonly used imputation methods. The study includes methods such as missRanger, mixgb (both random forest-based imputation methods), MICE (Multiple Imputation by Chained Equations) and the naive imputation based on the arithmetic mean and mode as a benchmark. We use both simulated and real-world data sets from medical research. Based on these data sets, we demonstrate how to assess imputation methods based on their imputation accuracy. Since there is no unique definition of the accuracy of an imputation method, we focus on different goals that researchers might have when imputing missing values. We assess the predictive accuracy (reconstruction of the actual values) using the Normalized Root Mean Squared Error (NRMSE) and the Proportion of Falsely Classified (here referring to falsely imputed values) (PFC). To assess how well the original distribution is preserved, we use distribution distance measures such as the uni- and multivariate Kolmogorov-Smirnov statistic.

Results: Preliminary results indicate that tree-based methods perform best. Our results suggest that no single method outperforms the others across all settings and evaluation goals.

Conclusion: The optimal choice of an imputation method should depend on the specific analysis goal and the evaluation criterion. This study can be seen as a guide for researchers selecting imputation strategies aligned with different research goals, with particular relevance for medical research and beyond.

Methods for clinical risk prediction under interventions: a comparative simulation study

Kostouraki, Andriana, Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK; Williamson, Elizabeth, Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, UK.

Background: When developing a risk prediction model on a data set of individuals intervention-free at baseline, later intervention 'drop-in' - i.e. high-risk patients preferentially receiving the intervention - may lead to underestimation of risk under no intervention for such patients. Causality concepts have been combined with risk prediction modelling (often called predictions under interventions) to correctly allow for this setting; however, these require stronger assumptions and more complex modelling strategies, while it is unclear what their comparative performance is.

Methods: Approaches to account for intervention drop-in, could be grouped into those based on modelling i. the intervention, or, ii. the outcome - both conditional on intervention, covariates and outcome histories (propensity score and g-formula related, respectively), iii. both the intervention and outcome, iv. the outcome conditional on predictors - while incorporating a time-dependent intervention effect, pooled from an external source, or, v. those not accounting for intervention drop-in, referred to as 'standard'. We perform simulations to evaluate and compare the methods in longitudinal observational settings with time-varying confounding and intervention strategies sustained over time – i.e. 'intervention always' vs 'no intervention always'. We focus on time-to-event outcomes. We explore scenarios where the development and validation data may/may not differ and apply the Brier score, the area under the curve and the calibration slope, adapted to assess predictive performance under 'intervention always' and under 'no intervention always', as recently proposed in the literature. To estimate the true predictive performance, we generate 'perfect' counterfactual validation data sets.

Results: Accounting for intervention drop-in enables better prediction under intervention performance; the 'standard' approach performs worse when the target is prediction of risk under 'no intervention always'.

Conclusions: When predicting risk under intervention given predictors, causal inference offers well-established methods to correctly target the population of interest. Evaluation measures must appropriately assess the performance in predicting risk under interventions; therefore, standard measures must be adapted accordingly to contrast the predicted outcomes against the outcomes that would have been observed had individuals been assigned the intervention of interest.

Randomised Comparison of Ponatinib Versus Imatinib in Combination with Chemotherapy in Patients 55 Years of Age and Older with Newly Diagnosed Ph+ ALL: Primary Outcome Analysis of the EWALL PH03 Study

Shi, Tongtong, Centre for Trials Research at Cardiff University, Cardiff, United Kingdom; Porkka, Kimmo, University of Helsinki, Helsinki University Hospital Comprehensive Cancer Center and Hematology Research Unit Helsinki, Helsinki, Finland; Lübking, Anna, Skåne University Hospital, Lund, Sweden; Clappier, Emmanuell, Hopital Saint-Louis, Paris, France; Hicheri, Yosr, Institut Paoli-Calmettes, Marseille, France; Leguay, Thibault, Hôpital Haut Lévêque, Bordeaux, France; Lemasle, Emilie, Centre Henri Becquerel Rouen, Rouen, France; Burns, Sarah, Centre for Trials Research at Cardiff University, Cardiff, United Kingdom; Letterme, Charlotte, Centre Hospitalier de Versailles, Le Chesnay-Rocquencourt, France; Porter, Catharine, Centre for Trials Research at Cardiff University, Cardiff, United Kingdom; Sydenham, Mia, Centre for Trials Research at Cardiff University, Cardiff, United Kingdom; Thomas, Ian, Centre for Trials Research at Cardiff University, Cardiff, United Kingdom; Cayuela, Jean Michel, Hopital Saint-Louis, Paris, France; Rousselot, Philippe, Centre Hospitalier de Versailles, Le Chesnay-Rocquencourt, France; Sophie Rose, Centre for Trials Research at Cardiff University, Cardiff, United Kingdom; Ottmann, Oliver G., Cardiff University, Cardiff, United Kingdom

Background: BCR::ABL1 tyrosine kinase inhibitors (TKIs) plus reduced-intensity chemotherapy are standard of care for older patients with newly diagnosed Philadelphia chromosome-positive acute lymphoblastic leukemia (Ph+ALL). The third-generation TKI ponatinib (PON) has demonstrated superior molecular response rates over imatinib (IM) in adults with Ph+ALL. The EWALL PH03 trial was designed to evaluate this comparison in an older population, with an exploratory chemotherapy-free arm subsequently discontinued, necessitating a mid-trial protocol amendment from a three-arm to two-arm design.

Methods: Patients ≥55 years were randomised 1:1 to reduced-intensity chemotherapy with either PON or IM, using unweighted minimisation (20% random element) stratified by white blood cell count, bone marrow blast percentage, and age group. The primary endpoint was measurable residual disease (MRD) response (BCR::ABL1 ≤0.01%; MR4) after consolidation cycle 2, centrally assessed. Treatment groups were compared using a one-sided Mantel-Haenszel chi-squared test adjusted for stratification factors. Event-free survival (EFS) and overall survival (OS) were analysed using the Kaplan-Meier methods.

Results: Seventy-two patients (median age 66.8 years; 40.3% female) were randomised (n=36 per arm), with 66 evaluable for the primary endpoint after excluding patients who died or withdrew prior to primary endpoint assessment. MR4 response was achieved in 64.7% (22/34) for PON and 43.8% (14/32) for IM (adjusted p=0.11). Complete molecular remission rates were 35.3% and 28.1%, respectively. Withdrawal from study treatment was lower with PON (11.1%) than IM (25%). Survival data remain immature, limiting reliable estimation of time-to-event endpoints.

Conclusions: PON combined with reduced-intensity chemotherapy yielded a higher, though not statistically significant, deep molecular response rate compared to IM in older patients with Ph+ALL. Interpretation is limited by small sample size and incomplete outcome data. Longer follow-up is required to determine the impact on survival outcomes.

Joint modelling of longitudinal biomarkers and time to healing in venous leg ulcers

Zhan, Xinyue, Centre for Biostatistics, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK; Harvey, Joe, NIHR Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester UK, Division of Cell Matrix Biology and Regenerative Medicine, School of Biological Sciences, University of Manchester, Manchester, UK; Dumville, Jo, Manchester Centre for Health Psychology, University of Manchester, Manchester, UK; Sergeant, Jamie, Centre for Biostatistics, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK

Introduction: Venous leg ulcers (VLUs) are the most common chronic leg ulcer type, characterised by prolonged healing and substantial physical and financial burden. Prognostic biomarkers derived from longitudinal wound dressing samples may improve prediction of healing and enable personalised management. This study investigated whether longitudinal biomarker trajectories are associated with time to complete healing in patients with VLUs.

Methods: Biomarkers (coded to protect intellectual property) were measured longitudinally. The primary outcome was time to complete healing. Analyses included flexible parametric survival models, Cox proportional hazards models, and accelerated failure time (AFT) models to assess baseline associations. Joint models combining linear mixed-effects sub-models with Cox and AFT survival sub-models were then used to evaluate associations between longitudinal biomarker trajectories and healing time. Missing data were addressed using joint model multiple imputation, with Blom normal score transformation applied to all biomarkers.

Results: A total of 100 participants were included, with a median follow-up of 21 weeks (IQR 14–42 weeks), and 75% achieving healing within 60 weeks. After adjustment for baseline prognostic factors (age, wound area, ulcer duration, infection status, and compression therapy), biomarkers B2 and B5 remained significantly associated with time to healing (adjusted HR [95% CI], log-transformed: B2 0.528 [0.387–0.720]; B5 0.691 [0.495–0.964]).

Conclusions: The sample comprised predominantly older adults (≥60 years) with large, chronic VLUs receiving compression therapy and without infection, representing a higher-risk subgroup for delayed healing. This limits generalisability to the broader VLU population. The small sample size resulted in wide confidence intervals, and findings should be interpreted cautiously. In addition, joint models assume similar biomarker trajectory shapes across groups, differing only in magnitude, which may not reflect real-world variability. Biomarker predictive performance may also be influenced by storage duration of wound dressing samples.

Joint Bayesian modelling of antibody and PCR test data

Wang, Larry, MRC Biostatistics Unit; Jackson, Christopher, MRC Biostatistics Unit

Background: In infectious disease research, an increasing variety of data streams has become available. Modeling them jointly in a coherent and efficient manner is vital for understanding the underlying disease dynamics. In this work, motivated by the SARS-CoV-2 Immunity and Reinfection Evaluation (SIREN) study of healthcare workers in the UK, we propose a Bayesian framework that jointly models quantitative antibody and binary PCR test data. The primary aim of the model is to improve upon previous methods for estimating prevalence and incidence based solely on PCR test data. The model may also describe the evolution of immunity within individuals after exposure events.

Methods: Antibody trajectories are modeled using a smooth function that represents a boost following an exposure and a subsequent wane over time until plateau. The probability of a positive PCR test is modeled as a smooth logistic function of time since infection, and heterogeneity in immune responses is represented by covariates on antibody-related parameters. The posterior distribution was estimated using Numpyro's No U-Turn Sampler (NUTS) and Mean-Field variational inference (VI), a faster approximation. The accuracy of these algorithms is evaluated using Simulation Based Calibration (SBC) on synthetic cohorts observed over periods of 30 months.

Results: Across all simulations, NUTS yielded low bias and accurate coverage of posterior uncertainty, whereas VI suffered from systematic underestimation of posterior uncertainty. Nevertheless, NUTS proved to be computationally prohibitive for large cohort sizes, with inference taking over 2 hours for cohorts of size 1000, whilst VI took ≈ 2 minutes. In addition, despite under-coverage of posterior uncertainty, VI produced good point estimates, proving itself to be a useful alternative for large datasets. –

Conclusions: Our study demonstrates a new approach of jointly modeling antibody trajectory data and PCR test data. In ongoing work, the model is being applied to characterise incidence of SARS-CoV-2 infection among healthcare workers during 2020 − 2021, using data from the SIREN study, and determining the added value of antibody data. Further development directions include adopting a biphasic structure for the antibody trajectory to account for short term vs. long term waning, and evaluating optimal surveillance strategies for future pandemics.

The cost of electronic bedside transfusion checks and its impact on the whole hospital transfusion pathway in England

Montasir Ahmed[1*], Laura Green[2,3,4*], Florence Oyekan[2,4], Louise Bowles[2], Josephine McCullagh[2,3,4], Michael F Murphy[3,5,6], Nathan Proudlove[7] , Florian Tomini[1] , Yan Feng1

1 Wolfson Institute of Population Health, Queen Mary University of London, London, UK; 2 Barts Health NHS Trust, London, UK; 3 NHS Blood and Transplant, UK; 4 Blizard Institute, Queen Mary University of London, London, UK; 5 Oxford University Hospitals NHS Foundation Trust; 6 University of Oxford; 7 Alliance Manchester Business School, The University of Manchester, Manchester, UK

Background: Bedside electronic transfusion checks (BETC) for group and screen (G&S) sample labelling and blood administration are considered good clinical practice, improving patient safety and workflow efficiency, and potentially reducing costs. Despite widespread adoption, evidence on the economic impact of BETC across the transfusion pathway remains limited. This study evaluated the costs of BETC compared with manual transfusion checks and assessed its impact on workflow at Barts Health NHS Trust (BH), England.

Methods: We conducted a before-and-after micro-costing study using structured process mapping and time-and-motion observation across ten transfusion steps defined by Serious Hazards of Transfusion (SHOT). Data were collected across four hospitals within the BH before BETC implementation (May 2022–April 2023) and after implementation (May 2023–October 2025). BETC replaced manual processes for G&S sample collection and labelling (Steps 1–3) and blood administration and traceability (Steps 9–10), while Steps 4–8 remained unchanged. Direct clinician time and consumable use were measured per transfusion episode from the NHS provider perspective and valued using 2022/23-unit costs.

Results: BETC substantially reduces staff time by integrating electronic verification into previously manual stages, particularly group and screen (G&S) sample labelling and blood administration. Implementation of the BETC saved an average of 3.47 minutes (SD 0.36) per G&S sample. For blood administration, BETC reduces time by 4.93 minutes (SD 0.77) per routine blood unit transfusion and 2.24 minutes (SD 0.06) saved per emergency unit, corresponding to a weighted average saving of 4.79 minutes per unit transfused. Improvements in traceability further reduced staff time by 5.15 minutes for incomplete tags and 1.29 minutes for completed tags. Accounting for additional efficiencies— (including reduced duplication, rework, additional testing, and fewer mislabelled samples)—BETC reduces total transfusion pathway time by 17.59 minutes per transfusion episode, equivalent to approximately £10.62 in staff time savings.

Conclusions: This study provides one of the most detailed process-based economic evaluations of BETC to date and supports wider adoption across NHS hospitals to strengthen patient blood management and deliver system-wide cost savings.

Automated marking of Very Short Answer Questions in an applied health setting

Deeks, Oscar, University of Birmingham

Background: Very short answer questions (VSAQs) offer a flexible alternative to multiple-choice formats for assessing applied knowledge, but present challenges for scalable, consistent, and statistically valid marking. Free-text responses introduce variability that can undermine reliability and complicate quantitative analysis unless carefully managed. Automated marking approaches offer a potential solution but require transparent design and validation to ensure alignment with expert judgement and suitability as outcome measures. This work describes the development and validation of an automated mark scheme for VSAQs.

Methods: VSAQs were designed to assess specific, well-defined knowledge domains, with each question mapped to expert-defined scoring criteria. An automated marking framework was implemented using fuzzy matching techniques, allowing controlled flexibility in phrasing while maintaining interpretability. Question-specific features were incorporated, including synonym handling, tolerance for spelling variation, and rules for conditional credit. Matching strategies and thresholds were iteratively refined using pilot response data. Validation was conducted through comparison with human marking, assessing agreement at the individual item level and examining patterns of disagreement to identify sources of error. Final scores were structured for use in standard statistical analyses.

Results: Automated marking demonstrated good agreement with human scoring across multiple question types, while substantially reducing manual marking burden. When evaluated against human marking, the algorithm achieved a sensitivity of 96.6% and a specificity of 97.7%.Variation in performance across knowledge domains was observed, illustrating the ability of the scoring approach to discriminate between topic areas. The resulting scores were stable and interpretable, supporting their use as quantitative outcomes in analyses examining change in knowledge over time or exposure.

Conclusion: Question-specific fuzzy matching methods can be used to construct and validate automated mark schemes for VSAQs with explicit attention to measurement validity. This approach enables scalable scoring while supporting reliable statistical inference, including analyses of learning effects, and is applicable where free-text assessment is desirable but traditional marking is impractical.

Posters

Model selection tests for truncated vine copulas under nested hypotheses

Nishi, Ichiro, The Graduate University for Advanced Studies; Kawasaki, Yoshinori, The Institute of Statistical Mathematics;

Background: Copulas are widely used to model multivariate dependence because they separate marginal distributions from the dependence structure on a standardized uniform scale. Vine copulas provide a particularly flexible framework for modeling complex multi-dimensional dependencies by decomposing a multivariate copula into a sequence of bivariate building blocks. This flexibility, however, is accompanied by rapidly increasing model complexity as the dimension grows. Truncated vine copulas address this issue by limiting the number of vine trees included in the model, thereby improving parsimony. A key methodological challenge is to determine the appropriate truncation level. Although Vuong's model selection test has been proposed for this purpose, its use in the vine copula literature has been largely heuristic, because competing truncated models are inherently nested, whereas existing applications often assumed strictly non-nested hypotheses.

Methods: This study systematically investigates Vuong's model selection framework for truncated vine copulas under both nested and strictly non-nested settings. In particular, we compare the Vuong-N test, designed for nested hypotheses, and the Vuong-SNN test, designed for strictly non-nested hypotheses but previously applied heuristically in truncation problems. Extensive simulation studies were conducted for vine copula models under a range of dependence scenarios. The empirical performance of both tests was evaluated using p-values, rejection frequencies, and mean empirical Kullback-Leibler information criterion (KLIC).

Results: The simulation results show that the relative performance of the two tests depends strongly on the strength of dependence within the vine structure. When pairwise dependencies are weak, the nested hypothesis is more likely to hold, and the Vuong-N test tends to produce lower p-values, higher rejection rates, and improved mean empirical KLIC. In contrast, when dependencies are stronger, the Vuong-SNN test provides often superior model distinctions.

Conclusions: These findings clarify the roles of Vuong-N and Vuong-SNN in vine copula truncation. They show that the nested testing framework is preferable when the nested hypothesis is appropriate, while the strictly non-nested approach can still yield informative and practically useful guidance, even when applied heuristically. This contributes to both methodological understanding and practical model selection for truncated vine copulas.

Augmenting Randomized Trials With Real-World Controls for Negative Binomial Endpoints Using Bayesian Mixture Priors

Jiapeng Xu, MRC Biostatistics Unit; Pavel Mozgunov, MRC Biostatistics Unit; Jaki Thomas, MRC Biostatistics Unit

Negative binomial endpoints commonly arise in clinical studies involving counts with overdispersion, such as event rates, exacerbations, or healthcare utilization outcomes. When external real-world or historical control data are available, borrowing such information can improve efficiency, but inappropriate borrowing may inflate bias and error when the external and current data sources are not sufficiently comparable. We propose a robust Bayesian borrowing framework for negative binomial data based on exchangeable/nonexchangeable (EX/NEX) mixture priors. The method allows dynamic borrowing from historical or real-world controls into the concurrent control arm of a randomized trial while protecting against prior-data conflict. The proposed framework is built on the log-mean and log-dispersion parameterization of the negative binomial model and supports several borrowing strategies, including borrowing of the mean only, borrowing of both mean and dispersion separately, and joint borrowing structures. Exchangeability indicators determine the extent to which historical information is incorporated, allowing the data to drive posterior borrowing weights. This structure accommodates heterogeneous external sources and provides robustness when dispersion differs across studies. We evaluate the operating characteristics of the proposed approach under scenarios with varying degrees of agreement and drift between historical and current controls. Results show that the method can improve estimation precision and decision efficiency when exchangeability is plausible, while limiting excessive borrowing when incompatibility is present. The framework is particularly useful for augmenting randomized trials with real-world data in settings where count outcomes exhibit overdispersion and between-study heterogeneity. This work extends robust Bayesian borrowing methods to an important class of non-Gaussian endpoints frequently encountered in practice.

Exploration and Exploitation under Uncertainty: A Comparative Simulation Study of the Thompson Sampling Family in Multi-Arm Trials

Lema Martinez, Jose Manuel, University of Oxford, Newcastle University; Wilson, Kevin, Newcastle University; Wilson, Nina, Newcastle University

Background: Response-adaptive randomization (RAR) methods reallocate trial participants dynamically, directing more towards better-performing arms as evidence accumulates. This creates a fundamental tension: exploitative methods maximise benefit for participants within the trial, while explorative methods preserve the ability to correctly identify the best treatment for future patients. Thompson Sampling (TS) balances this trade-off through posterior sampling, while Accelerated Thompson Sampling (ATS) introduces a time-varying acceleration parameter to progressively shift allocation towards exploitation. The practical implications of method selection, particularly under prior misspecification and across varying clinical scenarios, remain incompletely characterised.

Methods: Four RAR strategies were evaluated: standard TS, a myopic strategy (ATS, $\rho = 0$), and two ATS variants ($\rho = 0.5$ and $\rho = 3$). Four clinical scenarios were examined, varying the size and clarity of treatment differences across five arms. Three prior specifications were assessed: flat, correctly centred informative, and incorrectly centred informative. Performance was evaluated across a range of prior variances using three operating characteristics: Probability of Correct Selection (P(Opt)), proportion allocated to the best arm (p^*), and the Maximum Misallocation Rate (MMR), defined as the maximum proportion allocated to any single suboptimal arm.

Results: No single method was universally superior. Explorative methods (standard TS and ATS, $\rho = 3$) consistently achieved higher P(Opt); exploitative methods (myopic and ATS, $\rho = 0.5$) allocated more participants to the best arm. Correctly specified informative priors produced near-perfect performance across all methods. Incorrectly centred priors were the most discriminating setting: standard TS maintained good P(Opt) while the myopic strategy and ATS ($\rho = 0.5$) performed poorly. Standard TS was the most robust strategy across all scenarios and prior specifications.

Conclusions: The choice of RAR method is a strategic decision that must align with the trial's primary objective. Standard TS is recommended as the default strategy, particularly when prior information is uncertain, due to its resilience to misspecification. The acceleration parameter ρ requires deliberate calibration and is not uniformly beneficial. MMR is introduced as a novel operating characteristic capturing the risk of persistent over-concentration on a single suboptimal arm.

From Hybrid to Fully Bayesian: A Simulation Study Evaluating Operating Characteristics and Efficiency in an Adaptive Platform Trial

Lema Martinez, Jose Manuel, University of Oxford, Newcastle University; Wilson, Kevin, Newcastle University; Wilson, Nina, Newcastle University

Background: The adoption of fully Bayesian adaptive trial designs has been limited by regulatory concerns over frequentist operating characteristics, particularly type I error control. This has driven the use of hybrid designs, which apply Bayesian methods at interim analyses and a frequentist final analysis. This paper contributes to the evidence that fully Bayesian adaptive designs are a viable alternative, and investigates efficiency gains through prior specification and participant allocation.

Methods: Using a Phase II, multi-arm, multi-stage platform trial as a case study, we converted the original hybrid design into a fully Bayesian framework by replacing the final frequentist test with a posterior probability decision rule, calibrated to match the original significance levels. Interim efficacy and futility rules were retained. A simulation framework (100,000 synthetic trials per configuration) compared both designs on four operating characteristics: assurance, Bayesian type I error, and their frequentist counterparts. Expected sample size (ESS) was estimated as a measure of efficiency. Informative truncated design priors governed data generation; non-informative analysis priors governed within-trial decisions. Five participant allocation configurations were evaluated per arm.

Results: At zero prior variance, the fully Bayesian and hybrid designs yield identical frequentist performance, showing that the Bayesian framework can exactly reproduce its hybrid counterpart when well calibrated. As prior variance increases, the frameworks diverge: Bayesian type I error declines sharply toward zero; the hybrid design's frequentist type I error remains substantially higher. Assurance shows a modest initial dip before stabilising above its starting level. Balanced allocation configurations consistently produced the lowest ESS and highest assurance across both arm types and all variance levels.

Conclusions: A fully Bayesian adaptive design can meet the frequentist operating characteristics that motivate hybrid designs, directly addressing the central regulatory concern. Beyond this equivalence, it offers substantially better false positive control. Assurance provides a more realistic measure of trial success than frequentist power by formally incorporating prior uncertainty. Balanced participant allocation is a straightforward design optimisation that reduces ESS and increases assurance, and should be considered as a default in multi-stage adaptive trials.

Recurrent Hamiltonian Directional Echo Learning (RHDEL): A Time-Reversible Framework for Memory-Efficient EEG Classification

Bhattacharya, Swastik, University of Manchester

Background: Deep learning methods for electroencephalography (EEG) classification typically rely on recurrent neural networks trained using Backpropagation Through Time (BPTT), which requires storing all intermediate hidden states and results in memory complexity scaling linearly with sequence length. This limits their applicability in real-time and resource-constrained settings such as wearable neurotechnology. Recent advances in physics-informed machine learning indicate that Hamiltonian systems, due to their time-reversible structure, offer a promising alternative for memory-efficient learning. This work builds on such ideas to develop a biologically grounded and computationally efficient framework for multi-channel EEG classification.

Methods: We propose Recurrent Hamiltonian Directional Echo Learning (RHDEL), (Refer to https://arxiv.org/abs/2506.05259 which shows the development of RHEL) a framework that models EEG signals as a Hamiltonian dynamical system in phase space. Each channel is represented by instantaneous phase and angular velocity derived via the Hilbert transform. The system is governed by a Hamiltonian consisting of kinetic energy and a directional coupling potential inspired by the Von Mises distribution to capture phase synchronization. Training follows a four-stage procedure: forward symplectic integration, terminal loss computation, a small momentum perturbation, and reverse-time integration that generates local “echo” updates for the coupling parameters. This eliminates the need to store intermediate states, achieving O(1) memory complexity.

Results: RHDEL was evaluated on 8-channel EEG data over 50 epochs, demonstrating stable convergence and improved accuracy from 63.13% to 90.62%. Compared to a standard recurrent neural network trained via BPTT, RHDEL achieved higher peak accuracy (90.62% vs. 88.45%) while reducing memory requirements from O(T) to O(1). The model effectively captures phase-synchronization dynamics, highlighting the suitability of directional statistics for oscillatory neural signals.

Conclusions: RHDEL presents a time-reversible, physics-informed approach to EEG classification that overcomes the memory limitations of traditional methods. By integrating Hamiltonian dynamics with directional statistics, it enables accurate and resource-efficient learning, making it well-suited for real-time, low-power applications and advancing the development of neuromorphic and wearable AI systems.

Climate-driven modelling of dengue incidence using statistical and machine learning approaches

Alzahrani, Sahar, Cardiff University; Jones, Owen, Cardiff University; Cable, Jo, Cardiff University; Malik, M. Wasif, Health Services Academy, Islamabad

Background: Dengue fever is a climate-sensitive vector-borne disease transmitted by mosquitoes, whose populations are influenced by climatic conditions. In the absence of direct mosquito data, climate variables are used as proxies to model dengue transmission. Understanding these relationships is important for improving early warning systems and informing public health interventions. This study investigates the impact of climate factors on dengue incidence using statistical and machine learning approaches.

Methods: We analysed dengue case data alongside climate variables over multiple years. A Negative Binomial Generalised Linear Model (NB-GLM) was used to account for overdispersion in count data and to provide interpretable estimates of associations between predictors and dengue incidence. In parallel, an Extreme Gradient Boosting (XGBoost) model was developed to explore non-linear relationships and interactions among variables. Model development was conducted using historical data, with recent data used for preliminary evaluation. Lagged climate variables were also considered to capture delayed effects on dengue transmission. The dataset was updated to include recent observations. For the XGBoost model, earlier data were used for training, while recent data were reserved for testing. Initially, climate variables were spatially averaged across districts. To better capture spatial heterogeneity, a refined approach was implemented by fitting separate XGBoost models for each district and aggregating predictions across districts.

Results: Preliminary findings indicate that climatic factors are key predictors of dengue incidence. The inclusion of lagged variables improved the identification of these relationships, suggesting delayed environmental effects on disease transmission. The NB-GLM provided interpretable insights, while the XGBoost model captured more complex patterns. The district-level modelling approach improved forecasting compared to spatially averaged models, highlighting the importance of spatial variability. Evaluation on recent data is ongoing.

Conclusions: This study highlights the role of climatic factors in dengue incidence and the value of combining statistical and machine learning approaches. While results are preliminary, climate-based modelling shows potential for supporting early warning systems, with further work focused on model refinement and validation.

Machine Learning-Based Prediction of Paediatric Kidney Graft Survival: A Multi-Model, Multi-Horizon Analysis of the UNOS Registry

Burns, Constance Ella, NIHR Great Ormond Street Hospital Biomedical Research Centre, Great Ormond Street Hospital (GOSH) for Children NHS Foundation Trust, London, United Kingdom; Marks, Stephen, NIHR Great Ormond Street Hospital Biomedical Research Centre, Great Ormond Street Hospital (GOSH) for Children NHS Foundation Trust, London, United Kingdom; Brugha, Rossa, GOSH Data Research, Innovation and Virtual Environments, London, United Kingdom; Cortina-Borja, Mario, University College London Great Ormond Street Institute of Child Health, London, United Kingdom.

Aims/Purpose: Evidence on machine learning (ML) applications in paediatric kidney transplantation is limited, and no validated ML-based prognostic models currently exist for graft survival prediction in this population. We aimed to develop and benchmark ML models to predict graft failure at 1, 5, and 10 years after kidney transplantation using the United Network for Organ Sharing (UNOS) registry.

Methods: Paediatric recipients (≤18 years) undergoing primary single-organ kidney transplantation recorded in the UNOS registry between 1987 and 2025 were included. Graft failure was defined as return to dialysis or re-transplantation. Following missForest imputation and Boruta-based feature selection, six algorithm families (GLMNET, random forest, XGBoost, LightGBM, multilayer perceptron, and k-nearest neighbours) were benchmarked under up sampling and down sampling strategies to address class imbalance. All models were optimised via 5×5 repeated cross-validation with ANOVA racing, with discrimination assessed by AUC-ROC, clinical utility evaluated by decision curve analysis, and predictor contributions quantified using SHAP values.

Results: A total of 21,789 paediatric kidney transplant recipients (pKTR) were included in the 1-year analysis, 17,218 at 5 years, and 13,331 at 10 years with graft failure occurring in 7.3%, 27.9%, and 55.1% of patients, respectively. Tree based models with class balancing achieved the best performance across all time horizons, with AUC-ROC values of 0.74, 0.69 and 0.68 at 1, 5 and 10 years, respectively. Decision curve analysis demonstrated a net clinical benefit across a range of threshold probabilities, supporting potential utility in risk stratification. Key predictors across all three horizons included recipient age, serum albumin, panel-reactive antibody level, cold ischaemia time, donor age, and donor type.

Conclusion: Machine learning models trained on routinely collected pre-transplant data can identify pKTR at risk of graft failure across the timepoints after transplantation, supporting individualised risk stratification at the point of transplantation. External validation in other independent cohorts is required to assess model generalisability.

A joint meta-analysis framework for the accuracy of two diagnostic tests accounting for varying study designs

Hudak, Vera, University of Bristol; Welton, Nicky J., University of Bristol; Derezea, Efthymia, University of Bristol; Jones, Hayley E., University of Bristol

Background: Standard methods for the joint meta-analysis of diagnostic test accuracy typically assume tests are independent conditional on true disease status. This is often unrealistic and, if violated, will lead to biased estimates of the accuracy of tests used in combination. Some models have been proposed which account for dependence of tests. These mostly rely on joint classification data (Test 1 vs Test 2 in both disease states), which is often not reported in study publications. Moreover, existing methods typically account for conditional dependence by modelling joint rates directly. These are constrained by marginal rates, which can lead to estimation issues. We propose modelling conditional dependence through unconstrained log-odds ratios to resolve these issues while incorporating studies with missing joint data or alternative designs.

Methods: We propose a Bayesian hierarchical modelling framework for the joint meta-analysis of two binary diagnostic tests. This retains key characteristics of standard diagnostic test accuracy methods, while additionally modelling conditional dependence through study-specific odds ratios within each disease state. While our framework requires some studies with joint classification tables to estimate within-study dependence, it naturally extends to include varied study designs by jointly modelling varying likelihoods. This includes studies where tests are evaluated on the same or different people, and where true disease status is unobserved for a subset of individuals. We apply the model to two case studies: ultrasound markers for identifying Down syndrome and the McIsaac clinical prediction rule and rapid antigen detection tests for identifying Group A Streptococcus.

Results: Case study applications show that when conditional dependence is high, ignoring it leads to significant bias in joint accuracy estimates. Our framework successfully integrated studies with varying designs without relying on data imputation.

Conclusions: Results demonstrate that accounting for test dependence is critical when jointly meta-analysing tests. Our framework provides a flexible approach for incorporating varying study designs, using a log-odds ratio parametrization for the conditional dependence of tests to ensure computational stability and avoid the estimation issues common in existing methods.

Modelling Multiplayer FPS Player Performance Using Hidden Markov Models

Terry, Brian, University of Cambridge; Didelot, Xavier, University of Warwick

Background: Modern multiplayer video games are driven by the competitive ranking of players to allow for the balance of teams. This balance affects the overall enjoyment of the game by players and is a major point of research for game developers. Meanwhile, Hidden Markov models (HMMs) are a powerful tool within time series analysis. Their basic premise is that a 'hidden' state layer produces observable emissions which we can use to infer the state, or if needed, predict future observations. Hidden Markov models have previously been used for modelling player behaviour and skill, with notable examples including Matsumoto and Thawonmas in their 2004 paper using HMMs to classify play style in a massively multiplayer online game; Zamith et al. investigate HMMs as a method to dynamically balance single player games; Tamassia et al. used HMMs to predict player churn and offer incentives to those about to leave a game indefinitely. Interestingly, all previous work in this field has approached this topic from a Frequentist framework and has so far avoided the direct question of game balance in a multiplayer context.

Methods: Our work has two major novelties to it. The first is using the Bayesian paradigm to deal with potentially low sample sizes in practical data sets. The second is that we apply a hidden Markov model to the 'performance' of modern multiplayer first-person shooter (FPS) players for the first time. Our research takes the ideas of Zamith et al. in trying to predict if a game should be more or less challenging and combines them with the analysis done by Tamassia et al. who analysed data from an FPS.

Results: For this project, we have obtained a dataset of multiplayer matches played in the video game Call of Duty: Black Ops 6. We were then able to fit an HMM to typical multiplayer FPS data to model player performance levels, using Gibbs sampling. We also predict player performance levels in future games.

Conclusions: Our research demonstrates the feasibility of the approach to the challenge of inferring individual player skill. While open questions remain about extending the model to full-lobby balancing, our work serves as a proof of concept that HMMs can be used as a flexible and efficient tool for estimating player skill, which can in turn support dynamic balancing in competitive games.

Revising Use of F-Tests in Weak Instrument Practice

Lane, James, MRC Biostatistics Unit University of Cambridge; Stephen Burgess, MRC Biostatistics Unit University of Cambridge

Background: Instrumental Variables (IVs) are a common device used in economics and medicine to infer the causal effect of a risk factor on an outcome. When the association between the risk factor and IV is small, issues of weak instrument bias arise. The conventional practice is to use F-Tests to provide sufficient statistical guarantee on IV strength. We aim to expose some key flaws with this practice hitherto overlooked, suggesting this conventional practice requires alteration.

Methods: We develop the Weak IV asymptotics first derived in Staiger & Stock (1997), by accounting for an additional variance term and accounting for the publication bias induced by conditioning on passing the F-Test. We compare mean squared error (MSE) of the two-stage least squares (2SLS) estimator used in IV analysis, using IV strengths the F-test provides sufficient statistical guarantee for, with that of the ordinary least squares (OLS) estimator. Accounting for publication bias creates a trade-off between F-test critical values big enough to statistically guarantee sufficient IV strength and publication bias being aggravated by larger critical values. To investigate this, we look at a 'Control Function' that calculates the type I error rate of the Wald test for the existence of a causal effect conditional on passing the F-test at a given critical value, using the IV strength the F-test provides sufficient statistical guarantee for.

Results: When F-test is passed with Stock & Yogo (2005) critical values, 2SLS is frequently outperformed by OLS in terms of MSE, even if the bias in OLS is bigger. The optimal critical value to minimise the Control Function is zero. Or equivalently, not to perform an F-test at all.

Conclusion: Conventional practice of using an F-test for IV strength, including the common F greater than 10 rule of thumb, needs substantial alteration. In particular, when the F-test is passed, the OLS estimator still frequently outperforms IV analysis. Moreover, selection bias induced by the F-test aggravates type I error rates, leaving one better off not performing an F-test at all than using currently recommended critical values. Research is required to find alternative solutions to guarantee sufficient IV strength.

Collaborative Targeted Maximum Likelihood Estimation for High-Dimensional Confounding: A Simulation Study and Julia Implementation

Sarpong, Juliet Asantewaa, University of Edinburgh; Beentjes, Sjoerd Viktor, University of Edinburgh; Khamseh, Ava, University of Edinburgh; Ponting, Chris P., University of Edinburgh

Background: Estimating causal effects from observational healthcare and genomic data is complicated by high-dimensional confounding, potential positivity violations, and model misspecification. Targeted Maximum Likelihood Estimation (TMLE), a standard semiparametric, doubly-robust estimator, can suffer in these settings because its propensity score is optimised independently of the causal target parameter, introducing instability when treatment probabilities become extreme. Collaborative TMLE (C-TMLE) addresses this by embedding outcome-guided variable selection directly into the estimation of the propensity score. C-TMLE constructs a nested sequence of candidate propensity models and uses cross-validated loss on the causal parameter to select the best-performing combination, aligning nuisance estimation with inferential goals.

Methods: Three C-TMLE variants were implemented from first principles in R: Greedy C-TMLE (forward stepwise covariate selection), C-TMLE1 (LASSO-regularised collaborative selection over a grid of penalty parameters), and C-TMLE0 (extended targeting via a second clever covariate accounting for EIF sensitivity to the penalty parameter). Simulation studies were conducted under high-dimensional conditions: p=100 covariates, Toeplitz correlation structure (ρ=0.9), n=5,000, and 100 replications. Performance was evaluated on bias, variance, coverage, and MSE. C-TMLE1 was re-implemented in Julia and integrated into TMLE.jl, with bootstrap procedures used to validate stability.

Results: All three C-TMLE variants outperformed standard TMLE in bias and coverage under high-dimensional confounding. C-TMLE1 achieved the lowest bias and MSE; C-TMLE0 the highest coverage (94%); and Greedy C-TMLE the lowest variance with the tightest confidence intervals, illustrating a clear bias-variance trade-off across variants. No single variant dominated across all criteria, suggesting selection should be guided by the inferential goal. Bootstrap validation confirmed stable performance of the Julia implementation in TMLE.jl.

Conclusions: C-TMLE offers a principled improvement over standard TMLE in high-dimensional settings by directly aligning propensity score estimation with the causal target parameter. The observed bias-variance trade-off across variants suggests that method choice should be guided by inferential priorities. The Julia implementation in TMLE.jl provides a scalable foundation for planned application to UK Biobank genomic data.

Quantifying the Long-Term Quality-of-Life Impact of ABO-Incompatible Transfusion Events Using SHOT Data and Clinical Evidence

Montasir Ahmed[1], Florian Tomini[1], Florence Oyekan[2,4], Laura Green[2,3,4], Yan Feng[1], Louise Bowles[2], Michael F Murphy[3,5,6], Nathan Proudlove[7], Josephine McCullagh[2,3,4]

1 Wolfson Institute of Population Health, Queen Mary University of London, London, UK; 2 Barts Health NHS Trust, London, UK; 3 NHS Blood and Transplant, UK; 4 Blizard Institute, Queen Mary University of London, London, UK; 5 Oxford University Hospitals NHS Foundation Trust; 6 University of Oxford; 7 Alliance Manchester Business School, The University of Manchester, Manchester, UK

Background: ABO-incompatible (ABOi) red blood cell transfusions are rare but preventable patient safety events associated with substantial morbidity and mortality. Our previous systematic review of 27 years of UK Serious Hazards of Transfusion (SHOT) showed that ABO-incompatible (ABOi) transfusions occur in 0.67 per 100,000 transfusions and are associated with 6.3% mortality and ~24% major morbidity. However, the longer-term health and economic burden of these events are not well understood.

Methods: We developed a decision-analytic model to estimate the per-event health and cost burden of confirmed ABOi transfusions using 27 years of individual-level data from the UK Serious Hazards of Transfusion (SHOT) scheme. Events were modelled using a decision tree with mutually exclusive clinical outcomes, from no reaction to major morbidity and death, with transition probabilities derived directly from pooled SHOT data. Health outcomes were measured in quality-adjusted life years (QALYs) using a piecewise area-under-the-curve approach that captures the acute and recovery phases, informed by EQ-5D evidence. Costs were estimated from the UK National Health Service and Personal Social Services perspective. The base-case analysis used a one-year time horizon. Parameter uncertainty was assessed using probabilistic sensitivity analysis, while scenario analyses explored assumptions about recovery and baseline quality of life.

Results: The expected cost per ABOi transfusion event was £3.486, with an associated health loss of 0.24 QALYs. Probabilistic sensitivity analysis showed that costs remained substantial, with a mean of £4.790 (95% uncertainty interval [UI] £3.848–£5.869, while the mean QALY loss was 0.19 (95% UI 0.15–0.24). Severe outcomes were infrequent but accounted for the majority of costs and QALY losses, particularly events involving acute kidney injury and critical illness. Less severe reactions contributed minimally to the overall burden. Results were robust across alternative modelling assumptions. Discussion By leveraging national haemovigilance data, this study provides policy-relevant inputs for economic evaluations of transfusion safety interventions and supports explicit consideration of opportunity costs in patient safety decision-making.

Decomposing the Uncertainty of Dynamic Risk Predictions using Value of Information

Petersen, Teun, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK; Chen, Sida, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK; Wood, Angela, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; Jackson, Christopher, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK; Barrett, Jessica, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

Quantifying the uncertainty of clinical risk predictions is essential for determining a model's reliability and showing the strength of evidence. In dynamic risk prediction, longitudinal trajectories are used to make risk predictions as new data becomes available. These are often estimated using joint models (JMs) for longitudinal and time-to-event data. JMs consist of two sub-models: a mixed-effects model to describe the repeated measurements (e.g. of a patient's biomarkers), and a Cox proportional hazards model, which are interlinked by shared random effects. The uncertainty in risk predictions from JMs stems therefore from multiple sources which could be disentangled for more effective clinical interpretation. We propose a framework to decompose the uncertainty of dynamic risk estimates into two key components: 1) patient-specific uncertainty, which arises from estimating an individual's trajectory from a limited number of longitudinal measurements, and 2) model-specific uncertainty, which is inherent to the underlying prediction model fitted on a finite dataset. To this end, we borrow the concept of the expected value of partial perfect information from value-of-information literature and apply it to the reduction in uncertainty as a consequence of knowing a patient's random effects. We apply this method to predicting hospital readmission for heart failure patients, using electronic health record data from 32,471 patients admitted for heart failure in the East of England region. Given that early heart failure rehospitalisation is frequent and might often be preventable, a clearer understanding of individual readmission risk and its sources of uncertainty could aid decisions on post-discharge care and monitoring strategies. High patient-specific uncertainty may indicate that collecting additional longitudinal measurements for a patient could improve the precision of their risk estimate. While high model-specific uncertainty highlights the limitations of the sample size of the JM. Illustrating this distinction allows clinicians to better assess the reliability of individual predictions, supports more informed, shared decision-making, and could improve sample size calculations for future dynamic risk models.

A note on diffusive/random-walk behaviour in Metropolis-Hastings algorithms

Zhou, Peiyi, University College London; Liu, Yuxin, University College London; Livingstone, Sam, University College London

Background: Metropolis-Hastings (MH) algorithms are widely used in Monte Carlo sampling, but their efficiency depends on how quickly the associated Markov chain mixes. Particularly, diffusive or random-walk behaviour can lead to slow exploration of the target distribution, whereas non-reversible methods with momentum may achieve faster, more ballistic motion. This research studies when MH algorithms inherit poor mixing from their proposal mechanism, and when non-reversible sampling methods, specifically the guided walk Metropolis (GWM), offer advantages over the classical random walk Metropolis (RWM).

Methods: The research first establishes a general theoretical result for MH algorithms in a high-acceptance regime. Specifically, it proves that if the proposal kernel is not geometrically ergodic and the acceptance probability goes to one sufficiently quickly in the tails, then the MH transition is also not geometrically ergodic, whereas a counterexample is constructed to show that a weaker condition based only on the average acceptance rate tending to one is not sufficient. We also study extensions where acceptance probability converges to one, but the target is asymptotically flat in some directions, connecting this to acceleration in sampling and optimisation. The analysis then focuses on two one-dimensional algorithms: the reversible Gaussian RWM and the non-reversible GWM, comparing their mixing behaviour under heavy, polynomial-tailed and light, strictly log-concave target distributions.

Results: The general theorem shows that high acceptance in the tail can cause the MH chain to behave too similarly to a non-geometrically ergodic proposal, hence preventing geometric ergodicity of the chain. For polynomial-tailed targets, we show that the GWM achieves a polynomial convergence rate twice than that of RWM. By contrast, for strictly log-concave targets, the random walk Metropolis behaves asymptotically like a (1/2)-lazy version of the guided walk Metropolis, so both algorithms move with similar ballistic speed in the tails.

Conclusions: The results show that the benefits of non-reversibility depend crucially on the tail structure of the target distribution. Non-reversible GWM method provides a clear advantage for heavy-tailed or asymptotically flat targets, but much less so for light-tailed, strictly log-concave targets, where reversible RWM and non-reversible (1/2)-lazy GWM dynamics become asymptotically similar.

Modeling Prediction Markets as Defaultable Bonds

Göransson-Gaspar, Erik, Lund University

During the last few years platforms have emerged that host a large number of liquid prediction markets spanning a range of topics. In this context, it is natural to model prediction market prices using tools developed for derivative pricing. I present the first price process-model suited to prediction markets that are based on questions of the type "Will event A occur before time T?". This model is adapted from the literature on valuing bonds with counterparty risk. The event underlying such a prediction market is assumed to arrive with a doubly stochastic Poisson process. By supposing that its intensity process is governed by jump-diffusion Cox-Ingersoll-Ross dynamics we get an analytic pricing equation. The process' parameters can be estimated using a Kalman-filter quasi-maximum likelihood procedure. I show that this model fits market data well by studying the coverage rates of approximate prediction intervals. Once fitted to some market, the model allows us to reconstruct and forecast the intensity driving event arrival. This characterization of the price dynamics allows for efficient market making, and the estimation of the probability of the underlying event arriving in any arbitrary time interval.

Re-randomisation tests in clinical trials: a simulation case study and practical guidance

Wilson, Huw, Veramed; Fletcher, Alice, Veramed

Background: At the end of a clinical trial, a key consideration is whether the results reflect a true treatment effect, or if they are in fact spurious and a Type 1 error has occurred. A re-randomisation test can help to distinguish between these possibilities. The method behind this test is straightforward: fix each participant's observed response, simulate many alternative treatment allocations under the original randomisation scheme, and then apply the same analysis for each simulation. If many of these simulations produce a result as extreme as the original, then it is likely that a Type 1 error has occurred (or the randomisation scheme somehow favours one treatment group over the other). Alternatively, if few of these simulations produce a result as extreme as the original, then confidence in the original result is increased. Methods In this case study, we will simulate data from a parallel-arm, placebo-controlled clinical trial using dynamic allocation with 4 stratification factors. Firstly, we will simulate a null scenario where no true treatment effect exists, and secondly, an alternative scenario where a true treatment effect does exist. For both scenarios, we will estimate the treatment effect difference from the observed results, apply a re-randomisation test, and then give an interpretation of the results.

Results: This case study will show how a re-randomisation test can successfully identify a Type I error, or alternatively, strengthen the credibility of the original result. We will also provide practical guidance and highlight points to consider when applying the test in a clinical trial context. For example, how to determine if the test is appropriate and how to prepare for the test while blinded (as the test can only be performed after unblinding).

Conclusions: Re-randomisation tests can be very useful at the end of a trial and are explicitly recommended for certain types of randomisation schemes by regulatory authorities such as the European Medicines Agency (EMA). However, care must be taken to ensure their correct application and interpretation.

Methods for Informative Priors via Predictives using Predictively Oriented and Generalised Bayes

Dunipace, Craig, MRC-BSU, University of Cambridge, Goudie, Robert, MRC-BSU, University of Cambridge

Direct prior elicitation on the parameter space is often difficult (Mikkola et al, 2024). To address this challenge, recent work has focused on eliciting predictive distributions on the observable space (Bockting et al,2024), then seeking to translate these distributions directly into informative priors on the parameter space. However, existing approaches are either limited to simple models or are highly computationally expensive. In this work, we highlight the connection between prior specification via elicited predictives and recent 'post-Bayesian' developments: specifically, inferentially-oriented Generalised Bayes (Bissiri et al, 2016) and Predictively Oriented Posteriors (McLatchie et al, 2025). Both approaches aim to condition prior parameter distributions on elicited predictive distributions. The Predictively Oriented approach has been proposed as a way to capture the epistemological uncertainty of a miss-specified model, a setting with relevance when predictives, or experts, express partially contradictory information. The Generalised Bayes maintains the interpretable inferential focus of standard Bayes and clearer interpretation of information quantification, via generalisations of ideas such as effective sample size, and greater flexibility in choice of loss function. We compare the behaviour of these approaches in several contexts: varying degrees of model misspecification, combinations of parameter and predictive elicitation, and predictives defined on different levels of marginalisation. Evaluation focuses on faithfulness to elicited information, computational tractability, and suitability of distributions as informative priors. We illustrate with a real-world biostatistical regression example with expert elicitations.

Divide and conquer sequential Monte Carlo for chained Markov melding

Liu, Yixuan, MRC Biostatistics Unit, University of Cambridge

A practical motivation for Bayesian inference is to construct models that integrate information from multiple heterogeneous data sources. To this end, the fully Bayesian analysis requires a joint model that incorporates all parameters and separated datasets. Thus, the joint model often becomes complex and computationally expensive. Alternatively, a more feasible way is to model the sources separately and subsequently combine all submodels. Within this framework, Goudie et al. (2019) propose Markov Melding, a generic method for integrating multiple sources that share a common parameter, and Manderson and Goudie (2023) later extend this approach to more general relationships between submodels. However, this method becomes challenging when the number of submodels is large. Our current work focuses on addressing this problem by adopting a Sequential-Monte-Carlo-based method proposed by Lindsten et al. (2017), which offers efficient sampling for hierarchical structures. Based on the real-world and simulation studies, we demonstrate that our method is competitive with the original Markov Melding approach in settings with three submodels and exhibits comparable performance to full MCMC in cases involving five submodels. Ideally, this new extension will remain robust, enabling applications to scenarios with an arbitrarily large number of submodels.

Equitable cardiovascular risk stratification in Type 2 Diabetes: development of a population-specific prediction model in a large, ethnically diverse London cohort

Schanhave, Santhirasekaran, Department of Population Health & Policy, St George's School of Health and Medical Sciences, City St George's University of London, UK; Zeinab, Schaefer, Department of Population Health & Policy, St George's School of Health and Medical Sciences, City St George's University of London, UK; Alasdair, Warwick, University College London Institute of Cardiovascular Science, London, UK, NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; John, Anderson, Homerton Healthcare NHS Foundation Trust, London, UK; Anna, Pouncey, Cardiovascular and Genomics Institute, City St George's, University of London, UK; Iain, Roy, Cardiovascular and Genomics Institute, City St George's, University of London, UK; Joan, Morris, Department of Population Health & Policy, St George's School of Health and Medical Sciences, City St George's University of London, UK; Alicja, Rudnicka, Department of Population Health & Policy, St George's School of Health and Medical Sciences, City St George's University of London, UK; Christopher, Owen, Department of Population Health & Policy, St George's School of Health and Medical Sciences, City St George's University of London, UK

Background: QRISK algorithms, particularly QRISK3, are recommended by the National Institute for Health and Care Excellence (NICE) to estimate 10-year cardiovascular disease (CVD) risk in adults aged 25–84 in England and Wales. Although widely implemented within the NHS Health Check programme, their performance is reduced in populations already at elevated risk, including individuals with type 2 diabetes (T2D). In addition, many risk prediction models are derived from predominantly White or non-UK populations, limiting their generalisability to ethnically diverse settings such as London. Given the rising prevalence of T2D and its strong association with CVD morbidity and mortality, there is a need for robust, population-specific prediction tools tailored to high-risk and diverse populations.

Methods: We conducted a retrospective cohort study using linked primary and secondary care electronic health records from WSIC/OneLondon. A total of 164,650 individuals aged 18–90 years with T2D, diagnosed between 1 January 2000 and 1 July 2025, were included after excluding those with prior CVD. During follow-up, 29,920 incident CVD events occurred. Flexible parametric survival models were used to estimate 5-year CVD risk, accommodating non-linear baseline hazards. Candidate predictors included traditional cardiovascular risk factors (systolic blood pressure, smoking status, total and HDL cholesterol), diabetes-specific variables (age at diagnosis, glycated haemoglobin [HbA1c], and estimated glomerular filtration rate [eGFR]), and medication use (antihypertensives, lipid-lowering therapies, and antiplatelet agents). Missing data were addressed using multiple imputation by chained equations, and continuous variables were modelled using appropriate functional forms.

Results: Apparent validation demonstrated moderate discrimination, with Harrell's C-index ranging from 0.64 to 0.67 across models. This is consistent with the performance of general-population tools such as QRISK3 when applied to high-risk groups, but reflects estimates derived within a T2D-specific, ethnically diverse cohort. Calibration and subgroup analyses, including stratification by ethnicity and deprivation, are ongoing.

Conclusions: We present the development and initial validation of a population-based model for predicting 5-year CVD risk in individuals with T2D in a large, diverse UK cohort. This approach supports more accurate and equitable risk stratification in a high-risk population.

Predicting MACE and Mortality in Type 2 Diabetes: A DeepHit Competing Risks Model in North West London EHR Data

Zeinab, Schaefer, Department of Population Health & Policy, St George's School of Health and Medical Sciences, City St George's University of London, UK; Schanhave, Santhirasekaran, Department of Population Health & Policy, St George's School of Health and Medical Sciences, City St George's University of London, UK; Alasdair, Warwick, University College London Institute of Cardiovascular Science, London, UK, NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; John, Anderson, Homerton Healthcare NHS Foundation Trust, London, UK; Anna, Pouncey, Cardiovascular and Genomics Institute, City St George's, University of London, UK; Iain, Roy, Cardiovascular and Genomics Institute, City St George's, University of London, UK; Joan, Morris, Department of Population Health & Policy, St George's School of Health and Medical Sciences, City St George's University of London, UK; Alicja, Rudnicka, Department of Population Health & Policy, St George's School of Health and Medical Sciences, City St George's University of London, UK; Christopher, Owen, Department of Population Health & Policy, St George's School of Health and Medical Sciences, City St George's University of London, UK;

Background: Type 2 diabetes mellitus (T2DM) increases the risk of major adverse cardiovascular events (MACE), but predicting individual-level risk in routine clinical populations remains challenging. In competing-risk settings, death before MACE removes patients from the risk set, and treating mortality as independent censoring, as in Cox models, can overestimate cumulative incidence. Conventional survival models also rely on proportional hazards and log-linear assumptions that are difficult to justify in high-dimensional electronic health record (EHR) data. We therefore applied DeepHit, a deep learning survival framework that jointly models MACE and all-cause mortality while capturing non-linear relationships without parametric assumptions.

Methods: We identified patients with T2DM from the North West London EHR dataset using a 2-year landmark design, retaining individuals event-free at the landmark date. The final cohort included 55,438 patients: 4,391 experienced MACE (7.9%), 4,919 died without prior MACE (8.9%), and 46,128 were censored (83.2%) over a mean follow-up of 7.1 years. Features included demographics, BMI and HbA1c at diagnosis and landmark, biomarker change over time, cardiovascular medication exposure, and GLP-1 receptor agonist use. DeepHit's shared and cause-specific architecture generated event probabilities across 39 discrete time bins with a joint softmax layer enforcing competing-risk constraints. The model was trained using a 70/15/15 train-validation-test split with weighted log-likelihood and ranking losses. Feature contributions were assessed using SHAP values.

Results: On the held-out test set (n=8,316), DeepHit achieved a C-index of 0.656 for MACE and 0.654 for all-cause mortality, with integrated Brier scores of 0.164 and 0.213, respectively. Predicted cumulative incidence curves separated clearly across risk quintiles. SHAP analysis identified age, glycaemic trajectory, and pre-diagnosis cardiovascular medication burden as major predictors. GLP-1 receptor agonist use contributed minimally, likely reflecting low exposure prevalence (1.4%, n=789).

Conclusions: DeepHit provided well-discriminated competing-risk predictions for MACE and all-cause mortality in a large real-world T2DM cohort. Future work will focus on missing-data imputation, feature enrichment, and propensity-score-matched comparisons of GLP-1 users and non-users.

Using individual participant data meta-analyses to explore the effectiveness of weight loss interventions for people with knee OA

Kavi Mathiyalagan, Ram Bajpai, Emma Healey, Danielle van der windt, Mel Holden

Background: Knee osteoarthritis (OA) is a leading cause of pain and disability worldwide and commonly coexists with overweight and obesity, which worsen symptoms and disease progression. International guidelines recommend weight management as a core treatment; however, while randomised controlled trials (RCTs) consistently demonstrate weight loss, effects on pain and physical function are inconsistent. These mixed results may arise from heterogeneity in individual treatment response and uncertainty over the amount of weight loss required for clinically meaningful benefit. Conventional meta-analyses and single trials are underpowered to investigate these questions.

Aims: To identify individual-level characteristics associated with short-, medium-, and long-term effects of weight loss interventions on pain, physical function, and weight loss in people with knee OA who are overweight or obese, and to determine the amount of weight loss needed to achieve clinically meaningful improvements in symptoms.

Methods: A systematic review and individual participant data meta-analysis (IPD-MA) will be conducted in accordance with PRISMA-IPD guidelines. RCTs comparing interventions primarily focused on weight loss with non–weight loss controls will be included. Participants will be adults aged ≥45 years with knee OA and ethnicity-adjusted overweight or obesity. Pain and physical function outcomes will be harmonised and standardised to a 0–100 scale. Analyses will follow a two-stage IPD-MA approach using random-effects models. Trial-specific linear mixed-effects models will estimate intervention effects and treatment–covariate interactions, which will then be pooled using restricted maximum likelihood. Prespecified moderators include age, sex, ethnicity, deprivation, body mass index, adiposity, mental health, OA severity, and comorbidities. Dose–response relationships between achieved weight loss and outcomes will be examined using mixed-effects regression, and results interpreted using established minimal clinically important differences.

Conclusion: This IPD-MA will clarify which individuals benefit most from weight loss interventions and the degree of weight loss required to achieve meaningful improvements in knee OA symptoms.

Progression of Multiple Long-Term Conditions in Type 1 and Type 2 Diabetes: Insights from A CPRD Matched Cohort Study

S Hassen, MPH [1,2], J Miksza, MSc[1,2], S Shabnam, PhD[1,2], G Kacha[1], Y Chudasama, PhD[1,2], F Zaccardi, PhD[1,2], P Choudhary, PhD[1], N Kaur, PhD[1,2], EW Gregg, PhD[4], Kamlesh Khunti, MD[1,2], Cameron Razieh, PhD[1,2], Lauren O'Mahoney, PhD[1,3], Clare L Gillies, PhD[1,2]

1 Diabetes Research Centre, University of Leicester, Leicester, UK; 2 Leicester Real World Evidence Unit, Diabetes Research Centre, University of Leicester, Leicester, UK; 3 NIHR Leicester Biomedical Research Centre, Leicester, UK; 4 School of Public Health, Imperial College London, London, UK

Background: People with diabetes frequently develop additional long-term conditions, with the co-existence of two or more defined as multiple long-term conditions (MLTCs). This study aimed to compare the prevalence and trajectories of MLTCs in people with type 1 diabetes (T1DM) and type 2 diabetes mellitus (T2DM).

Methods: A retrospective cohort study was conducted using linked primary and secondary care data from the UK Clinical Practice Research Datalink (CPRD). Individuals with T1DM and T2DM were matched with up to two controls without diabetes on age, sex, and general practice. MLTC burden was categorised from ≥1 to ≥5 conditions and followed up for 15 years. Regression models were used to examine age-adjusted trajectories and patterns of MLTC accumulation.

Results: The study included 55,288 individuals with T1DM and 422,912 with T2DM. At all time points, MLTC prevalence was higher among individuals with diabetes than among controls. By 15 years of follow-up, 61·2% of individuals with T1DM and 82·5% of those with T2DM had three or more long-term conditions, compared with 24·1% and 41·4% in their respective matched control groups. Age-adjusted models showed greater MLTC accumulation in T2DM compared with T1DM, with a difference of 0·49 conditions at 15 years (95% CI 0·41, 0·57). Higher burden of MLTCs was observed among females, individuals living in more deprived areas, and those with longer follow-up. Hypertension was the most prevalent condition in both diabetes types, although cardiometabolic conditions were more common in T2DM and autoimmune conditions relatively more frequent in T1DM.

Conclusions: Individuals with T1DM and T2DM experience a substantial and progressive burden of MLTCs, with more rapid accumulation in T2DM. Our findings extend evidence beyond T2DM to show that MLTCs is also an early and sustained feature of T1DM, highlighting the need for earlier and more integrated approaches to managing multimorbidity in diabetes.