

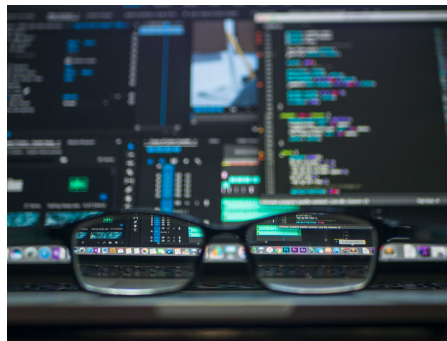


THE UNIVERSITY *of* EDINBURGH
Centre for Statistics

CENTRE FOR STATISTICS ANNUAL CONFERENCE 2026

June 19, 2026

Nucleus Building, 1.15 Larch Lecture Theatre
University of Edinburgh, King's Buildings, EH9 3FD



Hosted by the **Centre for Statistics**

<https://centreforstatistics.maths.ed.ac.uk>

Organised by Jordan Richards (jordan.richards@ed.ac.uk), Sara Wade (sara.wade@ed.ac.uk), and Timothy Cannings (timothy.cannings@ed.ac.uk).

The **Centre for Statistics** unites data-driven researchers from across the University of Edinburgh and Associated Institutions. We organise several **cross-disciplinary research days** throughout the academic year, provide free statistical advice at our regular **consultancy drop-in clinic**, and host an annual distinguished lecture series in honour of **Prof David Finney**.

Our **Annual Conference** showcases the interdisciplinary data-driven research being carried out in and around Edinburgh. The one-day event features a number of **invited speakers** from academia and industry, a **poster session**, and plenty of time for **networking**.

Centre for Statistics

THE UNIVERSITY *of* EDINBURGH

Edinburgh, UK

<https://centreforstatistics.maths.ed.ac.uk>

Programme:

- 09:00-09:25: Registration.
- 09:25-09:30: Housekeeping and Welcome (Jordan Richards, School of Mathematics).
- Session 1:
 - 09:30-10:05: Marie Michaelides (Heriot-Watt) *Balance and fairness through multicalibration in nonlife insurance pricing*
 - 10:05-10:40: Nadine Seward (Centre for Clinical Brain Sciences) *How and for Whom Do Perinatal Depression Interventions Work in LMICs? A Causal Inference and Machine Learning Analysis of Socioeconomic and Psychosocial Drivers*
- 10:40-11:10: Coffee break.
- Session 2:
 - 11:10-11:45: Steven McDonagh (School of Engineering) *Navigating Reliability and Vulnerability in Generative Paradigms*
 - 11:45-12:20: Adriana Duta (Moray House School of Education and Sport) *Do Children Start School on Equal Footing? Social Inequalities in Early Childhood Cognitive and Non-Cognitive Outcomes*
- 12:20-14:20: Posters and Lunch/photo.
- Session 3:
 - 14:20-14:55: Xuebin Zhao (School of Geosciences) *Variational Bayesian Seismic Imaging and Efficient Assessment of Prior Hypotheses in a Spatially 3D World*
 - 14:55-15:30: Edoardo M. Ponti (School of Informatics) *Adaptive Foundation Models for Efficient and Long-Horizon AI*
- Session 4 - Snap talks (School of Mathematics):
 - 15:30-15:45: Lambert de Monte *Multivariate extremal index and temporal clustering in multivariate extremes: A case study of Dutch hydrological extremes*
 - 15:45-16:00: Lanxin Li *Estimating hidden population sizes using multiple systems estimation with censored data*
 - 16:00-16:15: Claudia Collarin *Scalable Estimation of Large Generalized Additive Models*
 - 16:15-16:30: Torben Sell *Priors in Bayesian Deep Learning*
- 16:30-16:35: Closing (Jordan Richards, School of Mathematics).

Invited talks are 25 min + 10 min Q&A and changeover. Snap talks are 12 min + 3 min Q&A and changeover. The full programme can be found on the CfS 2026 website:

<https://centreforstatistics.maths.ed.ac.uk/events/upcoming-events/cfs-annual-conference-2026>

Invited talks feature speakers from six Schools and Institutes at the **University of Edinburgh** and Maxwell Institute, including the School of Engineering, School of Geosciences, School of Informatics, School of Mathematical and Computer Sciences (Heriot-Watt), Centre for Clinical Brain Sciences, and Moray House School of Education and Sport.

Balance and fairness through multicalibration in nonlife insurance pricing

Marie Michaelides (m.michaelides@hw.ac.uk)

School of Mathematical and Computer Sciences, Heriot-Watt

Abstract: Autocalibration is known to be an important requirement for insurance premiums since it guarantees that premium income balances corresponding claims, on average, not only at portfolio level but also inside each group paying similar premiums. Also, fairness has become a major concern because unfair treatment may expose insurers to lawsuits or reputational damage. Translating fairness into conditional mean independence allows actuaries to combine autocalibration and fairness into the multicalibration concept. This paper studies the properties of multicalibration in an insurance context and proposes practical ways to implement it, through local regression or bias correction within groups including credibility adjustments. A case study based on motor insurance data illustrates the relevance of multicalibration in insurance pricing.

[†]**Time slot:** 09:30-10:05

How and for Whom Do Perinatal Depression Interventions Work in LMICs? A Causal Inference and Machine Learning Analysis of Socioeconomic and Psychosocial Drivers

Nadine Seward (nseward@ed.ac.uk)

Centre for Clinical Brain Sciences, University of Edinburgh

Abstract:

Authors

Nadine Seward, Alexis Elias, Ahmed Waquas, Wen Wei Loh, Ricardo Araya, Angus McBeth, Charlotte Hanlon*, Crick Lund*

Background

Numerous randomized trials have evaluated the effectiveness of psychological interventions to treat perinatal depression in low- and middle-income countries (LMICs), yielding highly heterogeneous findings. This variability is unsurprising given the diverse intervention architectures and distinct socio-cultural contexts in which they are implemented. Utilizing robust causal inference and causal machine learning frameworks to analyse individual and pooled trial datasets can reveal the mechanisms (how) and specific subgroups (for whom) these interventions work, ultimately providing actionable insights to optimize global maternal mental health strategies.

Methods

We systematically searched MEDLINE, Embase, PsycINFO, CINAHL, Web of Science, and the Cochrane Library from database inception to November 19, 2025, for randomised controlled trials evaluating psychological interventions for perinatal depression in LMICs. For trials where investigators provided individual participant data (IPD), we assessed treatment effect heterogeneity (moderation) using causal forests and mechanisms (mediation) via interventional effects. IPD were subsequently pooled across clinically relevant trials to conduct moderation and mediation analyses mirroring the single-trial approaches.

Findings

We identified 11 eligible studies, seven of which provided IPD, yielding a total pooled sample of 3,154 participants (1,700 intervention, 1,454 control). Heterogeneity analyses demonstrated that the most vulnerable participants, specifically those with lower educational attainment, poor social support, and food insecurity, derived the greatest benefit from the interventions. Causal mediation analyses revealed that the primary mechanism was through enhanced social support. Furthermore, we found

evidence that reductions in debt and food insecurity mediated improvements in depressive symptoms.

Interpretation

Our findings indicate that psychosocial support and reduces levels of poverty are critical pathways to recovery from perinatal depression in resource-constrained settings. Although psychological interventions rarely address structural poverty directly, these results suggest that targeting socioeconomic determinants more aggressively, rather than focusing exclusively on clinical symptomatology, will substantially amplify treatment impacts and accelerate symptom reduction.

†**Time slot:** 10:05-10:40

Navigating Reliability and Vulnerability in Generative Paradigms

Steven McDonagh (s.mcdonagh@ed.ac.uk)

School of Engineering, University of Edinburgh

Abstract: As generative models find application across increasingly diverse and high-stakes fields, they provide a basis for probing the state of reliable and persistent data-driven reasoning. We consider a unifying challenge of concept preservation: ensuring that underlying semantics remain invariant even when subjected to sparse labelling, adversarial manipulation, or dynamic environmental changes. In this talk, I will cover some recent works that examine how probabilistic frameworks can leverage weakly labelled multimodal data, probe vulnerabilities, and simulate persistent environments. We firstly address clinical AI reliability, using weakly labelled data to bridge visual abnormality detection and report generation for more accurate diagnostics. Probing the robustness of these semantic correspondences leads us to redefine the nature of adversarial threats in an era of generative models, showing that classifiers can be misled by concept-based samples that maintain semantic meaning. Finally, we extend generative paradigms to world simulation by discussing how the evolution of latent 3D scene states can replace pixel-based context windows to ensure geometric and temporal consistency.

†**Time slot:** 11:10-11:45

Do Children Start School on Equal Footing? Social Inequalities in Early Childhood Cognitive and Non-Cognitive Outcomes

Adriana Duta (Adriana.Duta@ed.ac.uk)

Moray House School of Education and Sports, University of Edinburgh

Abstract: This talk presents statistical evidence on how early-life inequalities shape children's cognitive and socio-emotional outcomes. Drawing on large-scale longitudinal studies from the United Kingdom (UK Household Longitudinal Study, UKHLS; Millennium Cohort Study, MCS) and Ireland (Growing Up in Ireland, GUI), I show how methods such as multilevel modelling and mediation analysis can be used to disentangle the complex role of family background in early childhood outcomes.

First, using a sibling random-effects design with UKHLS data, I demonstrate how much of children's socio-emotional difficulties at age five can be attributed to shared family environments, both observed and unobserved. By decomposing family background into multiple dimensions such as socio-economic circumstances, parenting practices, emotional environment, I highlight which aspects matter most and how much remains unaccounted for. Second, I turn to cross-country comparative analysis to examine inequalities in early cognitive skills at school entry. I explore how factors such as the home learning environment and early childhood education mediate the relationship between family resources and children's skills, and how these processes differ between Scotland and Ireland.

Across both studies, a consistent message emerges: inequalities are already substantial before formal schooling begins, and no single factor explains them. Instead, statistical evidence points to a web of influences operating at different levels and stages of childhood. Overall, the talk illustrates how quantitative methods can inform policy-relevant questions by identifying when and where interventions may be most effective, while remaining accessible to a broad audience interested in the real-world impact of data.

†**Time slot:** 11:45-12:20

Variational Bayesian Seismic Imaging and Efficient Assessment of Prior Hypotheses in a Spatially 3D World

Xuebin Zhao (Xuebin.Zhao@ed.ac.uk)

School of Geosciences, University of Edinburgh

Abstract: Three-dimensional seismic full waveform inversion (3D FWI) is a highly nonlinear and computationally demanding inverse problem that constructs 3D subsurface seismic velocity structures using seismic waveform data. To characterise non-uniqueness in the solutions, we demonstrate Bayesian 3D FWI using an efficient method called physically structured variational inference, and apply it to 3D acoustic Bayesian FWI. The results provide reasonable posterior uncertainty estimates, at a computational cost that is only an order of magnitude greater than that of standard, deterministic FWI. Furthermore, we deploy variational prior replacement to calculate Bayesian solutions corresponding to different classes of prior information at low additional cost. We further analyse a variety of prior hypotheses by constructing Bayesian L-curves, which reveal the sensitivity of the inversion process to different prior assumptions. To our knowledge, this is the first study that allows such prior hypotheses to be compared in probabilistic 3D FWI at feasible computational cost. This work shows that fully probabilistic 3D FWI can be performed and can be used to test different prior hypotheses, at a cost that may be practical, at least in small problems.

† **Time slot:** 14:20-14:55

Adaptive Foundation Models for Efficient and Long-Horizon AI

Edoardo M. Ponti (eponti@ed.ac.uk)

School of Informatics, University of Edinburgh

Abstract: The recent revolution in generative AI has been driven by the rapid growth in training and inference compute for Foundation Models (FMs). This scaling, however, brings unintended consequences, including high energy demand and associated carbon emissions. My research sets out to reverse this trend by targeting a fundamental inefficiency in dominant FM architectures such as Transformers. Currently, these map data into sequences of internal representations, whose length bottlenecks both prompt processing (compute-bound by typically quadratic attention) and output generation (memory-bandwidth bound by key-value cache reads). Yet this length is largely determined upfront by input segmentation (tokenisation) and typically remains fixed across layers; during decoding, it also accumulates unchanged in the key-value cache. I have prototyped a new class of FM architectures that learn, end-to-end, to compress sequences of internal representations, effectively redefining the model's "atomic units" for processing and memorising information. To accelerate adoption, I have repurposed existing state-of-the-art open-weight FMs (such as Qwen 3 in collaboration with NVIDIA and OLMo in collaboration with AI2) into adaptive variants and released them publicly. This leads not only to substantial gains in efficiency (with $8\times$ speedups without accuracy degradation) but also to the emergence of new capabilities: adaptive FMs can operate over broader effective horizons, as they can perceive longer inputs and generate longer outputs under a fixed budget. This enables (1) lifelong learning via a permanent, sub-linearly growing memory, (2) inference-time hyper-scaling for reasoning-intensive tasks (maths, science, coding), and (3) enhanced world modelling for multimodal planning and simulation. Adaptive FMs thus open a path towards greener, more capable generative AI.

† **Time slot:** 14:55-15:30

Snap talks feature four early career researchers from the School of Mathematics.

Multivariate extremal index and temporal clustering in multivariate extremes: A case study of
Dutch hydrological extremes

Lambert de Monte (l.demonte@ed.ac.uk)

School of Mathematics, University of Edinburgh

Abstract: The (relatively) recent re-emergence of a geometric approach to multivariate extreme value theory has led to new theoretical and methodological considerations for the study of extreme multivariate events. However, most of these recent methods have neglected the important issue of temporal dependence which may induce clustering of extreme events. In this talk, I will introduce the directional multivariate extremal index (MEI), a measure of the severity of clustering of extreme events which is applicable to all of these recent methods, and propose a Bayesian inference method based on the SPDE/GMRF framework. The usefulness of the MEI is demonstrated on an in-depth case study of compound hydrological extremes of the Rhine and the Meuse rivers for which both low and high discharges have important downstream impacts on Rotterdam in the Netherlands. This is joint work with Bas Matsuura and Ioannis Papastathopoulos.

†**Time slot:** 15:30-15:45

Estimating hidden population sizes using multiple systems estimation with censored data

Lanxin Li (Lanxin.Li@ed.ac.uk)

School of Mathematics, University of Edinburgh

Abstract: Estimating the size of hidden or hard-to-reach populations, such as drug users, is an important problem in epidemiology and public health. In such situations, multiple systems estimation (MSE) is often applied. In MSE, multiple data lists observing individuals in the target population are linked and collated to form an incomplete contingency table, providing the number of individuals observed by each distinct combination of lists. Log-linear models are fitted to the observed cell counts, allowing for interactions between different lists. The number of individuals not observed by any list can then be estimated from the fitted model and combined with the number of observed individuals to provide an estimate of the total population size.

The observed cell counts, however, are often imperfect or only partially available. One common issue in epidemiological data is that lists may include individuals who are not members of the target population, leading to censored cell counts that provide only upper bounds on the true cell counts. Using the censored counts for statistical modelling leads to biased population size estimators. We extend the standard log-linear modelling approach in the classical framework, taking into account the censoring information, to obtain estimates of the corresponding total population size as well as the proportion of the target individuals identified by the censored lists. We demonstrate the substantial level of bias that can be introduced when ignoring the censoring for simulated data; before applying the approach to real data to estimate the number of individuals with opioid use disorder within New York State.

†**Time slot:** 15:45-16:00

Scalable Estimation of Large Generalized Additive Models

Claudia Collarin (claudia.collarin@ed.ac.uk)

School of Mathematics, University of Edinburgh

Abstract: In large generalized additive models (GAMs) with high-dimensional parameter vectors, forming and factorizing the penalized Hessian incurs an $O(np^2)$ computational cost, creating a bottleneck that makes model estimation impractical. To overcome this issue, we combine the generalized Fellner–Schall smoothing parameter update with stochastic trace estimation (such as Hutch++) and preconditioned conjugate gradient methods, thereby avoiding both the construction and the Cholesky factorization of the GAM penalized Hessian. The resulting procedure relies only on matrix–vector products, enables low memory-bandwidth parallelization, exploits model term sparsity with minimal fill-in, and achieves an $O(np)$ computational cost. The performance of the approach is demonstrated on the NMMAPS respiratory mortality data with over one million observations and more than 20,000 coefficients, fitted in just over half an hour.

†**Time slot:** 16:00-16:15

Priors in Bayesian Deep Learning

Torben Sell (torben.sell@ed.ac.uk)

School of Mathematics, University of Edinburgh

Abstract: In classical Bayesian statistics, a prior represents our initial beliefs about the parameters of a given statistical model before observing data. In this snap talk, I will give an overview of different priors in Bayesian deep learning, with a particular focus on Bayesian neural networks. I will highlight computational and practical considerations when choosing a prior, and illustrate how to include Bayesian statistics in an applied machine learning context.

†**Time slot:** 16:15-16:30

CFS Annual Conference 2026: Contributed Posters

We welcome 14 contributed poster presentations from a variety of Schools and Institutions across the University of Edinburgh, as well as Cytel, UK. Posters will be presented outside of the Larch Lecture theatre during the lunch break, 12:20–14:20.

Interpreting Physics-Informed Neural Networks
Michael Herrmann (michael.herrmann@ed.ac.uk)
School of Informatics, University of Edinburgh

Bayesian Kolmogorov-Arnold Neural Model with Picking-and-Pruning Prior
Johnny Lee (johnny.myungwon.lee@ed.ac.uk)
School of Mathematics, University of Edinburgh

Modelling Tourism Demand Using Generalised Additive Models
Haoxiang Xie (Haoxiang.Xie@ed.ac.uk)
Business School, University of Edinburgh

Disentangling non-crossover and crossover interaction types
between genetic and environmental effects
Daniel Tolhurst (dtolhurs@ed.ac.uk)
The Roslin Institute, University of Edinburgh

Collaborative Targeted Learning in Genomic
Juliet Asantewaa Sarpong (jsarpong@ed.ac.uk)
School of Mathematics, University of Edinburgh

The generalized underlap coefficient
Zhaoxi Zhang (Z.Zhang-156@sms.ed.ac.uk)
School of Mathematics, University of Edinburgh

CART for Survival Trees vs Traditional Cox Models:
A Simulation Study in Oncology Trial Settings
Conan Yunxi Liang (Zconanliang4@gmail.com)
Cytel, UK

Noise-Aware Clustering via Predictive Resampling and Posterior Summarisation
Lily Qing Li (qli10@ed.ac.uk)
School of Mathematics, University of Edinburgh

Multidimensional (Dis)advantage and its impact on School Engagement from Primary to
Secondary School in Scotland

Kerr Lumsden (K.Lumsden-1@sms.ed.ac.uk)

Moray House School of Education and Sport, University of Edinburgh

Evaluating re identification risks scores in publicly available clinical trial datasets:
Insights and implications

Aryelly Rodriguez (aryelly.rodriguez@ed.ac.uk)

Usher Institute, University of Edinburgh

Realising a global lidar for biomass change mapping

Steven Hancock (steven.hancock@ed.ac.uk)

School of Geosciences, University of Edinburgh

Tricks to improve MCMC convergence

James Chok (james.chok@outlook.com)

School of Mathematics, University of Edinburgh

Stochastic Climate Change Scenarios and Forecasts

Gian Luca Vriz (gvriz@ed.ac.uk)

Business School, University of Edinburgh

Reverend Bayes, we have a problem - and a solution

Andrew Curtis (andrew.curtis@ed.ac.uk)

School of Geosciences, University of Edinburgh