CDSDM 2026

Conference on Data-driven Statistical Decision Making

27–28 March 2026

Department of Mathematics

Indian Institute of Technology Hyderabad

ABSTRACT BOOK

Key Theme Areas

- Finance
- Reliability Engineering
- Clinical Studies
- Genomics
- Methodology

Table of Contents

21	**Ms. Nandini Bhosale, Pfizer**	*Efficacy optimized eligibility criteria with improved generalizability of emulated trial results*	Clinical Studies
22	**Dr. Palash Ghosh, IIT Guwahati**	*From Side Effects to Safer Care: Monitoring Drug Safety - A Statistical Lens on Pharmacovigilance*	Clinical Studies
23	**Dr. Damitri Kundu, Eli Lily**	*PREDOSE : Pharmacometrically-Refined Early-phase Dose Optimization design for Oncology Study Enhancement*	Clinical Studies
24	**Dr. Anirban Mitra, Johnson & Johnson**	*Declaring Doses as Safe in Ongoing Oncology Dose-Escalation Trials: Are They Truly Safe? A Critical Assessment of Safety Criteria*	Clinical Studies
25	**Dr. Satya Prakash Singh, IIT Kanpur**	*Exploiting Symmetry in Optimization: Applications of the Purkiss Principle to Optimal Experimental Design*	Modelling/Optimization
26	**Dr. Arnab Hazra, IIT Kanpur**	*Large Wave Direction Data Modeling Using Wrapped Spatial Gaussian Markov Random Fields*	Modelling/Optimization
27	**Dr. Moumita Chatterjee, ISI Hyderabad**	*Modelling alternately recurring events: a frailty based hazard estimation approach*	Modelling/Optimization
28	**Dr. Achintya Roy, NIT Warangal**	*Reliability study of a parallel system equipped with two warm standby components*	Modelling/Optimization

#01 INAUGURAL

Statistics and Operations Research in Applied Work: Reflections from Theory to Practice

Name	Prof. GSR Murthy
Affiliation	Indian Statistical Institute Hyderabad

ABSTRACT

The ideas presented in this talk stem from my four decades of experience as an academician and a consultant to industry. I wish to emphasise that Statistics and Operations Research (OR) are inseparable in applied work, and that academicians working closely with industry must be strong in both for effective impact. I reflect on my experience as an applied statistician and OR professional, balancing theoretical research with real-world applications. A central message I convey is that applied work itself involves substantial research and innovation—often through simple but effective and elegant ideas—and that my real satisfaction has come from providing elegant solutions to complex problems, even when such work has faced barriers in traditional publication. I shall present some important contributions, with a focus on the intelligent tricks that made each successful, hopefully inspiring young scientists in academia and industry.

#02 RELIABILITY ENGINEERING

Convex Ordering in Gamma Family and its Application in the Problem of Selecting the Most Reliable System

Name	Prof. Neeraj Misra
Affiliation	Indian Institute of Technology Kanpur
Subject Area	Reliability Engineering

ABSTRACT

We will discuss the Convex Ordering among distributions in gamma family and illustrate its application in the problem of selecting the most reliable system, where system lifetimes follow independent gamma distributions.

#03 RELIABILITY ENGINEERING

On Some Important Reliability Aspects of Coherent Systems

Name	Dr. Tanmay Sahoo
Affiliation	Indian Institute of Technology Palakkad
Subject Area	Reliability Engineering

ABSTRACT

Developed Sequential Order Statistics (DSOS) provide a flexible framework for modeling lifetimes of systems with dependent components, particularly when the failure of one component affects the performance of the remaining survivors. We establish several stochastic comparison results for DSOS in both one-sample and two-sample settings. Various ageing properties of DSOS are also investigated. Our results extend to generalized order statistics and ordinary order statistics under dependence, thereby unifying and broadening existing ordered models.

In an enormous number of real-world systems, each component makes a distinct contribution to the system, and the system's performance is decided not just by its working but also by the sum of the contributions made by all of its components. To capture such heterogeneity, weighted k-out-of-n; systems were introduced, where each component is assigned a weight and system failure occurs when the total weight of failed components exceeds a predefined threshold. However, this measure is not meaningful for an arbitrary coherent structure as it does not involve the structure of the system. To overcome this drawback, we introduce here a new notion of performance measure (namely, the structural capacity) and then define three different notions of random weighted coherent systems, namely, Type-I, Type-II and Type-III systems. We then study various reliability aspects of these systems.

#04 GENOMICS

Genetic architecture of complex traits in the Indian population

Name	Dr. Bratati Kahali
Affiliation	Indian Institute of Science, Bangalore
Subject Area	Genomics

ABSTRACT

Large-scale genomic studies in underrepresented populations, such as India are essential to uncover population-specific genetic architecture and improve global risk prediction. Here, we elucidate the genetic underpinnings of cardiometabolic traits in ~8,000 Indians. We identify novel genome-wide significant loci, several showing marked population specificity. Fine-mapping highlights functionally relevant variants with distinct linkage disequilibrium patterns stratified by linguistic groups. Comparative analyses with ancestry-matched datasets reveal heterogeneity in effect sizes at key loci. Together, these findings underscore the genetic diversity of the Indian subcontinent and the need for population-specific genomic discovery.

#05 GENOMICS

Statistical methods for integrating genomics data to identify pleiotropy across complex traits

Name	Dr. Samsiddhi Bhattacharjee
Affiliation	National Institute of Biomedical Genomics
Subject Area	Genomics

ABSTRACT

Genome-Wide Association Studies (GWAS) have identified hundreds of associated variants and thus helped elucidate the genomic architecture of most complex traits. Complex traits can often be grouped into classes that are perceived to be more related based on existing knowledge of trait etiology. Joint genetic or genomic analysis of multiple related traits, also termed as horizontal integration, has helped to uncover 'pleiotropic' loci associated with multiple such related complex traits. Pleiotropy analysis presents several statistical challenges from choice of null and alternative hypotheses, heterogeneity and directionality of effects, measurement resolution (e.g. locus specific or genome-wide) and challenges of scalability to large number of traits. In this context, we shall discuss these analytical challenges in Pleiotropy Analysis and a few statistical methods that have been used to overcome these challenges. We shall also discuss novel statistical methods that can capture exact horizontal pleiotropy between traits at a SNP level and distinguish it from mediated pleiotropy thus helping to sharpen inference of shared biological basis among complex traits.

#06 GENOMICS

Associating Non-coding Enhancer Variants with Altered Transcriptional Regulation in ALS Pathogenesis

Name	Dr. Rahul Kumar
Affiliation	Indian Institute of Technology Hyderabad
Subject Area	Genomics

ABSTRACT

Regulatory regions, such as enhancers, crucially control gene expression, including transcription and post-transcription. While variants within regulatory regions can disrupt gene dynamics and drive disease, their role in Amyotrophic Lateral Sclerosis (ALS) remains largely underexplored. This study aimed to identify enhancer-associated variants in ALS patients and evaluate their impact on disease-relevant gene expression. To achieve this, whole-genome sequencing (WGS) data from the Answer ALS consortium for ALS cases were obtained as gVCF files. Alongside, a curated set of brain and spinal cord associated enhancer regions was obtained from public resources (dbSUPER, seDB2, and GEO). Variants were intersected with these enhancer annotations to identify variants located within regulatory elements. Statistical filtering was applied to identify those with significantly elevated allele frequency in ALS cases (FDR < 0.05). Variants were further classified as annotated or non-annotated based on their presence in population-scale genetic databases such as dbSNP, gnomAD, ExAC, and 1000 Genomes. The associated genes of these variants were further evaluated using RNA-Seq data to assess their expression patterns. The analysis identified 12,742 enhancer-associated

variants enriched in ALS, including 3,833 non-annotated and 8,909 annotated variants. Non-annotated variants represent previously unreported alterations absent from population-scale

resources, suggesting potential disease specificity. Several of these identified variants were associated with genes implicated in brain physiology and neurodegenerative processes, including QKI, ARID5A, ARGLU1, FAF1, LTN1, GAB2, and NAV2, suggesting these regions

may act as regulatory hotspots. In conclusion, the study highlights regulatory enhancer-associated novel genetic variants (NGVs) in ALS. These variants could potentially dysregulate

gene expression by hijacking epigenetic regulation and promoting ALS pathogenesis. Further functional genetic studies implicating these variants will help in a more profound understanding of the molecular complexity of ALS.

#07 METHODOLOGY

Modeling Zero-Inflated Longitudinal Circular Data Using Bayesian Methods: Application to Ophthalmology

Name	Dr. Jayant Jha
Affiliation	Indian Statistical Institute Kolkata
Subject Area	Biostatistical Method

ABSTRACT

This work proposes a model for circular data with excess zeros under a longitudinal framework, where the response is a circular variable and the covariates can be both linear and circular in nature. Motivated by a real case study, a mixed-effects two-stage model based on the projected normal distribution is proposed. The interpretation of the model parameters is discussed and identifiability conditions are derived. A Bayesian methodology based on Gibbs sampling technique is developed for estimating the associated model parameters. Simulation results show that the proposed method outperforms its competitors in various situations. A real dataset on post-operative astigmatism is analyzed to demonstrate the practical implementation of the proposed methodology. The use of the proposed method facilitates effective decision-making for treatment choices and in the follow-up phases.

#08 METHODOLOGY

Healthy aging as information divergence in the multiplex brain

Name	Dr. Moumita Das
Affiliation	Indian Institute of Management Udaipur
Subject Area	Biostatistical Method

ABSTRACT

Understanding how the human brain's structural scaffold and functional traffic co-evolve across the adult lifespan remains a fundamental challenge in neuroscience. While age-related changes in grey matter and functional activation are well-documented, the joint trajectory of the structural (SC) and functional (FC) connectomes is often overlooked due to the lack of an integrative framework. Here, we model the brain as a multiplex network to quantify the information-theoretic interdependencies between these two layers in a cross-sectional cohort of 590 healthy individuals (ages 18–88) from the Cam-CAN dataset.

Using Jensen-Shannon Divergence and relative entropy metrics, we identify a fundamental organizing principle of healthy aging: a progressive information divergence where functional dynamics increasingly "untether" from their underlying structural constraints. Our results reveal that this decoupling follows a significant linear trajectory across the lifespan. Notably, this reorganization is spatially heterogeneous. Meso-scale community analysis using the Multiplex Map Equation identifies subcortical hubs—specifically the putamen, pallidum, caudate, and thalamus—as the primary epicenters of age-related divergence. These "switchboards" transition toward a state of heightened functional independence from structural wiring. Conversely, the limbic core, including the hippocampus and entorhinal cortex, exhibits remarkable stability, suggesting a neurobiological strategy for preserving memory-critical circuits amidst global communicative rewiring.

By characterizing healthy aging as a systematic subcortical decoupling alongside limbic resilience, our work provides a new framework for identifying biomarkers of neural longevity and distinguishes normative maturation from the early signals of neurodegenerative decline.

#09 METHODOLOGY

Optimal Adaptive SMART Designs with Binary Outcomes

Name	Dr. Rik Ghosh
Affiliation	Johnson & Johnson
Subject Area	Biostatistical Method

ABSTRACT

In a sequential multiple-assignment randomized trial (SMART), a sequence of treatments is given to a patient over multiple stages. At each stage, randomization may be used to allocate patients to the different treatment groups considered for the SMART. Despite SMART designs becoming popular among clinicians/clinical researchers, the methodologies for adaptive randomization at different stages of a SMART remain few and not sophisticated enough to handle the complexity of optimal treatment allocation at each stage of the trial. In this work, we have developed an optimal adaptive allocation procedure to minimize the expected number of treatment failures for a SMART that has a binary primary outcome. The applicability of the proposed methodology is also demonstrated using a recently conducted SMART study named M-Bridge for developing a universal and resource-efficient dynamic treatment regime (DTR) for incoming first-year college students.

#10 METHODOLOGY

Estimand and Causal Framework in the presence of a Non-Susceptible Fraction in Cardiovascular Trials

Name	Dr. Swarnendu Chatterjee
Affiliation	Eli Lilly
Subject Area	Biostatistical Method

ABSTRACT

Standard time-to-event estimands in cardiovascular trials assume all patients remain perpetually at risk, yet modern therapies increasingly produce long-term survivors whose event risk becomes negligible. The hazard ratio and restricted mean survival time conflate two distinct treatment effects: increasing non-susceptibility probability and delaying events among susceptible patients. This conflation undermines causal interpretation and clinical relevance. We develop cure-aware estimands within the ICH E9(R1) framework, decomposing treatment effects into cure probability and latency components. Using the mixture cure model, population survival is $S(t) = \pi + (1-\pi)\cdot Su(t)$, where π is the cure fraction and $Su(t)$ survival among susceptibles. We define the cure probability difference, cure odds ratio, and susceptible-specific hazard ratio as distinct estimands targeting separate scientific questions, embedded within a potential outcomes and principal stratification framework for causal interpretation. We show that RMST decomposes as $\pi\tau + (1-\pi)\cdot RMSTu(\tau)$, attributing benefit to its mechanistic sources. Simulations and illustrative trial data demonstrate that estimand choice materially affects efficacy conclusions and that conventional estimands can mask treatment effects when a cure fraction exists. This work provides practical guidance for aligning estimands with clinical objectives in cardiovascular trials with long-term survivors.

#11 RELIABILITY ENGINEERING

Relative Ageing Properties for Distorted Lifetimes with Heterogeneous Structural Functions

Name	Anoop V S
Affiliation	Vellore Institute of Technology, Vellore Campus
Subject Area	Reliability Engineering

ABSTRACT

In reliability theory, relative ageing describes the ageing characteristics of two systems or components in a relative manner. In many practical systems, the lifetime of a system/component does not follow a single homogeneous structure. Real-world factors such as material variability, environmental stress, or design complexity introduce heterogeneity, leading to distorted lifetime distributions. These distortions can significantly affect the ageing behavior and complicate reliability assessment. This work establishes the relative ageing stochastic comparison of random variables through general distorted distribution functions. We derive sufficient conditions for comparing two arbitrary distorted random lifetimes based on their distribution functions that define the distortions. The study focuses on various notions of relative ageing under different reliability measures.

#12 DECISION SCIENCES

Hybrid Nonparametric Approach for Stochastic Regression in Time Series

Name	Kunal Rai
Affiliation	Indian Institute of Management Bangalore
Subject Area	Decision Sciences

ABSTRACT

Understanding the behavior of a time series often requires more than studying its average pattern. In many applications, especially in climate science, finance, and engineering, it is equally important to understand how variability changes over time and how extreme outcomes behave. In this work we study a time series regression framework where our goal is to estimate three important quantities simultaneously: the conditional mean, the conditional variance, and conditional quantiles of a response variable given a set of covariates. These quantities provide a complete description of the conditional distribution and therefore offer deeper insight into the underlying dynamics of the data. We examine a stochastic regression model where the dependent variable is expressed as sums of conditional mean, also known as the drift function and conditional variance, also known as the volatility function, each conditioned on covariates. The error term is assumed to have a mean of zero and a finite variance. The conditional mean reflects the process's central tendency, whereas the conditional variance characterizes the dispersion around the mean. We also concentrate on estimating conditional quantiles, which offer insights into various segments of the conditional distribution, encompassing extreme occurrences. Traditional parametric quantile regression approaches generally use a predetermined functional relationship between the response variable and its predictors. Although these models facilitate straightforward interpretation, they frequently struggle to adequately represent nonlinear associations and intricate dynamics inherent in real life time series data. Nonparametric methodologies provide enhanced flexibility by permitting the data to dictate the form of the regression function. Nevertheless, numerous existing nonparametric quantile regression techniques necessitate the resolution of computationally demanding optimization problems and are frequently formulated under independence assumptions, which may be invalid within time series contexts. To mitigate these constraints, we introduce a methodology that integrates concepts from dynamical systems, machine learning, and nonparametric statistics. The central concept involves integrating data derived from a dynamical relations that characterizes the system's fundamental behavior. We employ machine learning to approximate these dynamical behavior due to computational advantages. These estimates reflect both the observed data and the assumed relationship. In our framework, we do not use these outputs directly as the final estimate. Instead, we use the error of measurement to determine how much weight each observation should receive in the nonparametric estimation procedure. Using this modified nonparametric approach, we construct estimators for the conditional mean and conditional variance. The conditional mean is estimated through a weighted kernel smoother, while the conditional variance is obtained from weighted residuals around the estimated mean. For quantile estimation, we employ a kernel-weighted pinball-type loss function and compute the estimates. The theoretical properties of the proposed estimators are developed with a time series dependence structure based on the functional dependence framework. Under suitable smoothness and bandwidth conditions,
we establish asymptotics for the conditional mean, the variance and the quantiles estimator. Finally, we apply our methodology to precipitation data from four locations in the Indian state of Karnataka: Bangalore, Belgaum, Gadag, and Mangalore.

#13 FINANCE

A Flexible Modeling of Extremes in the Presence of Inliers

Name	Shivshankar Nila
Affiliation	Indian Institute of Technology Tirupati
Subject Area	Finance

ABSTRACT

Many random phenomena, including life-testing and environmental data, show positive values and excess zeros, which pose modeling challenges. In life testing, immediate failures result in zero lifetimes, often due to defects or poor quality, especially in electronics and clinical trials. These failures, called inliers at zero, are difficult to model using standard approaches. The presence and proportion of inliers may influence the accuracy of extreme value analysis, bias parameter estimates, or even lead to severe events or extreme effects, such as drought or crop failure. In such scenarios, a key issue in extreme value analysis is determining a suitable threshold to capture tail behaviour accurately. Bulk model misspecification can affect the threshold, extreme value estimates, and, in particular, the tail proportion. There is no unified framework for defining extreme value mixture models, especially the tail proportion. This paper proposes a flexible model that handles extremes, inliers, and the tail proportion. Parameters are estimated using maximum likelihood estimation. Compared the proposed model estimates with the classical mean excess plot, parameter stability plot, and Pickands plot estimates. Theoretical results are established, and the proposed model outperforms traditional methods in both simulation studies and real data analysis.

#14 METHODOLOGY

A Optimal and robust stepped wedge designs with unequal cluster sizes

Name	Soumadeb Pain
Affiliation	Indian Institute of Technology Kanpur
Subject Area	Methodology

ABSTRACT

Stepped wedge designs (SWDs) are increasingly gaining popularity in cluster randomized trials. A key feature of SWDs is the allocation of clusters to specific sequences that involve transitioning control to the intervention over various time periods. During the design phase, it is often assumed that clusters are of equal size or that the number of clusters assigned to each sequence is balanced. Here, we focus on a cohort stepped wedge trial with unequal cluster sizes. The optimality criterion is a functional of the variance of the estimate of the treatment effect parameter. Our main objective is to identify the optimal design, which refers to the best allocation of clusters to the sequences. We introduce Min–Min (MM) designs, optimal methods for choosing cluster allocations and cluster sizes simultaneously. Optimal designs are function of three unknown correlation parameters. To address this dependency on the design, a Bayesian approach is adopted. Proposed designs are compared with designs proposed in literature. Some real studies are also reevaluated showing better performance of the proposed designs compared to the conventional designs.

#15 OTHERS

Adaptive Learning Architecture Combining Retrieval-Augmented Generation and Bayesian Knowledge Tracing

Name	Nikhil Jha
Affiliation	Ramaiah University of Applied Sciences, Bengaluru
Subject Area	Others

ABSTRACT

Large Language Models (LLMs) are rapidly entering classrooms, yet most educational deployments remain shallow, operating as generic chat interfaces that lack curriculum grounding, cognitive modeling, and institutional scalability. This work presents the ongoing development of a scalable adaptive AI learning platform designed to move beyond static chatbot assistance toward a mastery-aware educational infrastructure.

The system is engineered over a structured academic corpus, transforming unstructured textbooks, lecture materials, and technical documents into a citation-aware, semantically indexed knowledge base using layout-preserving parsing, metadata enrichment, and high-dimensional embeddings. A hybrid retrieval architecture combining a distributed vector database with a relational learner-state store enables grounded responses and persistent personalization.

Student learning is modeled as a dynamic probabilistic process using Bayesian Knowledge Tracing integrated with LLM orchestration. Each interaction updates the learner's mastery profile, enabling detection of knowledge gaps and automatic scaffolding interventions. The architecture further incorporates serverless orchestration, multi-provider routing, semantic caching, and streaming delivery to reduce latency and cost while maintaining academic reliability.

#16 GENOMICS

A unified hierarchical Bayesian approach to transcriptome-wide association study

Name	Arnab Kumar Khan
Affiliation	Indian Statistical Institute Kolkata
Subject Area	Genomics

ABSTRACT

The transcriptome-wide association study (TWAS) has identified novel gene-trait associations, providing essential biological insights. TWAS combines reference transcriptome and genome-wide association study (GWAS) data. Traditional TWAS methods construct a prediction model for gene expression based on the transcriptome data, which is then employed to impute gene expression in the GWAS data. The complex trait in GWAS is regressed on the predicted expression to identify gene-trait associations. Such a two-step approach ignores the uncertainty of the imputed expression and can lead to reduced inference accuracy. We develop a unified Bayesian approach for TWAS that avoids the need for a two-step approach by modeling the two datasets simultaneously. We consider the horseshoe prior to model the relationship between gene expression and local SNPs, and the spike and slab prior to test for an association between the imputed expression and trait to build an integrated Bayesian framework. We extend the method to conducting a multi-ancestry TWAS, focusing on discovering genes that affect the trait in either or both ancestries. Using extensive simulations, we demonstrate that the new approach performs better than existing methods in terms of correctly classifying non-null genes and the accuracy of effect size estimation. To demonstrate our approach, we perform a single and multi-ancestry TWAS for intraocular pressure (IOP), integrating the Geuvadis transcriptome and UK Biobank GWAS data

#17 METHODOLOGY

Stable Epidemiological Risk-Factor Discovery Under Multicollinearity: Elastic- Net Logistic Regression with Bootstrap Stability Diagnostics

Name	Sagnik Acharyya
Affiliation	Visva-Bharati University
Subject Area	Methodology

ABSTRACT

Logistic regression is a foundational tool in epidemiology for modelling binary outcomes and interpreting covariate effects via odds ratios. However, modern biomedical studies often include dozens to thousands of correlated predictors, where classical maximum-likelihood fitting can be numerically unstable, sensitive to perturbations, and prone to separation-driven pathologies. We develop a practical framework for stable risk-factor identification using elastic–net penalized logistic regression, which combines ℓ1 sparsity with ℓ2 grouping to mitigate multicollinearity. To address the inferential gap induced by regularization, we couple the fitted model with a nonparametric bootstrap stability analysis that reports feature inclusion frequencies, sign stability, and bootstrap percentile intervals for odds ratios. A controlled Monte Carlo study with block-correlated predictors demonstrates that, as within-group correlation increases, elastic net retains substantially higher selection stability and signal recovery than the lasso benchmark while maintaining comparable discriminative accuracy. We further illustrate the pipeline on the Breast Cancer Wisconsin (Diagnostic) dataset, where both elastic net and lasso achieve strong test-set discrimination (ROC–AUC around 0.97) and highly reproducible selected sets under bootstrap resampling (mean Jaccard similarity around 0.97), yielding interpretable and stable effect summaries for morphometric risk factors.

#18 FINANCE/TIME SERIES

Functional Data Analysis

Name	Dr. Rituparna Sen
Affiliation	Indian Statistical Institute Bangalore
Subject Area	Finance/Time Series

ABSTRACT

Functional data analysis deals with function valued random variables. We shall start with functional principal components analysis (FPCA) and illustrate the procedure with the example of daily volatility patterns. We extend the FPCA methodology to a time series of functions and illustrate the method with the example of prediction of yield curves. We then discuss the problem of FPCA approach in hypothesis testing and describe an alternative Bayesian procedure for testing Granger causality with an example of yield curves of different countries. We finally develop a method to detect change points in time series of functions and apply this to daily patterns of high frequency financial data.

#19 FINANCE/TIME SERIES

Early Warning Systems for Financial Crises: A Statistical and Econometric Approach

Name	Prof. Prabheesh
Affiliation	Indian Institute of Technology Hyderabad
Subject Area	Finance/Time Series

ABSTRACT

Predicting financial crises remains a formidable challenge for policymakers, particularly as increased financial integration rapidly transmit vulnerabilities across countries. Emerging market economies are especially susceptible to global shocks, particularly liquidity shocks resulting from changes in global monetary policy, which often culminate in boom-and-bust cycles in financial markets and asset prices, leading to financial crises. Given the substantial costs associated with financial crises for both national and international financial systems, the establishment of early warning systems (EWS) capable of identifying future crises has become crucial. This study reviews the statistical and econometric approaches used to predict financial crises and discusses the key methodological steps involved in constructing effective EWS models, including crisis definition, indicator selection, model estimation, and evaluation of predictive performance.

#20 FINANCE/TIME SERIES

Exploring Finite Gaussian Mixtures as Innovations in AR(1) Model with Explanatory Variable

Name	Dr. Anuj Nain
Affiliation	New Delhi Institute of Management
Subject Area	Finance/Time Series

ABSTRACT

Researchers have introduced and studied AR(1) model with explanatory variable to model a variety of time series data. A usual assumption in these models is that the innovations follow a single Gaussian distribution. There might occur situations in which errors in the model may arise from several subpopulations rather than a single density. In this study, a mixture distribution to model the innovations in an AR(1) Model with Explanatory variable has been introduced. A simulation study has been conducted to justify the proposed theory. An empirical analysis on a real-time dataset has also been conducted to provide application, focusing on "Sales" and "Discount" variables representing electronic item sales data with 250 data points.

#21 CLINICAL STUDIES

Efficacy optimized eligibility criteria with improved generalizability of emulated trial results

Name	Ms. Nandini Bhosale
Affiliation	Pfizer
Subject Area	Clinical Studies

ABSTRACT

There has been increasing industry-wide emphasis on designing more inclusive clinical trials to improve treatment access and representation. However, defining optimal eligibility criteria remains challenging. There is a need for systematic, data-driven approaches that identify patient subgroups with meaningful treatment benefit while relaxing overly restrictive enrolment rules to expand the eligible patient pool.

Prior work in oncology (Liu et al., 2021), demonstrates efforts to evaluate clinical trial eligibility criteria using real-world data and AI. Our work extends these methods to the Inflammation & Immunology (I&I) space, specifically COPD and Asthma and builds upon them by incorporating doubly robust effect estimators. Our workflow includes:

- Encode eligibility criteria from existing study protocol to apply on real-world data (RWD).
- Emulate existing trials from real-world data (RWD) under different combinations of eligibility rules (inclusion/exclusion criteria) with propensity score adjusted analysis and augmented inverse probability weighting (AIPW).
- Evaluate each individual eligibility rule in silico with Shapley value and suggesting data-driven inclusion/exclusion criteria.

We further quantify the representativeness using the Generalizability Index of Study Traits (GIST), demonstrating that relaxing overly restrictive eligibility criteria increase representativeness without compromising treatment efficacy. This framework supports systematic, evidence-based optimization of eligibility criteria to enable more inclusive and efficient clinical trials.

#22 CLINICAL STUDIES

From Side Effects to Safer Care: Monitoring Drug Safety - A Statistical Lens on Pharmacovigilance

Name	Dr. Palash Ghosh
Affiliation	Indian Institute of Technology Guwahati
Subject Area	Clinical Studies

ABSTRACT

Pharmacovigilance is the systematic monitoring of medicines after approval to ensure their safety and effectiveness in real-world use. This talk outlines my contribution to drug safety by addressing underreporting, and biased sampling in spontaneous reporting systems and observational studies, and by leveraging supplementary information to improve signal detection and risk estimation. I will present a coherent statistical framework for bias assessment and correction to enhance efficiency. The talk will also summarize the complementary roles of major regulators and coordinating bodies, such as the World Health Organization, the U.S. Food and Drug Administration, and India's Central Drugs Standard Control Organization, in setting standards, operating surveillance programs, and translating evidence into safety actions. Together, these methods and systems strengthen post-marketing surveillance and support safer, data-informed care.

#23 CLINICAL STUDIES

PREDOSE : Pharmacometrically-Refined Early-phase Dose Optimization design for Oncology Study Enhancement

Name	Dr. Damitri Kundu
Affiliation	Eli Lily
Subject Area	Clinical Studies

ABSTRACT

Traditional Phase I oncology trials aim to identify the maximum tolerated dose (MTD), assuming higher doses improve efficacy despite increased toxicity. However, this monotonic dose-response relationship often does not hold for newer anti-cancer agents, prompting a shift toward determining the optimal biological dose (OBD) that better balances efficacy and safety. The US Food and Drug Administration (FDA) now recommends incorporating pharmacokinetic (PK) and pharmacodynamic (PD) data and introducing randomization after dose escalation to support OBD selection. We propose PREDOSE, a Bayesian dose-optimization framework for early-phase oncology trials that integrates patient-level PK and PD data within a seamless two-stage design. In Phase Ia, a PK-informed toxicity model guides dose escalation, followed by a utility-based approach to identify dose schemas with favorable benefit-risk profiles. The framework's effectiveness will be assessed through simulation studies.

#24 CLINICAL STUDIES

Declaring Doses as Safe in Ongoing Oncology Dose-Escalation Trials: Are They Truly Safe? A Critical Assessment of Safety Criteria

Name	Dr. Anirban Mitra
Affiliation	Johnson & Johnson
Subject Area	Clinical Studies

ABSTRACT

Ongoing dose escalation trials pose challenges in assessing safety while aiming to identify the MTD and/or RP2D. Given their long duration, doses are often declared “safe” even as the adaptive, iterative escalation continues. Such declarations guide decisions such as backfilling prior cohorts, intra patient escalation, or starting combination therapy, to enhance efficiency and accelerate development. A common misperception is that once a dose cohort clears escalation, it is safe. But, escalation algorithms allow de escalations implying that a cleared dose may still be unsafe, necessitating quantification of this risk.

We study criteria for declaring doses safe under the BOIN design. Simulations identify conditions that keep patients from being treated above the MTD. Requiring clearance of two consecutive higher dose levels reduces above MTD doses misclassified as safe to <1% (vs 12.4% reported by Zhao et al., 2024, which is overly permissive). A practical rule is to evaluate ≥3 patients at the next higher level which keeps misclassification <5% while allowing, on average, 18% more doses to be deemed safe.

“Safe dose” declaration rules are rarely specified, and their implications are seldom examined, risking treatment at doses ultimately found above the MTD. Our comparative analysis indicates that evaluating at least three patients at the dose above the one to be declared safe is essential for robust safety assessment and for protecting patients from excessive toxicity.

#25 MODELLING/OPTIMIZATION

Exploiting Symmetry in Optimization: Applications of the Purkiss Principle to Optimal Experimental Design

Name	Dr. Satya Prakash Singh
Affiliation	Indian Institute of Technology Kanpur
Subject Area	Modelling/Optimization

ABSTRACT

We examine how inherent symmetries of an objective function can be leveraged to identify and characterize its optimizers. Building on the Purkiss principle, we present a concise set of conditions that force optimal solutions to inherit the problem's symmetric structure, and we show how these conditions dramatically reduce the feasible search space. Through theoretical development and representative examples, we demonstrate the principle's effectiveness in deriving explicit optimal experimental designs for several important classes of problems, often yielding closed-form or uniquely determined solutions. The Purkiss-based approach both simplifies computations and deepens structural understanding, offering a practical, broadly applicable tool for analysts and designers working with symmetric optimization problems.

#26 MODELLING/OPTIMIZATION

Large Wave Direction Data Modeling Using Wrapped Spatial Gaussian Markov Random Fields

Name	Dr. Arnab Hazra
Affiliation	Indian Institute of Technology Kanpur
Subject Area	Modelling/Optimization

ABSTRACT

Statistical modeling of dependent directional data remains relatively underexplored, particularly in high-dimensional spatial settings. Existing approaches for spatial angular data primarily rely on wrapped Gaussian process (WGP) models, which provide a coherent framework for capturing spatial dependence on the circle. However, WGP-based methods become computationally challenging when the spatial domain is large, and observations are available at high resolution. This limitation is especially relevant in the analysis of large-scale geological and climate phenomena, such as tsunamis and hurricanes, where directional measurements (e.g., wave or wind directions) may be available over an entire ocean basin. To address these challenges, we propose a wrapped Gaussian Markov random field (WGMRF) model for large spatial directional datasets. By exploiting the sparse precision structure inherent in Gaussian Markov random fields, the proposed approach achieves substantial computational gains while preserving flexible spatial dependence on the circular scale. We discuss key properties of the model, including its identifiability and dependence characteristics. The model fitting involves standard Markov chain Monte Carlo techniques. Through extensive simulation studies and an application to the wave direction data across the Indian Ocean during the 2004 Indian Ocean Tsunami, we compare the proposed method with both a non-spatial wrapped Gaussian model and a low-rank WGP alternative. The results demonstrate that the WGMRF offers improved predictive performance and scalability in large-domain applications..

#27 MODELLING/OPTIMIZATION

Modelling alternately recurring events: a frailty based hazard estimation approach

Name	Dr. Moumita Chatterjee
Affiliation	Indian Statistical Institute Hyderabad
Subject Area	Modelling/Optimization

ABSTRACT

The motivation is to account for subject specific variations in a Cox proportional hazard model for recurrent and alternating recurrent events. This is done through two sets of frailty components, whose marginal distributions are bound together by a copula function. The likelihood function involves unobservable variables, which requires the use of the EM algorithm. This leads to intractable integrals, which after some approximations, are solved using computationally intensive techniques. The results are applied to a real-life data. A simulation study is also carried out to check for consistency.

#28 MODELLING/OPTIMIZATION

Reliability study of a parallel system equipped with two warm standby components

Name	Dr. Achintya Roy
Affiliation	National Institute of Technology Warangal
Subject Area	Modelling/Optimization

ABSTRACT

This study investigates the reliability characteristics of a parallel system incorporating two warm standby components. The standby units are not activated simultaneously; instead, they are brought into operation sequentially, one after another. We derive the system's reliability function and evaluate three distinct mean residual life (MRL) functions. In addition, several relevant stochastic ordering results are established. Numerical examples are presented to illustrate and support the theoretical findings.