

e-Brochure

Conference for Statistics and Programming in Clinical Research

Organized by

**Indian Association for Statistics in
Clinical Trials (IASCT)**



Contents

Welcome Note from the President	4
From the Desk: Program Chair	6
Our Sponsors	7
Agenda	8
Pre-conference workshop	9
Keynote Speaker	10
Leadership session	12
IASCT Executive Committee	15
Scientific Program Committee	20
Organizing Committee	23
Biostatistics Talks	27
Statistical Programming Talks	39
Data Science Talks	62
Management Talks	76
Industry Posters	81
Students Posters	104
Thank you	119

Our Sponsors

GOLD



Life Sciences



inference



Johnson & Johnson

 **VAICS phastar**

VeraMed®

CLASSIC

SENSAN 
BIOSCIENCES

 **eClinical**
 Powered by elluminate®

 **SCL IT**
Accelerating Pharma with Technology

EPHICACY 

Pre-conference Workshop

Avinash Bandi

*Senior Advisor – Strategic capabilities & Innovation,
Eli Lilly & Company*

I am a Mechanical Engineer by training. With 19 years of experience, I have built deep expertise in clinical statistical programming, specializing in ADaM, SDTM, and TFL creation. My foundation is SAS, complemented by strong proficiency in R — and the evolving intersection of these disciplines with AI is what draws me to this conversation. My focus today is not the tools themselves, but the bigger question they demand: as AI reshapes statistical workflows, how do we ensure that human expertise, judgment, and critical thinking remain at the center — not on the sidelines?



Keynote Speaker

Seema Pai

Senior Director Clinical Site Operations

(India/ SouthEast Asia)

Global Site & Study Operations, Clinical Development & Operations, R&D , Pfizer

Senior Director, Clinical Site Operations (India & South East Asia; APAC Leadership Team member, Pfizer India Country Business Leadership Team member, and Board Member, Pfizer Products India Pvt. Ltd.

Has had leadership experience across India, Africa, SEA and Australia within Pfizer and previously with Daiichi Sankyo. President, Indian Society for Clinical Research (ISCR, 2025-2028), actively advancing clinical research, innovation, capacity building, and policy development in collaboration with industry, academia, and government stakeholders like DoP, ICMR, CDSCO,

Leads global digital innovation initiatives at Pfizer, including site intelligence platforms, decentralized clinical trials, virtual monitoring, MyTrialProfile, and next-generation operating models to enhance clinical trial delivery and patient access.

Physician and clinical research leader with 20+ years of experience across global clinical development. Holds an MBBS, MD (Pathology) from Mumbai University, and Postgraduate diploma in Clinical Research and Regulatory Affairs & Intellectual Property, with extensive experience across oncology, vaccines, neurology, infectious diseases, and rare diseases clinical development.



Leadership session

Vicky Marriott

Head of Statistics and Data Science, Phastar

Vicky Marriott is Head of Statistics and Data Science at Phastar, leading a global team that supports the design, analysis, and delivery of clinical trials. With more than 19 years of experience across pharmaceutical companies and contract research organisations, she has worked throughout the drug development lifecycle, from trial design and regulatory submissions to strategic leadership and quality oversight. Passionate about developing future talent, Vicky is committed to creating opportunities for women in statistics, data science, and clinical research, helping to build diverse and inclusive teams that drive innovation in healthcare.





Talent Technology Innovation



ABOUT US

VAICS is a Life Sciences and Technology company delivering intelligent solutions that accelerate clinical development and enterprise transformation.

We combine domain expertise, advanced technology, and AI-driven innovation to help organizations improve efficiency, enhance compliance, and unlock the value of data.

OUR SERVICES

BIOMETRIC SERVICES

- Clinical Data Management
- Biostatistics & Statistical Programming
- Metadata Services
- Medical Writing

TECHNOLOGY SERVICES

- Application Development
- Artificial Intelligence Services
- Data Engineering



ABOUT SMART

VAICS SMART(Standardized Metadata and Analysis Reporting Tool) is an AI-enabled clinical programming platform designed to accelerate clinical data standards and submission workflows through automation, reusable metadata, standardization, and traceability.

WHAT SMART OFFERS

- SDTM Automation
- ADaM Automation
- TLF Automation
- Global Metadata Hub
- Workbench
- Validation & Audit

WHY VAICS?

- Domain Expertise
- Innovation First
- Regulatory Excellence
- Flexible Engagement Models
- Global Delivery Models

CONTACT US

VAICS CONSULTING PVT. LTD.

WEBSITE: WWW.VAICS-CONSULTING.COM

PHONE: 76800 14414

EMAIL: INFO@VAICS-CONSULTING.COM

INNOVATION SHOULD REDUCE REWORK— NOT ADD ANOTHER BUZZWORD.



HOW SCL REDUCES PROGRAMMING REWORK



THE CHALLENGE

Repetitive programming and manual review activities consume valuable time, increase the risk of inconsistencies, and often lead to avoidable rework.



SCL'S APPROACH

SCL uses metadata-driven workflows, reusable SAS/R frameworks, automated QC checks and R Shiny review solutions to standardise repetitive activities and reduce manual effort.



HOW WE KEEP IT CONTROLLED

- Human reviewed
- Version controlled
- Traceable outputs
- Validated for intended use
- Controlled before production use



MEASURED IMPACT

- 60% reduction in manual review effort
- 100% decrease in rework
- 100% reduction in repeat QC findings



CORE EXPERTISE



Biostatistics



SAS & R



CDM



Submissions

Experienced teams. Controlled delivery.

Bringing Expertise Together. Delivering Impact.



Practical technology.



Controlled impact.



Measurable value.

SCALE WITH SCL IT.

Stable statistical programming teams for FSP and project delivery.

Biostatistics | CDM | SAS & R | Submissions

MORE HEADCOUNT DOES NOT AUTOMATICALLY CREATE MORE CONTROL.

When timelines tighten, CROs and sponsor teams face four recurring risks: repeated onboarding, inconsistent standards, rework, and visibility gaps.

SCL helps you add experienced statisticians and programmers maintaining continuity, governance and delivery visibility.



CONTINUITY

Teams designed to build and retain programme knowledge.



ACCOUNTABILITY

Named leads, defined checkpoints and clear escalation paths.



FLEXIBILITY

FSP, project-based and hybrid delivery aligned with your systems, SOPs and standards.

DELIVERY MODELS



DEDICATED FSP TEAMS

- Stable team working within your systems, processes and governance
- SCL supports end-to-end biostatistics and programming
- Reduces onboarding and rework while preserving knowledge



PROJECT-BASED DELIVERY

- Defined scope, agreed milestones
- Clear ownership of deliverables
- Proven methodology from initiation through delivery
- Transparent communication and status updates



HYBRID SUPPORT

- Specialist resources during submissions, workload peaks, migrations or recovery
- Scalable up or down as needed

**Start with one workstream.
Scale when the model proves itself.**

OUR SERVICES

Biostatistics

Statistical insights from Data to Decision

Clinical Data Management

Quality Controlled Data from Start to Submission

SAS & R Programming

Advanced Statistical insights

AI & Automation

Driving Smart, Efficient Clinical Solutions

SCL PERFORMANCE SCORECARD



100+
Studies Supported



50+
Regulatory Submissions Supported



95%
Milestones Delivered on Time



90%
Annual Employee Retention



8+
Team Members with 8+ Years at SCL

“SCL's teams are responsive, consistent and deeply invested in our success. They understand our standards and operate like an extension of our own team.”

— Director, Clinical Development, Global Biopharma Company



CONTACT US



www.sclicitechnologies.com



info@sclicitechnologies.com

Let's discuss how SCL IT can support your next study or programme.

Case Studies

CASE STUDY 1 SCALING A GLOBAL FSP PROGRAMME

BUSINESS NEED	Scale biostatistics and programming capacity across multiple therapeutic areas while maintaining continuity and quality.
SCL APPROACH	Deployed a dedicated FSP team with governance, knowledge transfer and standardised processes.
OUTCOME	• On-time delivery maintained • 30% improvement in cycle time • High team retention and stakeholder satisfaction

CASE STUDY 2 ACCELERATING SUBMISSION READINESS

BUSINESS NEED	Accelerate submission deliverables with limited internal bandwidth during peak workload.
SCL APPROACH	Provided hybrid support for programming, review and eCTD deliverables with clear ownership.
OUTCOME	• Submission delivered ahead of timeline • Zero critical findings at health authority review • Seamless collaboration with internal teams

IASCT Executive Committee

Nishanth Nalan, President

Director and Practice Head (Life Sciences), ACL Digital

Nishanth Nalan is the Senior Director and Practice Head – Life Sciences at ACL Digital, where he leads the strategic direction of technology-enabled CRO services and solutions. With over 20 years of experience across the Life Sciences and CRO industry, he brings extensive expertise spanning business, technology, clinical research, and organizational leadership.

Nishanth has built CRO organizations and led large-scale clinical accounts across Clinical Data Management, Biostatistics, Statistical Programming, Medical Writing, and technology-enabled CRO solutions. He is passionate about leveraging innovation and technology to deliver high-impact solutions that contribute to better patient outcomes, while empowering teams, driving excellence, and fostering knowledge sharing between industry and academia. He currently serves as President of the Indian Association for Statistics in Clinical Trials (IASCT) and contributes to academia as Visiting Faculty, Advisory Committee Member, and Board of Studies Member at the Department of Biotechnology, Bharathiar University.

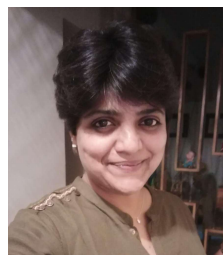
Nishanth holds a Master's in Computer Applications and a Post Graduate Diploma in Human Resource Development from PSG, Coimbatore, along with an Executive Education qualification from IIM Bangalore.



Pooja Shinde, Vice President

Senior Global Biometric Delivery Lead, Parexel

Pooja Shinde is a Senior Global Biometrics Delivery Lead at Parexel with over 14 years of experience in Clinical Research and Business Development. Supporting Real-World Evidence (RWE) studies, she specializes in global biometrics delivery, strategic account management, and client partnerships. She holds a strong track record of leading cross-functional teams, managing complex programs, and delivering successful outcomes across diverse therapeutic areas through collaboration and innovation.



ConSPIC to me - Shaping the future of clinical research through collaboration.

Satya Vyshnavi Thondapu, Secretary

Director – Statistical Programming, Eli Lilly

Satya Vyshnavi Thondapu is Director – Statistical Programming at Eli Lilly and Company, Bangalore, with around 14 years of experience in statistical programming supporting global clinical teams. She brings strong expertise in delivering high-quality statistical outputs, driving quality-focused execution, and enabling efficient, scalable programming practices. A seasoned people leader, she has extensive experience in building and mentoring high-performing teams, shaping strategic and technical solutions, and driving innovation through process improvement and automation across complex, cross-functional clinical environments. She holds a Master's in Pharmacy and has completed the Executive Programme in Leadership and Management from IIM Calcutta. As Secretary of the Indian Association of Statistical Computing and Technology (IASCT), she actively contributes to advancing the professional community through collaboration, knowledge sharing, and continuous learning.

**Jagannatha P S, Treasurer**

Statistics Consultant, ACL Digital

P S Jagannatha is currently Part-time Statistics Consultant, in Eurofins-Optimed Clinical Research. He holds a Master's degree in Statistics from Bangalore University, India. He has 30+ years of experience as a Statistician with Technical Knowledge. Before joining Eurofins, he was with Glaxo Smithkline Pharmaceuticals Ltd for 17 years. He was with the National Tuberculosis Institute, Min. of H&FW, Government of India for a period of 18yrs before joining GSK.

He has co-authored several publications in the reputed journal during his career, was involved in conducting biostatistics workshops for industry colleagues and students and presented at various professional conferences. He is a founder member of IASCT and has been actively involved with all IASCT activities for over 15 years. He started as the event committee chair for 2 years and for 10 years has been Treasurer. He is passionate about IASCT and likes to interact with colleagues across the industry.



Anurag Srivastav,

Associate Director - Biostatistics and Programming - Efficacy

Anurag Srivastav is Associate Director - Biostatistics and Programming at Efficacy Lifescience Analytics Pvt Ltd. He has a Masters in Statistics from Banaras Hindu University, Varanasi and over 17 years of industry experience in Clinical Research industry. He is skilled in people and project management, issue identification and resolution, clinical SAS programming across different therapeutic areas, biostatistics fundamentals, SAS(Base/Advance/Macro), statistical programming and analytics, CDISC standards, recruitment of statistical/R programmers and training management. Anurag has been associated with IASCT for more than a decade and currently serves as Events Committee Head. As working in the core industry, Anurag feels proud to be part of the clinical industry where he can help patients indirectly and would like to help in speeding up the trials to meet the expectation of industry. He loves exploring new food, new culture and meeting new people.

**Dimpy Modak, Admin Committee Head**

Director of Biometrics, ICON Clinical Research

Dimpy Modak is a Director of Biometrics at ICON Clinical Research with over 18 years of experience in statistical programming, clinical data sciences, and biometrics leadership. She leads global teams supporting clinical development programs, with a focus on operational excellence, innovation, and technology-enabled transformation.

She is an active contributor to the biometrics community, having presented at PHUSE and ISCR and serving on the Executive Committee of IASCT. Passionate about mentorship and knowledge sharing, she is committed to developing future-ready talent and advancing the role of data and analytics in clinical research.

ConSPIC to Dimpy?

ConSPIC is where conversations spark possibilities. It's a space to connect with peers, exchange ideas, learn from diverse experiences, and build relationships that inspire both personal and professional growth.



Ashish Charles, Ex-President

*Director Statistical Programming, Clinical Data Operations,
Novartis Healthcare*

Ashish Charles is Senior Director, Statistical Programming within Advanced Quantitative Sciences at Novartis Healthcare Pvt. Ltd. in Hyderabad. With more than 20 years of experience in clinical research, he leads global teams across a broad range of therapeutic and disease areas.

He holds a master's degree in computer applications from Bangalore University and has contributed to IASCT for 15 years, most recently serving as President. He now serves on the IASCT Executive Committee, where he continues to promote learning and knowledge sharing in clinical research. Outside his professional commitments, Ashish enjoys reading, archeology and music.



Harness the Power of Data to Transform Clinical Trials

Delivering Excellence in Clinical Analytics Since 2005

Ephicacy is a pioneering clinical and healthcare analytics company, delivering advanced data analytics services and technology solutions to pharmaceutical and biotech organizations worldwide.

Our Services

Driven by domain expertise and collaboration, we deliver impactful solutions that support every stage of clinical development

-  **BIostatISTICS**
-  **CLINICAL DATA MANAGEMENT**
-  **REAL WORLD EVIDENCE**
-  **PHARMOCOVIGILANCE**
-  **STATISTICAL PROGRAMMING**
-  **CDISC SERVICES**
-  **TECHNOLOGY SOLUTIONS**



20+

Years of Experience

100+

Clients Worldwide

700+

Studies Globally

500+

Experts Worldwide

Get in touch with us

info@ephicacy.com

Hema Krishnamoorthy

Programming Manager, Veramed

Hema Krishnamoorthy is a Programming Manager at Veramed with over 17 years of experience in the pharmaceutical industry. She leads teams of statistical programmers, bringing strong technical expertise and leadership to the delivery of high-quality clinical trial solutions. Throughout her career, she has successfully supported and managed multiple projects across diverse therapeutic areas, fostering collaboration and excellence in execution.

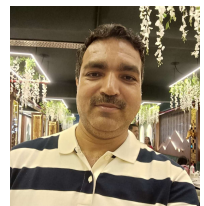
She is passionate about helping people reach their potential, encouraging continuous learning, and supporting teams in turning challenges into opportunities for growth and success. On a personal note, Hema enjoys exploring new places through travel and finds joy in experimenting with different styles of drawing.



Hitendra Pandey

Director of Statistics, Eli Lilly

Hitendra brings over 19 years of experience as a statistician in the pharmaceutical industry and currently serves as a Director of Statistics. He has led statistical delivery across multiple clinical development programs, contributing to regulatory submissions, key clinical trial milestones, and regulatory responses. With a strong focus on cross-functional collaboration, he partners with clinical, regulatory, medical, and other functions to enable data-driven decision-making. His career spans statistical and leadership roles across the pharmaceutical and services industries. Outside of work, Hitendra is a sports enthusiast with a particular passion for cricket and enjoys spending quality time with his family.



Kalpesh Valaki

Manager, Johnson & Johnson

Kalpesh Valaki serves as a Manager at Johnson & Johnson, supporting the Immunology portfolio and delivering high-quality analyses for clinical trials and regulatory submissions. With extensive experience across Johnson & Johnson, Fortrea, IQVIA and GSK, he brings strong expertise in statistical programming and clinical development. Passionate about innovation, collaboration, and continuous learning, Kalpesh actively contributes to the community through knowledge sharing and mentoring emerging professionals.



Manigandan Easwaran

AstraZeneca

Manigandan is a statistician by training, with nearly two decades of experience spanning academia, statistics, and statistical programming. He began his career as a Lecturer in Statistics before moving into the pharmaceutical industry, where he worked as a Trial Statistician and International Project Statistician, with significant involvement in statistical programming. He later transitioned into statistical programming and currently serves as a Statistical Programming Director. His experience spans clinical development and Established Brands, bringing together statistical insight, programming expertise, and cross-functional understanding across clinical and regulatory activities.

A lifelong learner and enthusiastic mentor, Manigandan enjoys both gaining knowledge and sharing it with others. Beyond work, he enjoys exploring mythology, literature, and history.



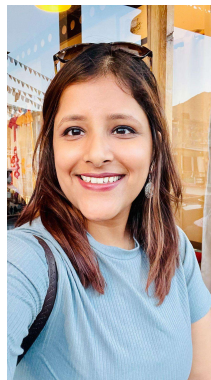
Neha Agarwal

Director, AstraZeneca

Director at AstraZeneca. She holds a Master's degree in Biostatistics and Demography from the International Institute for Population Sciences (IIPS) and a Bachelor's degree in Statistics (Hons.).

With several years of experience in biostatistics, she has served as a Lead Statistician on numerous late-phase clinical, real-world evidence, and bioequivalence studies. Alongside her project responsibilities, she actively contributes to leadership and strategic initiatives across the organization.

Neha is passionate about applying her statistical expertise to generate meaningful evidence, inform better decisions, and ultimately contribute to improving human health.



Shrutimita Pokhariyal

Biostatistician - FSP Manager, Phastar

Shrutimita Pokhariyal is a Biostatistician and FSP Manager at PHASTAR with over 15 years of experience in clinical research and biostatistics. She leads India operations for a global client account, supporting more than 50 statisticians and statistical programmers, with a focus on operational delivery, cross-functional coordination, resourcing, and client engagement. She is also Vice President – Member Value Offering at HBA India and an active mentor, speaker, and contributor to professional communities, including IASCT, ISCR, CWDS, and PHUSE. Passionate about developing the next generation of statisticians, Shrutimita actively supports mentoring, knowledge sharing, and industry–academia collaboration. She believes that strong professional communities can create opportunities, advance statistical science, and enable people to learn and grow together.



Vikramaditya Yandapalli

Senior Manager, Bristol-Myers Squibb

Vikramaditya Yandapalli is a Senior Manager, Statistical Programming at Bristol Myers Squibb (BMS), India, with over 14 years of experience in statistical programming and regulatory submissions. He joined BMS in 2024. He has led programming teams and deliverables across complex clinical trials and regulatory submissions in Cell Therapy, Hematology, Oncology, Virology, and Infectious Diseases.

His expertise includes programming strategy, automation, process standardization, SAS, R Shiny, and AI-enabled programming practices. He is also passionate about mentoring, capability building, and developing high-performing statistical programming teams. Vikramaditya holds an MS from the University of Georgia, USA, and a Biotechnology Engineering degree from JNTU Hyderabad. He has presented at ConSPIC and PHUSE SDE.



Satya Vyshnavi Thondapu, Scientific Committee Chair

Director of Statistical Programming, Eli Lilly

Satya Vyshnavi Thondapu, Scientific Committee Chair for ConSPIC 2026, is the Director of Statistical Programming at Eli Lilly and Company, Bengaluru. A seasoned executive with 15 years of global leadership experience, she specializes in driving strategic automation, process transformation, and scalable programming practices across complex clinical environments.

Known for building high-performing, cross-functional teams, Satya holds a Master's degree in Pharmacy, an Executive Leadership credential from IIM Calcutta, and serves as Secretary of the Indian Association of Statistical Computing and Technology (IASCT).



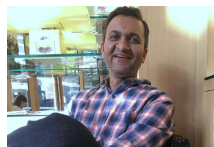
Organizing Committee

Hamel Nayak

Programming Manager, Phastar

Hamel Nayak is a Programming Manager at PHASTAR with over 18 years of experience in clinical research, including more than a decade in statistical programming and leadership. He holds a Master's degree in Computer Science and Industrial Systems. His expertise spans SAS programming, CDISC standards (SDTM and ADaM), regulatory submissions, and clinical trial programming across diverse therapeutic areas. He has led global programming teams with a strong focus on quality, compliance, operational excellence, and continuous improvement.

Passionate about mentoring and collaboration, Hamel brings a strong blend of technical expertise, leadership, and practical insights from global FSP and CRO environments.



Ketan Kapoor

Senior Programmer, Veramed

Ketan is a Senior Programmer at Veramed with a Master's degree in Statistics. With prior experience at Novartis and Syneos Health, he brings a strong foundation in biostatistical programming and a commitment to delivering high-quality work.

A passionate continuous learner, he believes in self-learning, collaboration, and continuously strengthening his knowledge and understanding of statistical programming practices.



Maria Susan Zachariah

Principal Biostatistician, Pfizer Inc.

Maria Susan Zachariah is a Principal Biostatistician at Pfizer. Her expertise spans clinical trial statistics and real-world evidence research, with a particular focus on vaccine development.

In her current role, she supports global clinical studies through statistical design, analysis, interpretation, and reporting of clinical data.

Her commitment to translating data into meaningful clinical insights reflects the purpose of modern clinical research - turning evidence into better outcomes for patients.

In her words: “Behind every data point is a patient waiting for a life-saving miracle.”



Pradeep Acharya

Associate Director, Efficacy

Pradeep Acharya holds a Master's degree in Biostatistics, and currently serves as Associate Director - Statistical Programming at Efficacy Lifescience Analytics Pvt. Ltd..

With over 17 years of experience in the clinical research industry, spanning both pharmaceutical companies and CROs, he brings extensive expertise in biostatistics and clinical SAS programming across therapeutic areas including Oncology, Cardiovascular and Metabolic diseases, Dermatology, and Vaccines.

Pradeep is passionate about presenting statistical research and has presented papers at several leading conferences, including IASCT-ConSPIC, PHUSE US Connect, PharmaSUG US, and PHUSE EU Connect. He has also conducted workshops on Statistics and Clinical SAS Programming at reputed universities and colleges and delivered a lecture on Statistics at the Patan Academy of Health Sciences (PAHS), Nepal.



Rammohan Rayicherla

Associate Director, AstraZeneca

Rammohan is an Associate Director of Statistical Programming at AstraZeneca, Bangalore, with an MSc in Biostatistics from Banaras Hindu University and 20 years of experience across biostatistics and statistical programming. Throughout his career, he has contributed to the successful delivery and submission of numerous early-phase and Phase III clinical trials across diverse therapeutic areas. Beyond his professional journey, Rammohan is an avid traveller, biker, and trekker. Among his notable achievements is completing the Everest Base Camp trek via the challenging Gokyo and Cho La Pass route.



Swathi Mandalapu

Programming Lead, Johnson & Johnson

Swathi Mandalapu holds a Master's degree in Biostatistics and currently serves as a Programming Lead in Statistical Programming at Johnson & Johnson. She supports the Oncology portfolio, delivering high-quality statistical analysis for clinical trials and regulatory submissions.

With extensive experience at Johnson & Johnson, Novartis, and Novo Nordisk, she brings strong expertise in statistical programming, clinical development, and regulatory submissions. She is passionate about innovation, collaboration, and continuous learning, with a strong commitment to advancing excellence in clinical research.



Sreyasi Bhattacharya

Associate Director, Eli Lilly and Company

Sreyasi Bhattacharya is an Associate Director, Statistical Programming at Eli Lilly and Company. She holds a Master's degree in Biotechnology and has experience across diverse therapeutic areas, including Rheumatoid Arthritis, Medical Devices, Hepatology, Hematology, Oncology, Diabetes, and Obesity. Passionate about mentoring and developing high-performing talent, she is committed to driving high-quality, impactful outcomes while advancing process improvement and strategic change through strong leadership.



Vijay Dharmaraj

Lead Statistical Programmer, ACL Digital

Vijay Dharmaraj is a Lead Statistical Programmer at ACL Digital with over 10 years of experience in clinical research and statistical programming. He brings strong expertise across diverse therapeutic areas, with particular experience in ISS/ISE analyses for oncology trials and regulatory submissions, including HAQ and DSUR. Known for his attention to detail, collaborative approach, and commitment to mentoring, Vijay is skilled at translating complex data into clear, actionable insights while maintaining a strong focus on accuracy and quality. Beyond work, he enjoys watching movies, cricket, and tennis, and playing chess.



Nevesh Saravanan

Senior Statistical Programmer, Eli Lilly and Company

Nevesh Saravanan is a Senior Statistical Programmer at SENSAN Biosciences with over 7 years of industry experience and a bachelor's degree in Biomedical Engineering. He brings strong expertise in biostatistical programming and has contributed to clinical development across Oncology, Neurology, and Gene Therapy.

A proactive self-learner, Nevesh is passionate about continuously expanding his knowledge and staying abreast of industry advancements. He also enjoys mentoring junior professionals and supporting their growth through knowledge sharing and guidance. Outside work, Nevesh enjoys exploring new places, listening to music, and spending time with friends.



Samrat Kumar Bollam

Business Development Manager, SCL IT Technologies Pvt Ltd

Samrat Kumar Bollam is a Business Development and Operations professional at SCL IT Technologies, contributing across business development, service quality, process improvement, compliance, HR operations, and stakeholder engagement.

At SCL IT, he works closely across teams to strengthen processes, quality practices, vendor coordination, client engagement, and organizational initiatives, while supporting the company's participation in industry forums and professional communities across the life sciences ecosystem.

With a strong focus on collaboration, continuous improvement, and operational excellence, Samrat is passionate about building meaningful connections, enabling effective processes, and contributing to SCL IT's growth and industry presence.



Ratandeep Pandey

Senior Manager, Bristol Myers Squibb

Ratandeep Pandey is a Senior Manager, Statistical Programming at Bristol Myers Squibb (BMS), bringing more than 15 years of industry experience in statistical programming. With an engineering background, he has built his career around delivering high-quality statistical programming solutions, primarily within the Oncology therapeutic area. Throughout his professional journey, he has successfully led teams and managed programming deliverables for complex clinical trials, contributing to the advancement of innovative therapies and data-driven decision-making.

Beyond his professional responsibilities, Ratandeep enjoys sports and travel, which provide an ideal balance to his work life and serve as his preferred ways to relax, recharge, and stay motivated.



Kannan Murugaiyan

Director Statistical Programming, Inference

Kannan brings over 20 years of experience in the clinical programming industry and currently serves as Director – Statistical Programming at Inference. He was previously associated with Quartesian (now part of Veranex) and Ephicity.

He is passionate about empowering teams through the development of tools and macros that optimize and standardize programming activities, with a consistent focus on delivering high-quality results.

Kannan holds a Master's degree in Computer Science from Bharathiar University.





Biometrics Built Better

Expert biostatistics, programming
and data management services
across Phases I-IV.

Dependable support, tailored to
your needs at any scale:



Project-based, consultancy, or FSP



Prioritization and dedicated focus



Trial design through to regulatory support



Right-first-time delivery



Never been rescued

Meet your global
biometrics partner

veramed.com

BIOS_PPT_001: Faster, Safer dose findings in Phase-I Trial with Time-to-Event Bayesian Optimal Interval (TITE-BOIN) design

Srinivas S

Precision For Medicine

Traditional BOIN design assumes dose limiting toxicities (DLT) occur quickly and it needs complete DLT data to compute toxicity rates. However, this assumption is increasingly violated in modern oncology and immunotherapy trials, where toxicities may emerge late due to immune mediated or metabolic mechanisms. When patient accrual is rapid and the DLT assessment window is long, many patients may have incomplete toxicity outcomes at the time dose-escalation decisions are required (Upcoming patients require immediate treatment), rendering the traditional BOIN design impractical. The Time-to-Event Bayesian Optimal Interval (TITE-BOIN) design extends the BOIN framework to accommodate delayed toxicities and pending DLT data. TITE-BOIN incorporates partial toxicity information by weighting each patient's contribution according to the proportion of the DLT assessment window completed. Patients with longer DLT-free follow-up provide stronger evidence of dose safety, allowing timely and informed dose-escalation decisions. Overall, TITE-BOIN offers a simple, robust, and efficient approach to dose finding, enabling faster trial progression, improved logistical feasibility, and better alignment with the realities of modern early-phase clinical trials. This presentation briefs in detail about TITE-BOIN design followed by a simulation study.

BIOS_PPT_009: Beyond Response Rates: Joint Modelling of Tumour Dynamics and Survival in Oncology Trials

Sakthivel Sivam

Bristol Myers Squibb

In oncology clinical trials, tumour measurements are collected repeatedly over time, while survival outcomes often require longer follow-up. Standard analyses frequently summarize tumour response using endpoints such as objective response rate or progression-free survival, which may not fully capture the longitudinal tumour trajectory or its relationship with survival. This presentation illustrates a practical joint modelling approach that links longitudinal tumour burden with time-to-event outcomes using a simulated oncology trial. A mixed-effects model estimates patient-level tumour trajectories and predicted tumour values are incorporated into a survival model as time-dependent covariates. The approach helps quantify how evolving tumour burden relates to event risk, supports interpretation of treatment effects, and addresses informative dropout. Joint modelling can strengthen oncology decision-making by connecting early tumour changes with long-term clinical outcomes

Keywords: Joint model, tumor dynamics, survival, oncology trials.

BIOS_PPT_010: Reconciling OS–PFS Discordance in Immuno-Oncology: A Multistate Modeling Framework

Neha Pakhle

In immuno-oncology, benefit may not always be captured by single time-to-event endpoints, and an OS advantage can emerge despite minimal progression-free survival (PFS) differences, complicating clinical interpretation and regulatory assessment. Using randomized Phase 2/3 data in first-line melanoma, we applied multistate modeling to simultaneously quantify treatment benefit across time-to-event outcomes and reconcile OS–PFS discordance. Patient status over time was partitioned into progression-free, post-progression, and death states, analyzed by biomarker subgroups. Transition-specific hazard ratios were estimated using Cox proportional-hazards models, and restricted mean time in state using Aalen–Johansen estimators. Distinct drivers of benefit emerged: in the negative biomarker subgroup, treatment reduced progression risk relative to control, whereas in the positive subgroup progression risk was similar but reduced post-progression mortality and extended post-progression survival as compared to control. Multistate modeling offers a transparent framework to reconcile OS–PFS discordance quantifying the exact treatment effect in each state and supporting robust, data-informed regulatory decision-making.

Keywords: Multistate models; Immuno-oncology; Biomarker ; OS–PFS discordance

BIOS_PPT_011: Application of the Partial Order Continual Reassessment Method (POCRM) of Continuous Endpoint for Dose-Combination Treatment Evaluation in a Clinical Trial

Ramya Bhat

GSK Pharmaceuticals

Dose-combinations studies are increasingly used in drug development to evaluate combinations of investigational and established therapies, with the aim of characterizing the dose-combination response relationship and identifying dose-combination suitable for further clinical development. Selecting an optimal dose-combination can be challenging when multiple agents are investigated simultaneously and the combination-response relationship is uncertain.

This abstract outlines the utility of the Partial Order Continual Reassessment Method (POCRM), a Bayesian model-based framework, for dose-combination selection using a continuous endpoint. In a clinical development setting, POCRM is designed to identify the optimal dose-combination when the ordering of dose-combination efficacy is only partially known by incorporating uncertainty in the dose-response relationship and provides a structured framework for comparing candidate dose-combinations. The method evaluates multiple potential complete orderings, estimates posterior probabilities for each ordering, and derives order-averaged response estimates to identify the Minimally Effective Combination (MEC) exceeding a clinically meaningful efficacy threshold

This presentation will describe the rationale, methodology, and practical implementation of POCRM for regimen evaluation in clinical development. Key considerations in defining regimen orderings, model specification, and interpretation of outputs for dose-regimen evaluation will be discussed.

BIOS_PPT_012: Supercharging Data by Implementing Machine Learning to Identify Drivers of Treatment Effect Heterogeneity

Sebin Thomas

Understanding treatment effect heterogeneity is important for interpreting clinical trial results and informing decision-making. Traditional subgroup analyses may be limited in identifying complex, non-linear interactions among baseline characteristics.

This presentation introduces the use of causal forests, a machine learning method for estimating heterogeneous treatment effects in randomized clinical trial data. Causal forests extend the random forest framework by optimizing splits based on treatment effect differences and applying an “honest” estimation strategy to reduce overfitting, enabling estimation of conditional average treatment effects across high-dimensional covariate spaces.

A combined analytical framework integrating regression-based and machine learning approaches will be presented to explore potential drivers of heterogeneity. Variable importance measures and split frequencies are used to identify baseline characteristics associated with differential treatment effects. While several candidate predictors were highlighted by the causal forest analysis, no clinically meaningful characteristics were identified that plausibly explained the observed differences between studies. This work demonstrates the potential of causal forests to explore complex treatment-response patterns while emphasizing the need for careful interpretation alongside clinical and scientific judgment.

BIOS_PPT_015: The Latent Variable Framework: A Method for Power and Precision Gains in Clinical Trials with Composite Endpoints

Ashok Gupta

GSK Pharmaceuticals

Clinical trials frequently use composite endpoints that combine multiple outcomes into a single binary responder/non-responder measure, particularly in chronic diseases. While practical, this approach often discards valuable information from continuous and ordinal components, reducing statistical efficiency and study power. A novel latent variable framework overcomes this limitation by incorporating data from all endpoint components into the analysis. The method treats binary and ordinal outcomes as observed manifestations of underlying continuous latent variables and assumes a correlated multivariate normal distribution across component errors. By jointly modelling observed continuous outcomes and latent variables, it preserves information that is typically lost through dichotomization. This enables more accurate estimation of treatment effects on the composite endpoint and substantially improves statistical power. Studies have demonstrated gains large enough to reduce required sample sizes by up to 33%. Consequently, this methodology supports more efficient, cost-effective, and informative clinical trials without compromising scientific rigor or interpretability.

BIOS_PPT_016: Federated Learning in Rare Disease Research: Opportunities and Challenges from a Biostatistical Perspective

Venkata Harish Challa

PAREXEL Pvt. Ltd

Federated Learning (FL) offers a promising approach for rare disease research, where small patient populations, fragmented datasets, and privacy regulations limit traditional collaborative analyses. FL enables institutions to retain patient-level data locally while sharing model parameters, supporting privacy-preserving multi-site research. Key biostatistical methods include mixed-effects and Bayesian hierarchical models to address site-level heterogeneity, distributed survival models for decentralized time-to-event analyses, and federated meta-analysis for collaborative estimation. Evidence suggests FL improves statistical efficiency, model robustness, and external validity compared with single-site analyses. However, challenges remain, including non-identically distributed data, variability in endpoint definitions, differing missing-data mechanisms, and reduced precision from summary-statistic aggregation. Regulatory acceptance of FL-based evidence is evolving and may require transparent documentation, data quality controls, reproducibility assessments, and sensitivity analyses demonstrating validity comparable to centralized approaches. When supported by standardized data processes, secure infrastructure, and robust governance, FL can facilitate reliable, reproducible rare disease research while preserving patient privacy.

BIOS_PPT_018: A Bayesian Gaussian Copula Framework for Evaluating Longitudinal Composite Clinical Endpoints

Amila Poddar

GSK Pharmaceuticals

Gaussian copula models provide a flexible statistical framework for jointly analysing outcomes that differ in scale, distribution, or clinical meaning. By separating the modelling of marginal distributions from the modelling of dependence, Gaussian copulas are particularly valuable in scenarios where multiple longitudinal endpoints—continuous, binary, or otherwise—must be evaluated together while preserving their inherent correlation structure. This approach is increasingly used in modern clinical development when interest lies in complex, composite outcomes that cannot be adequately captured by conventional univariate methods.

In the present work, a Gaussian copula-based Bayesian joint modelling strategy was implemented to estimate clinical benefit in a trial evaluating a drug XYZ. The composite endpoint was defined using two weekly longitudinal outcomes: (1) $\geq 20\%$ reduction from baseline in the Abdominal Pain Intensity (API) score, and (2) a Global Symptom (GIS) rating of “significantly relieved” or “moderately relieved.” A responder was defined as achieving both components for at least 5 of 10 weeks. Weekly API was modelled as a continuous normally distributed variable, while GIS was dichotomized and modelled

through a Bernoulli distribution. The Gaussian copula was used to capture temporal correlation across weeks as well as cross-outcome dependence. The model incorporated treatment, week, baseline factors, and treatment-by-week interactions, and posterior inference was obtained through MCMC with extensive posterior predictive simulation to estimate probabilities of composite response and risk differences relative to placebo.

Compared with traditional responder analyses, which typically dichotomize outcomes and ignore the joint evolution of endpoints over time, the Gaussian copula framework offers several important advantages. It preserves the full information in continuous measures like API, appropriately models correlations between API and GIS, naturally handles missing data under MAR, and more accurately propagates uncertainty through posterior predictive sampling. As a result, the approach enhances statistical efficiency and produces richer and more clinically interpretable estimates of treatment benefit. This joint modelling strategy therefore represents a more powerful and principled alternative to conventional responder-based analyses for evaluating complex, longitudinal composite endpoints.

BIOS_PPT_024: Code Less, Analyze More & Deliver Quickly: Building Modular R Framework for Scalable Real-World Evidence Generation & Analytics

Suraj Yadav, Arun Chhikara and Suraj Sharma

Veramed

Over the past few years, R has emerged as a powerful platform for real-world evidence (RWE) analytics, combining reproducibility, extensibility, and scalability within a single open-source ecosystem. However, common RWE workflows are often repeatedly programmed across studies, resulting in inefficiencies and variability in implementation. To address these challenges, a modular R package is developed which provides a standardized and reusable approach to RWE studies, enabling rapid implementation, consistent cohort evaluation, scalability to big data and efficient generation of key insights and outputs. By reducing repetitive programming and promoting a modular, plug-and-play design, it supports scalable execution of analyses ranging from quick exploratory requests to complex observational studies. Designed to operate within existing R environments with minimal dependencies, the module leverages R's strengths in functional programming and reusability to support consistent analytics across rapid data queries and complex observational studies. Through real-world use cases this framework represents a practical advancement in automation, open-source adoption, and operational efficiency for modern RWE programming teams.

Keywords: Real-World Evidence, R Programming, Modular Framework, Open-Source Analytics

BIOS_PPT_026: Predicting Individual Weight-Loss Trajectories Over Time: A Machine Learning Approach Using LASSO-Guided CART Modeling

Anisha Das

Eli Lilly

Background/Objective: Predicting individual weight-loss trajectories can support personalized treatment planning in obesity pharmacotherapy. This work develops a generalizable machine learning framework for predicting body weight over an extended treatment horizon from baseline characteristics, validated using data from a Phase 3 obesity clinical trial.

Methods: LASSO regression identified the most predictive baseline features across demographic, anthropometric, and metabolic domains, to train a CART regression model. Performance was assessed via a subject-level train-test split and confirmed through k-fold cross-validation, with no subject overlap between folds.

Results: LASSO selected a parsimonious subset of candidate predictors. CART showed strong near-term accuracy that attenuated gradually over longer horizons, reflecting increasing individual variability over time. Predicted trajectories closely tracked observed weight loss, with clear separation across treatment groups.

Conclusion: Interpretable ML models reliably forecast weight-loss trajectories from baseline data, offering a scalable, trial-agnostic tool for personalized trajectory prediction across obesity pharmacotherapy trials.

BIOS_PPT_028: Beyond Single Timepoints: A Case Study in AUC-Driven Analysis of Sustained Treatment Response in Chronic Care

Yogita Thoke

PAREXEL Pvt. Ltd

Background: Evaluating long-term efficacy requires approaches that capture sustained treatment benefit beyond single time-point endpoints. Response durability is important in psoriasis (PsO).

Methods: This phase 2 integrated PsO case study applied an Area Under the Curve (AUC)-based sustained response analysis using an as-observed framework. Missing data were handled conservatively using Last Observation Carried Forward (LOCF) for the first missing visit and nonresponse imputation for subsequent consecutive missing visits. Results are presented descriptively.

Results: AUC-based analysis illustrated sustained participant responses over time and captured clinically meaningful cumulative benefits beyond traditional single time-point endpoints. The pragmatic missing data strategy ensured conservative and interpretable estimates.

Conclusion: These findings demonstrate that AUC-based sustained response analysis captures the durability and cumulative impact of treatment over time, including benefits that may not be fully reflected by traditional single time-point endpoints. This approach offers a

clinically relevant framework for assessing long-term treatment benefit in patients with psoriasis.

BIOS_PPT_032: Accelerating Dose Escalation through Partial Follow-Up Information: Comparison of TITE-BLRM and Traditional BLRM

Sandipan Ghosh

Inference

Phase I oncology trials aim to identify the maximum tolerated dose (MTD) while minimizing the risk of exposure to unsafe dose levels. Conventional Bayesian Logistic Regression Models (BLRM) require complete dose-limiting toxicity (DLT) follow-up before dose escalation, potentially delaying trials. Time-to-Event BLRM (TITE-BLRM) addresses this by incorporating partial follow-up data for earlier Bayesian model updates.

This presentation will introduce the statistical principles underlying BLRM and TITE-BLRM, highlighting how time-to-event weighting enables the use of incomplete toxicity data in adaptive dose-escalation decisions. Through simulated Phase I oncology trial scenarios, the presentation will compare the two approaches with respect to MTD selection, overdose control, operational efficiency, and overall trial duration. Practical insights into TITE-BLRM implementation, covering Bayesian updating, Escalation With Overdose Control (EWOC) and critical model assumptions will also be presented.

BIOS_PPT_035: Project Optimus: Transforming Oncology Dose Selection Through Modern Statistical Designs

Sreejita Das

Inference

Project Optimus, an initiative led by the U.S. Food and Drug Administration (FDA), is redefining the approach to dose selection in oncology drug development. Moving away from the historical reliance on the Maximum Tolerated Dose (MTD), the initiative encourages a comprehensive evaluation of dose-response relationships to identify doses that maximize patient benefit while minimizing unnecessary toxicity.

This presentation will explore the regulatory drivers behind Project Optimus and its implications for clinical trial design and drug development strategies. The discussion will focus on the role of modern statistical methodologies, including model-assisted and Bayesian dose-finding approaches, in supporting dose optimization. Additionally, the application of adaptive designs and exposure-response analyses to integrate safety, efficacy, pharmacokinetic, and pharmacodynamic data will be highlighted. Through practical examples, the presentation will demonstrate how innovative statistical approaches can facilitate robust dose selection, align with evolving regulatory expectations, and support more efficient development of oncology therapies.

BIOS_PPT_038: Finding the Right Prior: Systematic Calibration in Bayesian Dose-Escalation Studies

Sudipta Banerjee

Merck

In Phase I dose-escalation trials, model-based Bayesian methods are increasingly used in identifying the maximum tolerated dose (MTD). The choice of priors for model parameters

plays a major role in the dose recommendations. While vague priors, meta-analytic predictive priors are commonly used approaches in these trials, we present an alternative approach that selects priors based on their performance across realistic trial scenarios. This presentation attempts to provide an overview of a systematic grid search prior calibration framework that leverages clinically relevant information including starting dose, tested doses, expected MTD; and are optimized through comprehensive simulations. The framework evaluates prior performance across a wide range of realistic scenarios, encompassing varying dose-limiting toxicity (DLT) patterns and participant allocations per dose level. The presentation includes an example study to demonstrate the grid search calibration approach for rigorous and transparent prior identification. Additionally, we explore the integration of large language models (LLMs) as tools to enhance and streamline the calibration framework, potentially making scenario generation and prior assessment more efficient.

BIOS_PPT_039: Stabilizing Multi-way Subgroup Analyses with Limited Sample Sizes

Gokul Pradeep

Pfizer

Multiway subgroup analyses are widely used to explore treatment effect heterogeneity but are often limited by sparse data, zero-count subgroups, and imbalanced sample sizes, leading to unstable estimates. Bayesian hierarchical models provide a principled framework to address these challenges through partial pooling. We applied hierarchical regression models to a rheumatoid arthritis clinical dataset to evaluate subgroup-level efficacy and safety across sex, age, body mass index, and disease duration. Efficacy was assessed using ACR70 response at Month 3, and safety using major adverse cardiovascular event incidence rates via logistic and Poisson regression. The hierarchical framework moderated extreme estimates in sparse subgroups while preserving signals when sample size was sufficient. These results demonstrate improved stability, interpretability, and clinical relevance for benefit-risk assessment in high-dimensional subgroup analyses with limited data. Additional methodologies may be incorporated as the project scope evolves.

BIOS_PPT_040: Evaluating the Causal Role of Early Biomarker Response in Mediating Long-Term Clinical Outcomes in Autoimmune Diseases

Shambhavi Pandey

Pfizer

Early biomarker response is frequently used to monitor treatment effectiveness in autoimmune diseases; however, its contribution to long-term clinical outcomes remains uncertain. This study evaluates whether early improvement in a candidate biomarker mediates the treatment effect on long-term disease progression, using a causal inference framework. Longitudinal biomarker measurements and clinical outcome data from a simulated randomized clinical trial, calibrated to reflect realistic effect sizes and variability reported in published autoimmune disease trials, were analyzed using mixed-effects models and counterfactual mediation analysis. Bootstrap resampling was applied to estimate confidence intervals for direct and indirect treatment effects, and sensitivity analyses assessed the robustness of the proportion-mediated estimate, particularly under conditions of small total treatment effect. The proposed framework quantifies the proportion of

treatment benefit explained through early biomarker response, providing a statistically rigorous method for biomarker evaluation. The analysis is implemented in R. This approach supports evidence-based endpoint selection and demonstrates a reproducible workflow applicable to autoimmune disease clinical trials and biomarker qualification studies.

BIOS_PPT_042: STATOMATE: One Click, Consistent Outputs

Vaishali Govindarajan and Nayana Tharayil Jayakrishnan

Zifo RnD Solutions

In TLF's maintaining consistency throughout all outputs is critical. When we are manually checking all outputs, missed details or mistakes can still occur. Also, it is time-consuming and requires multiple review cycles. It is challenging, especially during large-volume TLF submission. What if we have an automation tool to do all this? STATOMATE is a tool that we developed using R Shiny to make things better and faster.

STATOMATE will validate whether consistency and layout standards are met across TLFs and maintain consistency throughout headers, titles, abbreviations, references, and subscripts, and detect blank pages. Apart from this, it will check for missing abbreviations, outdated terms, and detect inconsistencies in font cases across the outputs. By using this tool, the time required for layout check was reduced by 40% and allowing statisticians to focus more on result accuracy. More details about the tool will be explained in the presentation.

BIOS_PPT_050: Causal decomposition of the MRD - Survival pathway in Multiple Myeloma

Nagadharshini M

Pfizer

Minimal residual disease (MRD) negativity is strongly associated with improved progression-free (PFS) and overall survival (OS) in multiple myeloma, but its causal role remains unclear. This presentation quantifies the extent to which treatment-related survival benefits are mediated through MRD negativity using a formal causal framework.

Using randomized trial data, a Directed Acyclic Graph (DAG) will model pathways from treatment to OS, adjusting for baseline confounders. Natural direct and indirect effects will be estimated on the restricted mean survival time (RMST) scale via pseudo-value based mediation analysis and g-computation. Sensitivity analyses, including E-values, will assess robustness to unmeasured confounding.

Results will provide causally interpretable estimates of mediation, clarifying whether MRD negativity acts as a true intermediate endpoint or a prognostic marker, informing its validity as a surrogate endpoint in clinical trials and regulatory decision-making.

BIOS_PPT_052: The PMDA TSV File: Small File, Big Submission Impact

Jasmin Jobanputra

Novo Nordisk

Electronic submissions to Japan's Pharmaceuticals and Medical Devices Agency (PMDA) require a Tab Separated Value (TSV) file that serves as the backbone of submission metadata. Despite its critical role, TSV preparation is often underestimated, with manual metadata entry restricted to the limited Gateway submission window before NDA filing. This presentation outlines a practical approach to understanding and preparing TSV files by explaining their structure, study- and file-level specifications, and key metadata requirements. It demonstrates how existing submission package metadata can be leveraged to automate much of the TSV generation process while identifying fields that require manual completion. The session also discusses common pitfalls that lead to validation failures, regulatory queries, submission delays, and increased operational effort. By adopting standardized and partially automated TSV preparation workflows, organizations can improve submission readiness, enhance data consistency, reduce manual effort, and strengthen compliance with PMDA submission requirements. Attendees will gain practical insights and best practices to support efficient, high-quality regulatory submissions.

BIOS_PPT_057: Estimation of Inter-Individual Variability in PopPK Models Under Sparse Sampling Scenarios: A Simulation-Based Evaluation

Anil Kumar Ganna and Dheeraj Gopu

Bristol Myers Squibb

Sparse pharmacokinetic (PK) sampling is widely used in clinical studies, but fewer observations may compromise reliable estimation of inter-individual variability. This study evaluated the impact of reduced sampling density on variability estimation in population pharmacokinetic (PopPK) models and compared the performance of First-Order Conditional Estimation with Interaction (FOCEI) and Stochastic Approximation Expectation Maximization (SAEM) methods.

True parameter values from a PopPK model developed using the Theophylline dataset were used to generate simulation datasets. For each sampling scenario (11, 6, 5, 4, 3, 2, or 1 samples per subject), 100 simulation–estimation replicates were performed. Performance was assessed using relative bias (RBias), relative root mean squared error (RRMSE).

Variability in volume of distribution was most sensitive to information loss, with RRMSE increasing from 10.2% to 94.5% for FOCEI and from 14.4% to 67.9% for SAEM. In contrast, variability in clearance remained comparatively robust. FOCEI generally performed better with richer data, whereas SAEM showed greater robustness under sparse sampling conditions.

Keywords: Sparse, variability, PopPK, FOCEI, SAEM

BIOS_PPT_060: Beyond Dichotomization: Win Probability Methods for Likert-based Ordinal Patient-Reported Outcomes in Clinical Trials

Snehal Dhamne and Sivam Sakthivel

Bristol Myers Squibb

Patient-reported outcomes (PROs) measured on Likert scales are increasingly used as primary endpoints in randomized clinical trials (RCTs). However, standard approaches — including dichotomization and linear regression treating ordinal data as continuous — are statistically inappropriate due to unequal gaps between scale categories, leading to loss of information and reduced statistical power.

We review and demonstrate win probability (WP) as a robust estimand for Likert-based outcomes. Win probability is defined as: $c = P(X > Y) + 0.5 \times P(X = Y)$, where X and Y represent outcomes of randomly selected patients from treatment and control groups respectively. Three companion measures — win probability, win difference, and win ratio — are presented alongside covariate-adjusted extensions and 95% confidence interval estimation using logistic transformation.

Applied to the MyTEMP dialysis trial, WP = 0.64 (95% CI: 0.57–0.70), demonstrating superior statistical power and interpretability compared to conventional methods.

Win probability provides a clinically meaningful, statistically rigorous, and assumption-free estimand for ordinal PRO endpoints in RCTs.

BIOS_PPT_064: Beyond Maximum Likelihood: A Minimum Distance Approach for High-Dimensional Cox Models

Mouli N

GSK Pharmaceuticals

The Cox proportional hazards model is widely used in pharmaceutical research for analysing time-to-event outcomes, including patient survival, disease progression, and treatment response. With advances in genomic technologies, clinical studies now generate high-dimensional data where the number of genomic features greatly exceeds the number of patient samples ($p \gg n$). Traditional penalised Cox methods, such as LASSO, SCAD, and MCP, are commonly used for variable selection but remain sensitive to outliers, measurement errors, and model misspecification. This study proposes a Minimum Distance Approach for high-dimensional Cox models by replacing likelihood-based estimation with a robust distance-based criterion. The proposed High-Dimensional Minimum Distance Cox Estimator estimates regression coefficients by minimising the distance between the observed Kaplan–Meier survival function and the model-based survival function. The proposed framework provides a robust and reliable approach for identifying genomic biomarkers and improving survival prediction in precision medicine applications.

BIOS_PPT_065: Bayesian Dynamic Borrowing from External Controls to Improve Efficiency in Rare Disease Clinical Trials

Somnath Karan and Shreya Singh

Inference

Concurrent, adequately powered control arms are often infeasible in rare disease trials due to small patient populations. This talk presents Bayesian dynamic borrowing as a solution, framed against a real regulatory precedent in which a placebo-controlled rare disease trial augmented its control arm with historical placebo data through an FDA innovative-trial-design engagement. Using Robust Mixture Prior and Commensurate Prior frameworks - where the borrowing weight adapts to the observed similarity between historical and current control populations - we present a simulation study spanning a grid of drift, sample size, and effect-size scenarios, quantifying gains in effective sample size and statistical power alongside preserved Type I error control under heterogeneity. We close with practical guidance for pre-specifying the borrowing model and sensitivity analyses in the Statistical Analysis Plan, positioning Bayesian dynamic borrowing as a principled, transparent, and regulatory-defensible pathway for rare disease clinical development.

BIOS_PPT_071: 54 Days to Approval: Statistical Lessons from Novo Nordisk's First FDA National Priority Voucher Submission

Hanna Angelene K M

Novo Nordisk

The FDA Commissioner's National Priority Voucher (CNPV) programme represents a new paradigm in regulatory review, compressing timelines from months to weeks. Novo Nordisk's approval of semaglutide 7.2 mg (Wegovy®) through this pathway provided valuable insights into how statistical support, submission readiness, and cross-functional collaboration must adapt to meet the demands of accelerated review.

Unlike traditional regulatory pathways, the CNPV programme required responses to agency information requests within 24–48 hours, allowing minimal time for reactive analyses or last-minute planning. Drawing on experiences from the semaglutide 7.2 mg submission, this presentation highlights practical lessons in submission readiness, management of rapid information requests, and maintaining scientific and statistical rigour while delivering at speed.

The session will explore how accelerated review pathways are reshaping expectations for statistical readiness and regulatory interactions. Attendees will gain practical insights into preparing and leading teams, analyses, and decision frameworks to support rapid decision-making without compromising scientific quality.

BIOS_PPT_075: Calibrating Bayesian Success Criteria to Frequentist Operating Characteristics: Operationalizing the Draft FDA Guidance in Sections IV - A.1 and B.1 for Phase II studies with two binary endpoints

Solomon Christopher, Manusree Banerjee and Sankhadeep Kundu

Bristol Myers Squibb

Background: The recent FDA draft guidance on use of Bayesian methodology in clinical trials of drugs and biological products recommends calibrating the Bayesian success criteria to frequentist operating characteristics (OCs) when information is not borrowed. For hypothesis testing, the recommended OCs for calibration are type-1 error rate and power.

Methods: We illustrate a simulation-based calibration procedure for a phase-II setting with two binary endpoints, wherein correlation between the endpoints is accounted for. The model specification and methodology follows published Bayesian optimal phase-II (BOP2) design. In addition to BOP2 specifications, further details on controlling type-1 error rate and general methodology on simulation-based calibration procedure will be described.

Results/Conclusion: The calibration results will be reported for a hypothetical trial data. The simulation-based calibration framework is useful for designing studies that utilize Bayesian success criteria and therefore, facilitates regulatory alignment and engagement as per the recent FDA draft guidance.

BIOS_PPT_078: Quantifying Uncertainty in Treatment Rankings Generated Through Network Meta-Analysis-Advancing Beyond SUCRA with MCMC-Based Credible Intervals

Amritendu Bhattacharya

Eli Lilly

Treatment rankings from Bayesian network meta-analysis typically rely on SUCRA (or its frequentist counterpart, P-Score), yet these metrics poorly capture ranking uncertainty, risking overconfident, clinically misleading conclusions. Precision of Treatment Hierarchy (POTH) improved on this but still cannot characterize credible interval ranges around orderings or reflect varying decision-maker priorities. This work offers two advances. First, an MCMC-based framework samples approximate posterior distributions of treatment efficacy parameters to generate rank-specific credible intervals and highest posterior density (HPD) regions capturing the densest probability mass of rankings and offering more reliable uncertainty estimates than SUCRA point values, especially for skewed or multimodal distributions. Second, it extends POTH into personalized, multidimensional variants that weight efficacy and safety endpoints according to context-specific priorities, yielding tailored certainty metrics. Together, these methods reveal uncertainty masked by traditional approaches, giving decision-makers a clearer view of ranking confidence, particularly with few competing treatments and distinguishing robust orderings from those likely to shift depending on outcome measures or patient-specific priorities.

BIOS_PPT_080: Matching the Unmatched: Data-Driven Synthetic Control Arms via Mahalanobis Distance

Jothi Meenakshi S P and Chandrasekhar Chikatapu

Sanofi Healthcare

Synthetic Control Arms (SCAs) are emerging as a valuable alternative to conventional randomized control groups, particularly in rare diseases, oncology, and settings where placebo-controlled studies are operationally challenging or ethically constrained. This work demonstrates the application of Mahalanobis Distance Matching (MDM) for constructing SCAs using external real-world or historical patient data. MDM enables multivariable matching by accounting for the covariance structure among baseline characteristics, thereby improving comparability between treated and control populations. A representative oncology case study was used to match patients on key prognostic variables, including age, ECOG performance status, tumor burden, and prior therapies. Distance-based matching with caliper restrictions facilitated the identification of clinically comparable controls while excluding outliers. The resulting synthetic cohort supported meaningful comparative evaluation of treatment outcomes and illustrated the potential of SCAs to enhance evidence generation when randomized controls are unavailable. These findings highlight the utility of Mahalanobis Distance Matching as a transparent and robust framework for incorporating real-world evidence into modern clinical trial design and decision-making.

BIOS_PPT_081: Adaptive Sample size Re-estimation- Promising Zone Design Approach

Yogesh Pokharkar

Bristol Myers Squibb

Sample size in clinical trials is typically fixed using power calculations, but uncertainty in the assumed treatment effect can undermine this approach. Adaptive sample size re-estimation addresses this challenge through a staged investment strategy: trials begin with a smaller upfront sample size, with additional resources committed only if interim results are promising.

This case study presents the design evolution of a Phase III, open-label, multicentre randomized controlled device trial using sample size re-estimation via a Promising Zone Design. Compared to traditional designs, it delivered meaningful reductions in average sample size, trial duration, and upfront investment, while preserving the statistical rigor required for regulatory approval. It also lowers the risk of an underpowered study by allowing sample size to be reviewed and adjusted using accumulating data, and increases the probability of trial success by committing extra resources only when interim results are promising.



Advance Clinical Research. Advance Your Career.

Join Phastar's global experts in biostatistics, programming, and data science, advancing clinical development through statistical excellence.

Quality Data

Confident Decisions

Faster Approvals

phastar.com/careers | phastar.com



Ind_SP_PPT_003: Enhancing Pharmaceutical Code Generation: An LLM-Powered Bot for Submission-Ready SAS, R, and Python Analyses in Pivotal Studies

Saritha Raghava

AstraZeneca

Large language models (LLMs) are emerging as valuable tools for improving programming efficiency in pivotal clinical studies. We present an LLM-powered bot designed to generate submission-ready SAS, R, and Python code for pharmaceutical analyses, including data derivations, statistical outputs, and tables, listings, and figures. The solution integrates domain-specific prompting, standardized templates, and quality guardrails to support consistency, traceability, and reproducibility across programming workflows. In regulated environments, the bot is intended to augment programmer productivity by reducing manual effort, accelerating turnaround, and promoting alignment with study and submission standards. Human review and validation remain central to ensure quality, compliance, and fitness for purpose. This work demonstrates the potential of LLM-enabled code generation to streamline analysis development, improve cross-language standardization, and support high-quality regulatory submissions in pharmaceutical research.

Ind_SP_PPT_004: Making CGM Data Analysis-Ready: Approaches for Large-Scale Data Preparation

Bhumireddy Venkata Pavan Kumar Reddy

Novo Nordisk

Clinical trials increasingly generate large and complex datasets, including omics and high-frequency device data. Continuous Glucose Monitoring (CGM) is one such data source, widely used in diabetes research to provide real-time glucose measurements through a subcutaneous sensor that tracks glucose levels in interstitial fluid. Unlike traditional fingerstick methods, CGM captures continuous longitudinal data, enabling assessment of glycemic variability, time in range, hypoglycemia, hyperglycemia, and treatment response. However, the high volume and frequency of CGM data create challenges in storage, system performance, standardization, quality control, analysis, and interpretation. Missing values, noisy readings, inconsistent formats, and long processing times can further complicate clinical trial workflows.

This presentation highlights the clinical relevance of CGM data, key operational and analytical complexities, and scalable solutions such as validation checks, and optimized programming workflows to convert high-volume CGM data into meaningful clinical insights.

Missing values, noisy readings, inconsistent formats, and long processing times can further complicate clinical trial workflows.

Ind_SP_PPT_007: Linear Mixed-Effects Analysis of Longitudinal HGB Data from CDISC ADaM Datasets

Moheet Jain, Arati Todkar

ICON Clinical Research India Pvt. Ltd.

Longitudinal biomarker analysis plays a critical role in evaluating treatment efficacy and disease progression in clinical trials, where repeated measurements are collected from the same participants over time. This study demonstrates the application of Linear Mixed Effects Models (LMMs) using R Programming to analyze longitudinal HGB data derived from CDISC ADaM datasets. Subject level (ADSL) and laboratory (ADLB) datasets were integrated to evaluate changes in HGB measurements across scheduled study visits. The model integrates subject-level and laboratory data to evaluate HGB changes across the scheduled visits, incorporating fixed effects for treatment visit and treatment by visit interaction, along with a subject-level random intercept to account for within-patient correlation. Model estimate uses Restricted maximum likelihood, and inference is conducted using the lmerTest package (Satterthwaite approximation) to obtain valid p-values for fixed effects. The analysis identified HGB trajectories, quantified inter-patient variability, and evaluated whether treatment produced statistically significant changes compared with the control group. This approach demonstrates the value of mixed effects modeling for repeated measures clinical data, providing a reproducible and robust framework for longitudinal analysis. The findings support evidence-based interpretation of biomarker trends including visualization and informed decision-making in clinical research

Ind_SP_PPT_008: Implementing the Estimand Framework: Bridging Statistical Theory and Programming Practice

Ketan Kapoor, Hardik N Shah

Veramed

The ICH E9(R1) estimand framework has changed how clinical trial objectives are translated into statistical analyses, bringing statisticians and programmers closer together throughout study development. While the concept is well established, applying estimands consistently during programming can still be challenging.

In this ppt, we describe a practical programming approach for implementing estimands within ADaMs. The process starts by translating the key estimand attributes - population, variable, intervention effect, intercurrent event strategy and summary measure - into clear programming requirements. These are then built into metadata-driven specs and reusable derivation templates, making implementation more consistent across studies.

The approach strengthens traceability by linking ADaM derivations back to SDTM source data and SAP definitions, while built-in qc checks help ensure that programmed outputs match the intended estimand.

It provides a practical framework that helps implement estimands more efficiently while supporting regulatory expectations and high-quality clinical trial analyses.

Ind_SP_PPT_009: One Object, Every Output: Implementing CDISC's ARS with {cards} and {cardx} in R

Rasagna Navamu, Manasa Chikkela

SCL IT TECHNOLOGIES

The pharmaverse ecosystem provides an integrated suite of open-source R packages that support the entire clinical trial submission pipeline, from SDTM creation using {sdm.oak} and ADaM derivation with {admiral} to the generation of Tables, Figures, and Listings (TFLs) using packages such as {tfrmt} and {gt}. A key advancement in this ecosystem is the CDISC Analysis Results Standard (ARS), which provides a structured, machine-readable framework for storing statistical results and their metadata, ensuring traceability and reproducibility.

The R packages {cards} and {cardx} provide the reference implementation of CDISC ARS in R. {cards} generates Analysis Results Data (ARD) objects for descriptive statistics including means, counts, and percentages using functions such as ard_continuous() and ard_categorical(), while {cardx} extends this into inferential territory with t-tests, Wilcoxon Rank-Sum tests, and regression models — all within the same standardized structure. Together, these packages bridge the gap between ADaM datasets and standardized, submission-ready statistical outputs that comply with CDISC ARS.

Ind_SP_PPT_010: Clinical Intelligence Studio: From Protocol to Trial Design datasets in 2 Minutes each!

Madhuri Madhuri, Madhuri K C

IQVIA

Protocol is the heart and soul of every clinical trial. Every clinical study starts with a protocol, and SDTM Trial Design datasets describes the plan for the procedures to be followed in a clinical trial, including planned and actual timing of events, control group, method of allocating treatments, blinding methods, assignment of epochs that subject pass through in the course of the trial. Right now, this is done manually, one element at a time, straight from pages of protocol text. It works, but it is time consuming and can be optimized from days to minutes for faster review and approval by FDA.

I built a Clinical Intelligence Studio to close that gap.

Upload a protocol PDF, and it returns a structured, SDTM-compliant Trial design dataset- arms, elements, epochs sequence and everything in about two minutes.

Everything runs locally, so the protocol never leaves your machine! This highly matters because the drug/molecule and eventually protocol and the sponsor is highly confidential in every clinical trial.

I have tested it across parallel-group and single-arm designs, with consistent results, and I have extended the same approach to the rest of the Trial Design domains: Trial Elements, Trial Visits, Trial Inclusion/Exclusion, and Trial Summary.

The goal: A complete Trial Design dataset package straight from the protocol, without the manual grind.

Ind_SP_PPT_017: The Submission Confidence Index: One Auditable Number for Clinical Deliverables in R

Pawan Rama Mali

IIT-Ropar

Every clinical submission ends with the same question: is the data ready to go? Today that call is made from validation reports, spreadsheets, and experience, and two people can reach different answers. R4SUB (Ready for Submission) is an open source R framework that turns readiness into a number. It reads the outputs teams already produce, including Pinnacle 21 findings, define.xml, and ADaM and SDTM specifications, and scores four areas: data quality, traceability, statistical risk, and reviewer usability. Those combine into the Submission Confidence Index, a 0 to 100 score with four decision bands, from Not Ready to Ready for Submission. Every point lost is traceable to the finding behind it. The packages are open source under Apache 2.0, with the core on CRAN, and a dashboard and worked example carry one study from raw evidence to a reviewer ready report.

Ind_SP_PPT_020: Redefining Trust, Quality and Accountability in R-Based Clinical Programming: Open Source Doesn't Mean Unvalidated

Pooja Kumari

Eli Lilly

Open-source R is increasingly used in global regulatory submissions, yet a persistent misconception remains : that open-source means unvalidated. This conflates a software license with a quality process. Regulatory accountability has always rested with the organization, not the tool. Addressing this perception is less a technical challenge and more an opportunity to strengthen process maturity across the industry.

Drawing on direct experience as a core {admiral} contributor and Study Lead Programmer on an early R regulatory pilot study, in which both the primary analysis and independent validation were conducted entirely in R within a regulated submission, this presentation offers a grounded, practical perspective on what responsible R adoption looks like.

The talk focuses on three areas. (i) What regulatory agencies require: they mandate a quality process, not a specific programming language, as confirmed by the R Consortium's successful submission pilot studies. (ii) A practical five-tier quality framework covering reproducible environments using {renv} or company-specific controlled environments, risk-based package qualification through the R Validation Hub, unit testing as inspection-ready evidence, QC parity in R-only workflows, and structured audit documentation with clearly defined ownership across programmers, reviewers, and biostatisticians. (iii) AI as a practical enabler, with applications mapped to each tier, including unit test scaffolding, QC discrepancy triage, and audit completeness checking.

Ind_SP_PPT_022: Treat-Through Analysis: Joining an Induction Study and Its Long-Term Extension into One Continuous Dataset

Pallavi Yadav

Eli Lilly

Long-term extension (LTE) studies follow induction and maintenance phases in autoimmune disease clinical trials to assess the durability of treatment response. Although patients progress seamlessly through these phases, parent studies and LTEs are often analyzed separately, obscuring long-term efficacy.

Separate analyses introduce a new baseline at LTE entry, restart the visit schedule, and fragment patient history. From an estimand perspective, re-baselining conditions inference on LTE entrants rather than estimating treatment effect from the original randomized baseline.

We present a treat-through framework integrating the parent study and LTE into a single longitudinal ADaM dataset by retaining the original baseline, aligning visits relative to first

dose, and preserving history, exposure, and intercurrent events, including rescue medication and discontinuation.

In an autoimmune disease development program with a longitudinal efficacy endpoint, we show this approach better assesses long-term efficacy while preserving the parent-study estimand, maintaining traceability to source studies, and reflecting the patient journey.

Ind_SP_PPT_024: Time as a Structural Determinant in Population Pharmacokinetic Modeling

Gayatri Wadghule, Abhishek Dey

Novartis Healthcare Pvt. Ltd.

Population Pharmacokinetic (PopPK) modeling depends on accurate representation of time, as time defines the relationship between drug dosing, systemic exposure, biological variability, and treatment response. Despite its importance, time is often treated as a routine dataset derivation rather than a key structural component of model development. This educational presentation introduces a parameter-centric framework for deriving, visualizing, and interpreting time variables commonly used in PopPK analysis, including time after first dose, time after most recent dose, and component-specific clocks in complex dosing regimens. The framework demonstrates how different concentration–time regions inform distinct pharmacokinetic parameters, with early observations supporting absorption processes, peak concentrations informing bioavailability and distribution, and terminal or trough observations supporting clearance, half-life, accumulation, and steady-state estimation. The presentation also highlights how harmonized time definitions and multi-study pooling improve parameter identifiability, model robustness, and interpretation, enabling reliable pharmacokinetic inference and decision-making. Keywords: Population Pharmacokinetics, Concentration-Time Profile, Time After Dose, ADME Kinetic

Ind_SP_PPT_025: The Art of Debugging in R: From Frustration to Mastery

Madhan Kumar N

ICON Clinical Research India Pvt. Ltd.

In today's programming landscape, developers increasingly turn to search engines and AI assistants by default at the first sign of an error - even when the underlying issue is trivial to diagnose independently. This growing dependency, while efficient in the short term, quietly erodes diagnostic reasoning and deepens frustration when context-specific bugs arise that no forum post or chatbot can resolve.

This presentation champions the idea that debugging is not merely a technical skill but a disciplined mindset - one that can be systematically taught, practiced, and mastered using

R's own native toolkit. The session walks through R's built-in debugging ecosystem: *traceback()* for tracing error origin, *browser()* and *debug()/debugonce()* for interactive step-through inspection, *trace()* for non-invasive code instrumentation, and R's condition system (*tryCatch()*, *withCallingHandlers()*) for structured error handling. It weaves these functions together into a repeatable diagnostic workflow - *reproduce*, *isolate*, *inspect*, *hypothesize*, and *verify*.

Real-world case studies anchor the discussion, each illustrating how a structured, tool-driven approach uncovers root causes that a shallow "copy-paste-the-error-message" search would likely miss. By the end of the session, attendees will learn the practical debugging framework and the confidence to solve problems independently, even in an era of instant answers.

Keywords: R Programming, Debugging Framework, Structured Error Handling, Root Cause Analysis

Ind_SP_PPT_026: AI-Augmented Quality Control for ADaM Datasets in SAS-to-R Migration: An Independent Verification Framework

Shrishti Shandily

Statnomos

Pharmaceutical organizations are increasingly migrating statistical programming from SAS to R, driven by the maturity of the Pharmaverse ecosystem and a broader industry shift toward open-source, reproducible tooling. This migration introduces quality control challenges that conventional independent double programming was not designed to address: differences in missing-value handling, tie-breaking, merge-versus-join semantics, and rounding conventions can produce output differences that are language-semantic artifacts rather than logic errors. This paper proposes an AI-augmented QC framework for ADaM datasets that preserves the blind-independence principle of conventional double programming while addressing these migration-specific failure modes. The framework separates a source-to-ADaM traceability check, an independent AI-built QC implementation blind to the production code, deterministic output comparison, and AI-assisted discrepancy triage that classifies each flagged difference by root cause. We situate this framework relative to recent AI-augmented SDTM review and migration literature, and outline an evaluation plan, governance, and limitations for adoption in a GxP-regulated environment.

Ind_SP_PPT_030: From Manual QC to automated TLF Review: Integrating Rule-Based Validation, AI Agents, Risk-Based Analytics, and Interactive Dashboards

Monica Monica, Samiksha Patil

Novartis Healthcare Pvt. Ltd.

The review of Tables, Listings, and Figures (TLFs) remains a resource-intensive step in clinical trial reporting, relying on repetitive manual quality-control activities. This work

proposes an intelligent review framework that combines automated rule-based validation, AI-assisted review, Failure Mode and Effects Analysis (FMEA)-based risk scoring, and interactive dashboard analytics to enhance review efficiency.

R-based validation modules evaluate population consistency, denominator calculations across statistical outputs, cross-TLF reconciliation, metadata compliance, and run-to-run changes. These findings are consumed by a review agent that interprets outcomes, identifies potential issues, summarizes findings, and recommends outputs requiring manual review. A risk scoring engine aggregates findings across multiple validation domains to prioritize high-impact outputs like efficacy, safety, and time-to-event analyses. An interactive dashboard provides real-time visibility into review status, validation failures, and reviewer workloads.

Ind_SP_PPT_031: Living Longer, Living Better? Decoding Health Utility Beyond Survival Curves

Falguni Dutta

Merck

A patient survives longer. But is that patient truly living better?

Traditional survival analyses tell us how long patients live, yet they often fail to capture how well patients live. What happens when two treatments show similar survival, but patients experience very different daily lives? And how can a single number represent mobility, self-care, pain, anxiety, and overall well-being?

This presentation explores the Health Utility Index (HUI) and its critical role in health technology assessment and value-based healthcare. We will examine how utility measures influence decisions by organizations such as NICE, G-BA, HAS, and ICER, and how instruments including EQ-5D-5L and EQ VAS contribute to Quality-Adjusted Life Years (QALYs).

Through practical examples and reproducible R programming workflows, attendees will learn how utility scores are derived, interpreted, and implemented — bridging the gap between statistical significance, survival benefit, and what truly matters to patients.

Ind_SP_PPT_039: SDTM IG 4.0 and VLM: Redundant or Required?

Mahendrakar Chanukya Samrat, Vipin Gupta, Hema Chimkod

AstraZeneca

The release of SDTMIG 4.0 would represent one of the most significant evolutions in CDISC submission standards in recent years. With extensive metadata restructuring, improved support for machine-readable specifications, and the transition from Supplemental Qualifiers

(SUPPQUAL) toward more integrated metadata approaches, organizations are re-evaluating long-established metadata practices. Among the most debated topics is the future role of Value-Level Metadata (VLM).

Traditionally, VLM has been essential for documenting variable attributes that differ across subsets of data values, particularly within Findings domains where test-specific properties such as units, methods, controlled terminology, derivations, and origins must be clearly defined. Define-XML further established VLM as a critical mechanism for communicating metadata to regulatory reviewers.

As the industry moves toward SDTMIG 4.0, enhanced metadata structures and increasing automation capabilities raise an important question: Is VLM becoming redundant, or does it remain a regulatory and operational necessity? This presentation examines both perspectives.

Ind_SP_PPT_043: ADBASE: Bridging Real-World Data to CDISC for Regulatory-Ready Submissions

Kaushik Dixit

Ephicity Lifescience Analytics

Real-world clinical studies increasingly leverage retrospective healthcare data from diverse sources such as insurance claims, electronic health records, and registries. The heterogeneity of these sources presents significant challenges in harmonizing data to CDISC standards and preparing analysis-ready datasets. In particular, baseline covariates influencing treatment outcomes are often study-specific, non-standard, and require extensive derivations (Summarization of records) and robust cross-domain integration. To address these challenges, ADBASE an analysis-ready baseline dataset designed to standardize and consolidate derived covariates into a single, traceable repository was created. ADBASE streamlines complex derivations, promotes consistency and reusability across analyses, and enables seamless integration with the ADEFF dataset. This framework enhances efficiency, reproducibility, and providing a scalable approach for generating high-quality real-world evidence while aligning heterogeneous healthcare data with CDISC-compliant analysis workflows.

We present challenges with case studies, real examples, and practical approaches to baseline covariates, this framework streamlines derivations, ensures consistency, and enables CDISC-compliant evidence generation.

Ind_SP_PPT_044: AI Has Changed the Rules – Is Your Data Governance Keeping Up?

Himani Narangi

Janssen (Johnson & Johnson)

Data governance has long rested on one premise: de-identify the data, secure it, control access, and it's protected. AI challenges that premise. It can infer sensitive information from anonymous-looking data. Prompts can leak confidential information without anyone noticing. Traditional governance wasn't built for these risks.

This session explores how data governance must evolve to enable responsible AI adoption

while protecting patient privacy, ensuring regulatory compliance, and maintaining trust in clinical data. Drawing on real examples, it exposes where AI outpaces traditional safeguards. A harmless prompt can become high risk when patient context, unapproved tools, unverifiable outputs, or missing audit trails converge. We'll walk through a practical framework spanning data classification, prompt governance, enterprise AI platforms, human oversight, and auditability, plus the AI literacy your teams need.

Attendees will walk away with a checklist to pressure-test their own governance against AI-era risks, before today's innovations become tomorrow's vulnerability.

Ind_SP_PPT_045: SMART TFL Studio

Madhukar Sende, Nanditha Chinni, Shashidhar Berelly

Vaics Consulting.

TFL development in clinical trials remains fragmented, rebuilt study by study with no shared standard, driving programming redundancy, inconsistent outputs, and delayed timelines. SMART TFL Studio addresses this by transforming TFL development into a standardized, reusable, and governed process built on a define once, reuse everywhere principle.

Global templates progress through a governed Development, Review, and Production lifecycle before reaching study teams, anchoring every output to one approved source of truth. At the study level, teams configure a Reporting Area, inherit approved templates, and layer study-specific adaptations through mockshells, without altering the underlying standard. The platform then generates structured metadata from which internally developed functions auto-generate production-ready R code, RTF, and PDF outputs, eliminating manual programming from scratch.

Built on Analysis Results Standard (ARS) principles, SMART TFL Studio delivers efficiency through automation, quality through governance, and scalability.

Ind_SP_PPT_048: Clinical Trial Delivery Digital Twin: A Predictive Decision Framework for Statistical Programming Operations

Vyom Bhatt

Optimapharm

As clinical trials become increasingly complex, statistical programming teams face growing pressure to deliver high-quality outputs within accelerated timelines. Traditional

project tracking methods provide visibility into current study status but offer limited ability to anticipate future risks or evaluate the impact of operational decisions. This paper introduces a Clinical Trial Delivery Digital Twin, a virtual representation of study execution that integrates delivery metrics, resource capacity, quality indicators, and milestone progress into a predictive framework. The digital twin enables scenario-based simulations to assess the impact of changes such as resource constraints, scope increases, review bottlenecks, or process automation. A proof-of-concept implementation demonstrates how predictive analytics and simulation can support proactive decision-making, improve delivery confidence, and identify potential delays before they affect key milestones. By shifting from retrospective reporting to predictive operational intelligence, this approach offers a practical roadmap for transforming statistical programming operations.

Ind_SP_PPT_050: Accelerating Define-XML Creation and Version Comparison Through Automation

Vaisint Singh Muthudurai, Albin Praveenkumar Irudayaraj

Zifo RnD Solutions

In Clinical trials, Define-XML is an important regulatory deliverable, and creating specifications for Define-XML with all standards is very time-consuming. Reviewing define-XML for each update needs extensive manual effort. Here, we developed a tool that automates these tasks and reduces time.

The tool assists in Define Specification generation by taking a few inputs such as SDTM Datasets, aCRFs, CDISC SDTM IG & CT. It creates Dataset, Variable, Code list, and Value level sheets, including length, data type, origin, and aCRF page numbers, etc., in the respective sheets of the pinnacle specification. This tool speeds up Define-XML spec offers a practical roadmap for transforming statistical programming operations.

Creation, reducing the time from DBL to final CDISC package delivery. It also compares Dtwo define-XML versions, highlights changes in separate style sheets, and simplifies review, helping reviewers identify changes faster.

This tool improves both Define-XML creation and review processes and helps developers and viewers work more efficiently, consistently, and with better quality.

Ind_SP_PPT_052: Modernizing Clinical Reporting in R: A Two-Part Approach to Innovation and Scalable Automation

Amisha Kaswankar

Janssen (Johnson & Johnson)

Traditional oncology analyses rely on progression-free survival (PFS) and overall survival (OS) to evaluate treatment benefit. While these endpoints remain fundamental, they summarize outcomes through a single event and may not fully capture how patients transition through clinically meaningful disease states over time.

Beyond knowing whether a patient progressed or survived, what if we could understand where patients are throughout their treatment journey? Multi-state modelling provides a framework to characterize transitions between Response, Stable Disease, Progression, and Death, offering a comprehensive view of disease evolution. This presentation demonstrates a practical R-based workflow using the mstate package to transform oncology efficacy data into multi-state models and generate patient journey visualizations.

State Occupation Probability plots illustrate the probability of patients occupying each disease state over time, providing insights into treatment pathways that may not be apparent through PFS or OS alone. The session will introduce statistical programmers to multi-state modelling in oncology.

Ind_SP_PPT_053: Connecting the Dots: Clinically Meaningful Cross Domain Validation Beyond Rule Based SDTM Checks

Sarika Selvaraj

Precision For Medicine

Rule-based validation tools such as Pinnacle 21 are indispensable for SDTM compliance, identifying structural issues, controlled terminology deviations, and standards conformance. However, because they primarily assess domains independently, clinically meaningful inconsistencies across datasets can remain undetected.

This presentation describes a practical Neo4j-based graph validation framework that complements conventional SDTM review. SDTM domains are converted into a patient-centric knowledge graph, linking subjects, visits, and clinical events across Exposure (EX), Adverse Events (AE), Laboratory (LB), Concomitant Medications (CM), and Disposition (DS). Configurable graph queries evaluate cross-domain scenarios, including adverse events predating treatment, laboratory abnormalities lacking associated adverse event records, medications without supporting clinical evidence, and disposition records inconsistent with exposure history.

A clinical trial case study demonstrates how graph-based validation uncovers cross-domain data quality issues that traditional checks may miss, improving review efficiency and patient-level data integrity. Implementation considerations, lessons learned, and implications for regulatory submission readiness are also discussed.

Ind_SP_PPT_057: Reproducible Machine Learning Framework for Stable Biomarker Selection and Classification in Clinical Trial Data

Hamsika Akkala, Vamshi Reddy Madugula

SCL IT TECHNOLOGIES

Machine learning is increasingly used in clinical research to identify predictive biomarkers, but reproducibility and feature stability remain major challenges in regulated environments. We developed a reproducible machine learning framework integrating Random Forest (RF) and Extreme Gradient Boosting (XGBoost) to classify treatment groups using multimodal biomarker data. The workflow applies standardized preprocessing within nested cross-validation, including missing-value filtering, median imputation, and median absolute deviation (MAD)-based Winsorization. A shadow feature is incorporated as a null reference to distinguish meaningful biomarkers from random noise. Feature importance is aggregated across repeated nested cross-validation runs, and only biomarkers consistently outperforming the shadow reference in both RF and XGBoost are retained, improving selection stability. Final models are evaluated using balanced accuracy, area under the ROC curve, confusion matrices, and bootstrap confidence intervals. This framework provides a transparent, reproducible, and regulatory-friendly approach for robust biomarker discovery and interpretable predictive modeling that can be adapted across therapeutic areas and multimodal clinical datasets.

Ind_SP_PPT_064: Advancing Clinical Research: Harmonizing Analysis Result Standards (ARS) and CDISC 360i

L P Vistaja Singamsetty

GSK Pharmaceuticals

In the complex landscape of clinical research, the integration of standardized frameworks with advanced implementation methodologies is critical for optimizing data management and analysis processes. This presentation explores how Analysis Result Standards (ARS) and CDISC 360 Implementation (CDISC 360i) can collectively advance the quality and efficiency of clinical trials. ARS provides a robust structure for consistent reporting of analysis results, ensuring clarity, reliability, and comparability across studies. CDISC 360i takes this standardization further by enabling executable metadata and automated workflows, thereby streamlining data handling and facilitating adherence to regulatory requirements. The combined application of ARS and CDISC 360i results in improved data integrity, traceability, and reproducibility, paving the way for smoother regulatory submissions and more credible research findings. This integration not only supports high-quality data processes but also advances the future of clinical trial methodologies, offering a pathway towards more efficient and standardized clinical research.

Ind_SP_PPT_069: From Manual Processing - Building reusable AI framework for clinical trial data workflows

Niyamat Sahi

Eli Lilly

Clinical trial workflows rely heavily on structured, repetitive, domain-specific data processes — spanning programming, coding, data manipulation, and submission-related

tasks. Without automation, this work stays manual, inconsistent across studies, and disconnected from clinical interpretation, consuming significant programmer and reviewer time on repetitive review.

This presentation introduces a unified AI automation framework built on two complementary layers: a Large Language Model (LLM) layer for clinical interpretation and semantic mapping, and a Machine Learning (ML) layer for anomaly detection and version consistency checks. By replacing manual, repetitive review with automated flagging, mapping, and narrative generation, the framework frees programmers and reviewers to focus on higher-value clinical judgment.

The framework is demonstrated through two case studies — automated PRO scoring and MedDRA dictionary automation — showing how the same LLM + ML architecture applies across different clinical programming tasks. Rather than a one-off fix, this represents a scalable, reusable, audit-ready AI architecture that turns manual clinical programming work into a consistent, automated process.

Ind_SP_PPT_078: Automated Bidirectional Verification of Origin References between Define.xml-to-aCRF using Python

Sonali Kimbhar, Amisha Kaswankar

Inference

Define.xml is a standardized XML metadata file used in clinical trial submissions to regulatory agencies such as the U.S. Food and Drug Administration (FDA). It provides metadata describing the structure, content, and origin of submitted datasets, supporting regulatory review and compliance with CDISC standards.

This presentation describes a Python-based automated tool for validating the consistency and completeness of origin references between Define.xml and the annotated Case Report Form (aCRF). This tool performs bidirectional verification by comparing Define.xml-to-aCRF and aCRF-to-Define.xml origin references to identify missing, inconsistent, or mismatched annotations.

The proposed approach parses Define.xml metadata, extracts origin and page reference information, and compares these with annotations extracted from the aCRF. This tool streamlines quality control and supports faster, more reliable regulatory submissions.

Ind_SP_PPT_079: Integrating Natural History and Rare Disease Data : A Programming Perspective

Swetha Balasubramanian

ACL Digital

Rare disease research is often hindered by limited patient populations, the absence of suitable control groups, and a lack of comprehensive understanding of disease progression.

Natural History (NH) data provides valuable insight into the untreated course of a disease and serves as an important benchmark for evaluating treatment outcomes.

This paper presents a programming-based framework for integrating NH and study drug data to support comparative analyses in rare disease development. NH and treated patient populations were aligned using baseline matching techniques, including propensity score methods, to improve comparability and minimize potential bias. Standardized data models, harmonized derivations, and rigorous quality control procedures were implemented to create a unified analytical dataset from multiple sources.

The integrated framework enabled consistent evaluation of disease progression and clinical outcomes, allowing treatment effects to be assessed against the expected disease trajectory. This approach improves data consistency, reproducibility, and efficiency, supporting evidence generation and informed clinical and regulatory decision-making in rare disease research.

Ind_SP_PPT_080: Precision Dose Finding in Alzheimer's Disease Clinical Trials: Implementing MCP-Mod Using R's Dose Finding Package

Gunasekaran Panchamoorthy

ACL Digital.

Selecting the optimal dose for Alzheimer's disease (AD) treatments remains a high-stakes hurdle in clinical development. Traditional pairwise comparisons often fail because they treat quantitative dose levels as isolated categories. This presentation demonstrates a practical application of the FDA/EMA-qualified MCP-Mod framework utilizing the open-source R Dose Finding package in a simulated Phase II AD trial. Using the package's `Mods()` function, a diverse candidate set of parametric models (including Emax and sigmoidal Emax) was pre-specified to manage model uncertainty. The `MCTtest()` function established a statistically significant dose-response signal under strict Type I error control. Finally, continuous curve fitting and model averaging (`fitMod`) were leveraged to seamlessly interpolate between tested cohorts and calculate the Minimum Effective Dose (MED). This programmatic workflow highlights how the R Dose Finding package provides a highly accessible, data-dense, and regulatory-compliant strategy to optimize dose selection for pivotal Phase III AD trials.

Ind_SP_PPT_084: Who Needs a Crystal Ball? How AI, ML, and NLP Are Transforming Real-World Data into Real-World Evidence

MAMATHA PANCHIKA, Pavani Dayanidhi, Srikanth Jula

PANAKEIA

Randomized clinical trials remain the foundation for evaluating the safety and efficacy of new therapies, yet they do not always reflect how treatments perform in routine clinical practice. The COVID-19 pandemic generated enormous volumes of Real-World

Data (RWD) from electronic health records, vaccination databases, insurance claims, and patient registries, creating both opportunities and challenges for evidence generation. This presentation examines how Artificial Intelligence (AI), Machine Learning (ML), and Natural Language Processing (NLP) transform these diverse data sources into reliable Real-World Evidence (RWE). Drawing on COVID-19 vaccine studies, it demonstrates how these technologies evaluated vaccine effectiveness, identified potential safety signals, and extracted clinically meaningful information from unstructured medical records. By complementing evidence from clinical trials, AI-driven RWE offers a broader understanding of treatment performance across diverse populations, enabling more timely clinical decisions, supporting regulatory evaluations, and advancing patient-centered drug development and future healthcare innovation.

Ind_SP_PPT_085: A Metadata-Driven Implementation Framework for Oncology PRO Programming

Divya Ramesh

IQVIA

Objective:

FDA's 2023 guidance for oncology patient-reported outcome (PRO) submissions introduced expectations for expected assessments, phantom records, analysis datasets, and supporting documentation. However, implementation often remains sponsor-specific and involves recurring study-level programming activities.

Approach:

This presentation describes a metadata-driven framework for recurring oncology PRO programming using realistic dummy studies with different instruments and visit schedules. Study-specific settings—including visit schedules, expected assessment rules, instrument-to-parameter mappings, completion thresholds, and supported scoring methods—are maintained in metadata. Reusable components use these definitions to generate expected and phantom records, derive analysis parameters and visits, and perform standardized ADQS quality-control checks.

Outcomes & Benefits:

The framework standardizes recurring implementation activities, producing consistent ADQS-ready outputs while reducing duplication and complexity, improving traceability and maintenance, and promoting reuse of validated components across studies.

Summary:

The framework supports consistent and maintainable PRO implementation, while study-specific scoring methods, derivations, analyses, and sponsor requirements remain under programmer and statistician oversight.

Ind_SP_PPT_086: HUMAN-GUIDED AI PROGRAMMING - From RECIST Interpretation Risk to Submission-Ready Outputs

VENKATA SATISH KONIKI

eClinical

Solutions

Clinical programming defects in oncology trials often begin before SAS code is written: in RECIST interpretation, response sequencing, cross-domain tumor assessment review, and PFS censoring decisions. This presentation demonstrates Human-Guided AI Programming as a practical way for statistical programmers to improve decision quality from SDTM to ADaM to TLF outputs. Using a RECIST-based oncology case study, conventional programming is compared with AI-generated and human-guided AI approaches for new lesion PD review, assessment-date selection, BOR confirmation, death-before-progression, missing-assessment censoring, and TLF denominator consistency. The focus is not replacing programmers with AI, but using AI as a structured review partner to challenge assumptions, identify missing edge cases, suggest reusable Base SAS patterns, and strengthen QC traceability. The proposed workflow supports earlier risk detection, consistent programming decisions, and submission readiness while preserving human accountability.

Keywords: Human-Guided AI Programming, RECIST, SDTM, ADaM, TLF, BOR, PFS, QC, submission readiness

Ind_SP_PPT_087: Creation of Trial Design Datasets & Minimize P21 Compliance Issues

Kiran Kumar Pendyala

eClinical Solutions

Trial Design datasets (TA, TE, TI, TS, and TV) are essential components of SDTM submissions, defining the planned study design and ensuring consistency across clinical trial data. Although these datasets are relatively small, errors in their structure, metadata, or controlled terminology frequently result in Pinnacle 21 (P21) validation findings, leading to increased rework and potential submission delays. This presentation discusses best practices for creating compliant Trial Design datasets while minimizing common P21 issues. It highlights frequently encountered validation findings, their root causes, and practical strategies to prevent them through standardized specifications, robust programming practices, early validation, and comprehensive quality checks. By adopting these approaches, clinical programming teams can improve first-pass P21 compliance, enhance submission readiness, reduce development effort, and deliver high-quality

SDTM Trial Design datasets that meet regulatory expectations and support efficient regulatory submissions.

Ind_SP_PPT_091: Biomarker Data Programming in Clinical Trials

Venu Barre

eClinical Solutions

Biomarker data are increasingly used to understand disease status, patient risk, treatment response, and subgroup effects in clinical trials. For SAS programmers, the challenge is not only to load vendor files, but to convert complex laboratory, tissue, genomic, and blood-based results into traceable and analysis-ready datasets. This presentation outlines a practical programming approach for handling biomarker data, starting with the analysis question and continuing through data transfer specification review, raw data checks, standardization, derivations, population flags, and output support. Generic examples show how expression results, mutation status, circulating tumor DNA, and laboratory markers can be classified and linked with efficacy endpoints such as response and time-to-event analyses. The key message is that effective biomarker programming requires technical accuracy, clear documentation, scientific understanding, and early collaboration with statisticians, data managers, vendors, and subject matter experts.

Ind_SP_PPT_092: Teal Programming Genomic Data for Clinical Trial Readiness
Chandrasekhara N
eClinical Solutions

Next-generation sequencing (NGS) and related genomic assays are increasingly used in clinical trials to support patient selection, disease monitoring, and exploratory treatment evaluation. For statistical programmers, these data are challenging because results can come from tissue, blood, external vendors, central laboratories, or diagnostic partners, and may vary in format, terminology, reference genome, assay timing, and quality-control status. This presentation outlines a practical SAS programming framework for preparing genomic data from receipt to analysis readiness. It covers transfer specification planning, source data profiling, specimen and assay metadata, gene and variant harmonization, mapping to standard findings structures, and creation of subject-, visit-, and variant-level analysis views. The focus is on traceability, missing-data handling, QC failures, derivation documentation, and review packages. The key message is that genomic programming requires technical discipline, scientific context, and collaboration with statisticians, vendors, bioinformaticians, and subject matter experts.

Ind_SP_PPT_095: Hybrid Analytics: A Scalable SAS-R Framework for Real-World Evidence Analytics
Kishore Kumar and Annai Esther
ACL Digital

Real-World Evidence (RWE) programs generate increasingly complex and heterogeneous clinical datasets that challenge traditional statistical programming workflows. Large-scale observational databases often contain fragmented patient records, longitudinal treatment histories, multiple relational data sources, and inconsistent clinical assessments, making scalable, reproducible analysis difficult. This work presents a hybrid analytics framework that combines SAS for regulatory-grade data processing, harmonization, and reproducible dataset derivation with R for advanced visualization and exploratory analytics. The framework incorporates a unified relational data architecture, staged data harmonization, predefined validation rules, and independent double-programming methodologies to improve analytical consistency and mitigate bias. Applied to a large medical device ecosystem comprising approximately 140,000 patients, four million treatment sessions, and over 130 relational tables, the framework enabled reliable longitudinal analyses and publication-ready outputs. The proposed approach demonstrates how hybrid statistical programming can strengthen governance, improve reproducibility, and transform complex real-world clinical data into trustworthy, decision-ready evidence for modern evidence generation.

Ind_SP_PPT_099: Metadata-Driven PRO Programming: A Reusable Framework for Clinical Trials

Sabharish Sekar, Pradeep Subramaniam

Inference

Patient-Reported Outcomes (PROs) are widely used in clinical trials to assess symptoms, functional status, and health-related quality of life. Despite differences in questionnaire design, instruments such as SF-36, EQ-5D, FACT-G, EORTC QLQ-C30, and PROMIS require similar programming activities, often resulting in duplicated effort and inconsistent implementations.

This presentation introduces a metadata-driven automation framework that separates questionnaire-specific business rules from programming logic. Centralized metadata and a generic scoring engine enable automated data validation, scoring, SDTM and ADaM derivations, and quality control across multiple PRO instruments. The framework improves consistency, traceability, CDISC compliance, and the efficiency of delivering submission-ready datasets.

Ind_SP_PPT_100: An efficient R Programming pipeline for Gene Set Enrichment Analysis: A Reproducible Workflow for Functional Interpretation of High-Throughput Gene Expression Data

Sagnik Datta

Avance

Although genome wide RNA expression analysis has become a routine tool in biomedical research, creating a need for robust computational methods to extract meaningful biological insights remains a major challenge. In this presentation we will describe an efficient analytical method named Gene Set Enrichment Analysis (GSEA) that has emerged as a powerful statistical approach for identifying coordinated changes in predefined gene sets, enabling researchers to interpret genome-wide expression profiles in the context of biological pathways and functional annotations. We will present a reproducible and efficient workflow for performing GSEA using the R programming environment.

The proposed workflow will integrate widely used Bioconductor packages to facilitate data preprocessing, gene ranking, enrichment analysis, and visualization. Differential gene expression results will be used to generate ranked gene lists, followed by enrichment analysis against curated gene sets. The workflow produces comprehensive outputs, including normalized enrichment scores (NES), false discovery rate (FDR)-adjusted p-values, enrichment plots, ridge plots, dot plots, and pathway network visualizations.

A publicly available RNA-sequencing dataset is used to demonstrate the applicability of the workflow. By leveraging open-source tools and standardized analytical practices, this workflow will provide researchers with a practical framework for functional genomics studies and enhance the biological interpretation of transcriptomic datasets.

Data Science Talks

DS_PPT_001: From Fragmented Handoffs to a Governed Digital Thread: A human-led, agentic AI model for the clinical study delivery lifecycle - from protocol to CRT package.

Yasa Sreekanth Reddy

IQVIA

Clinical study delivery is still more fragmented than we often admit. A single protocol moves through clinical, data management, biostatistics, statistical programming, standards, medical writing, and regulatory teams - but each function often interprets it separately, at different

points in time, and in different documents. As a result, alignment issues are frequently discovered late, during QC, audit, or submission preparation. A missed endpoint, unclear CRF requirement, inconsistent SDTM mapping, incomplete ADaM traceability, or reviewer guide gap can create rework across the SAP, CRF, SDTM, ADaM, TFLs, and CRT package. The issue is not lack of expertise; it is disconnected interpretation.

This proposal presents a human-led agentic AI model to create a connected digital thread from protocol to submission. Instead of relying on one black-box engine, the model uses a governed network of specialized agents, each with clear inputs, outputs, ownership, and human review gates. The lifecycle is organized into four phases: Plan - protocol, SAP, CRF design, and TFL shells; Build - database metadata, synthetic raw data, SDTM specification, and aCRF mapping; Program - SDTM, ADaM, and TFL programs; and Submit - cSDRG, ADRG, define.xml, validation outputs, and CRT package readiness.

A key feature is the early use of mock or synthetic raw data to test SDTM specifications and programming logic before live data is available. This helps teams identify traceability and standards issues earlier, when they are easier and cheaper to fix.

The model is built around six governance controls: human review, source-to-output traceability, version control and audit trail, independent QC, CDISC/ICH alignment, and clear positioning of AI as an accelerator, not a decision-maker.

The goal is not to replace clinical experts. It is to help them work from a shared, traceable, and defensible delivery model - reducing rework, improving consistency, and strengthening submission readiness.

DS_PPT_008: Using IPTW to Evaluate Treatment Effectiveness in Real-World Evidence

Soundharya M and Pooja S Hugar

Fortrea

Real-world evidence plays an increasingly important role in evaluating treatment effectiveness; however, observational studies are inherently prone to confounding. Propensity score-based methods, particularly inverse probability of treatment weighting (IPTW), offer a robust approach for reducing measured confounding and improving causal inference under appropriate assumptions. IPTW estimates the probability of receiving a treatment based on observed baseline characteristics and uses these probabilities to create a weighted pseudo-population in which measured covariates are balanced across treatment groups. This enables less biased estimation of average treatment effects while preserving the full study cohort. This presentation will demonstrate the practical application of IPTW using real-world healthcare data to evaluate the effectiveness of multiple therapeutic interventions and assess long-term patient outcomes. Through real-world examples, the presentation will highlight methodological considerations, assumptions, and challenges associated with IPTW, illustrating how these approaches transform complex healthcare data into actionable evidence that supports regulatory decision-making and informs clinical practice.

DS_PPT_014: AI is Accelerating RWE. Are We Accelerating Bias Too?

Yash Chaudhari

Fortrea

Artificial Intelligence is rapidly becoming part of the Real-World Evidence (RWE) ecosystem, helping analysts:

1. Generate cohort definitions
2. Develop code lists
3. Write SQL and programming code
4. Support interpretation of findings

While these capabilities promise faster study execution and improved productivity, they also introduce a new and often overlooked source of bias. AI-generated outputs may appear scientifically sound while embedding incorrect assumptions, incomplete clinical logic, hidden programming errors, or automation bias that can influence study results and conclusions.

As AI adoption grows, the focus must extend beyond efficiency to evidence credibility and reproducibility.

This presentation explores where bias can enter AI-assisted RWE workflows, highlights practical examples across study design and analytics, and discusses a pragmatic framework for validation, governance, and human oversight. The goal is not to slow AI adoption, but to ensure that accelerated evidence generation remains transparent, reliable, and scientifically robust in an increasingly AI-enabled future.

DS_PPT_016: A Metadata-Driven Approach to Automating SAP-to-ADaM Specification Generation

Ashwini Reddy and Ajay Tomar

SAS

The increasing complexity of clinical trials and the growing demand for standardization have highlighted the need to automate the translation of Statistical Analysis Plans (SAPs) into Analysis Data Model (ADaM) specifications. We present an automation workflow leveraging Natural Language Processing (NLP) and the CDISC Library API to systematically map SAP text to standardized ADaM variables and structures. It uses NLP techniques to extract key analytical concepts from SAP documents, including populations, endpoints, derivations, and timepoints. These concepts are then aligned with ADaM metadata retrieved dynamically through the CDISC Library API, ensuring compliance with CDISC standards and controlled terminology. The workflow generates draft ADaM specifications, including dataset structures, variable mappings, and analysis flags, significantly reducing manual effort and variability maintaining traceability.

This automation framework accelerates study build timelines, improves consistency across studies, and enables organizations to adopt a standardized, scalable, and intelligent metadata-driven programming model for clinical trial analysis.

DS_PPT_017: Development of a Computable Phenotype Algorithm to Identify Multimorbidity Patients Using Real-World

Khalid Amar

Fortrea

Real-world evidence (RWE) derived from electronic health records (EHRs) and administrative claims is increasingly used to support clinical research and healthcare decision-making. However, identifying patients with multimorbidity remains challenging due to inconsistent coding practices and heterogeneous data sources.

This study proposes a standardized, rule-based computable phenotype algorithm to identify multimorbid patients using structured real-world data. The framework integrates diagnosis codes, medication records, longitudinal observation, and temporal validation to distinguish persistent chronic diseases from transient conditions. Patients meeting validated criteria for two or more chronic diseases are classified as multimorbid.

The proposed approach provides a reproducible methodology that is adaptable across EHR and claims databases, facilitating reliable cohort identification for epidemiological studies, healthcare utilization research, and comparative effectiveness analyses. Standardized computable phenotypes have the potential to improve consistency, reproducibility, and validity of RWE studies in internal medicine.

DS_PPT_019: Testing what matters: Journey-Based Framework for eCOA UAT

Namev Kutte

Eli Lilly

Background: eCOA UAT often repeats functional and regression testing, consuming effort while missing issues that emerge in real-world study use.

Purpose: Reposition UAT from build verification to validating how patients, caregivers and sites experience the study end-to-end.

Methods: A journey-based approach in six moves: start at design review; write independent test scripts; cover complete user journeys; follow the data end-to-end; apply risk-based depth; extend same testing beyond go-live and mid-study changes.

Journey focus: The framework targets hand-offs, edge cases, integrations and endpoint risk that screen-by-screen checks will miss.

Outcomes: Journey-based UAT reveals defects such as hand-offs, safety alerts, offline sync, and data reconciliation before they reach patients. Catching issues early can reduce fix costs by up to 20 times while supporting 90%+ patient compliance, fewer PPCs, and fewer data change requests. Helpdesk tickets reduced by 60%.

Conclusion: Fewer post-go-live surprises, less rework, and measurably lower patient and site burden — a repeatable model any eCOA study can reuse.

DS_PPT_020: CRISPR-GPT: A Multi-Agent Large Language Model System for End-to-End Gene-Editing Automation

Karthick Sathiya

Pfizer

CRISPR-based genome engineering has revolutionized functional genomics and therapeutic research; however, designing gene-editing experiments remains complex and requires expertise in molecular biology, bioinformatics, and experimental optimization. CRISPR-GPT is an agentic artificial intelligence framework that automates end-to-end CRISPR workflow design through a four-agent architecture consisting of a User Proxy, LLM Planner, Task Executor, and Tool Provider. The framework integrates genomic databases, genome annotation resources, sgRNA design tools, off-target prediction algorithms, primer-design software, and protocol-generation modules. It supports major genome-editing modalities including gene knockout, base editing, prime editing, and CRISPRa/i applications. Evaluation across 288 gene-editing scenarios demonstrated high accuracy in target selection, guide RNA design, and validation planning. By combining large language model reasoning with biological tools, CRISPR-GPT enables reproducible, scalable, and expert-level genome-engineering workflows while reducing technical barriers and accelerating scientific discovery.

DS_PPT_021: A Conversational, DAG-Driven Framework for Clinical Programming

Vigneshwara Chandrasekaran and Sushanth Poojari

Ephicity Lifescience Analytics

Clinical programming is transitioning toward metadata-driven, AI-assisted development to support reproducibility, regulatory compliance, and open-source adoption. However, current large language model (LLM)-based code generation relies primarily on prompt engineering, limiting dependency awareness and workflow explainability. We developed a prototype **DAG-based conversational framework for clinical programming** that combines conversational AI with a multi-level dependency graph representing datasets, variables, derivation rules, and execution dependencies. Natural language requests are interpreted by a planner agent that traverses the dependency graph to identify prerequisite source domains, derivation functions, and execution sequences before generating executable R workflows. The prototype integrates Apache Arrow for clinical data management and the targets framework for reproducible workflow orchestration. Using SDTM-to-ADaM dataset generation as a representative use case, the prototype achieved dependency-aware workflow planning, metadata-driven validation, traceability, and reproducible execution, providing a modular and scalable foundation for conversational clinical programming.

DS_PPT_022: CT Studio: A Standardized Framework to Automate Controlled Terminology Mapping and Drive Scalable Data Governance

Fysal Kasim and Abin E Savio

Reconciling raw clinical trial data with CDISC Controlled Terminology (CT) is a resource-intensive step in SDTM development, requiring manual spreadsheet searches to identify compliant terminology and assess codelist extensibility. Conventional text-matching approaches may produce false-positive mappings by relying on lexical similarity alone. This presentation introduces **CT Studio**, a domain-agnostic framework that streamlines CT mapping using a multi-vector alignment strategy across CDISC submission values, synonyms, and clinical definitions. A term-specificity assessment distinguishes general concepts from modified parameters, while a confidence-based decision gate returns blank results for uncertain matches, routing ambiguous mappings for human review. CT Studio is designed as an extensible architecture supporting future machine learning integration using historical mappings and approved human overrides. The framework aims to reduce manual effort, improve mapping consistency, and provide a scalable, audit-ready foundation for controlled terminology governance across clinical development programs.

DS_PPT_023: Beyond Static Listings: Transforming Clinical Trial Safety Review Through Interactive Risk Intelligence

Darshana Sawant

Pfizer

Clinical trial safety review traditionally relies on static reports and manual exploration of large volumes of safety data, making identification of emerging risks increasingly challenging as trial complexity grows. This presentation introduces an interactive risk intelligence framework designed to support proactive safety surveillance through data-driven prioritization and visualization. The framework integrates adverse events, serious adverse events, adverse events of special interest, laboratory parameters, exposure information, vital signs, and patient-level safety profiles within a unified analytical environment. Configurable risk indicators combine multiple safety dimensions to highlight subjects, events, and trends requiring focused review. Interactive analytics enable movement from study-level surveillance to patient-level investigation, supporting rapid exploration of potential safety concerns. By transforming fragmented safety outputs into actionable intelligence, the framework enhances review efficiency, transparency, and decision support while establishing a scalable foundation for future predictive analytics and AI-assisted safety monitoring.

DS_PPT_025: From Fragmented Data to Focused Action: A Posit Framework for Recruitment Oversight

Rashi Rathi

Novo Nordisk

Patient recruitment can make or break a clinical trial, yet the information needed to monitor progress is often scattered across separate reports, making it difficult to answer key operational questions: Where is recruitment gaining or losing momentum? Which sites may need support? Are enrolled participants aligned with planned recruitment targets?

To address this challenge, we developed and deployed a Posit-based recruitment oversight framework that brings data from 13 clinical trials and six comparable benchmark studies across more than 20 countries into a standardised reporting pipeline. The framework combines automated weekly refreshes, data timestamping, quality checks and an interactive dashboard to provide timely, consistent visibility of recruitment progress across studies, countries and sites. In routine use, it has reduced reliance on manual report preparation and improved the availability of refreshed recruitment insights, helping teams identify emerging recruitment imbalances.

This presentation will share the data model, implementation lessons, and opportunities to extend the framework to other operational metrics.

DS_PPT_026: AI Never Sleeps: Turning Millions of Clinical Data Points into Intelligent Decisions: Connecting the Dots Unifying Clinical Data Beyond EDC

Rishi Raj

Sensan Biosciences

Julia is rapidly emerging as a powerful and exciting tool in clinical data analytics, combining high-performance computing with an intuitive, expressive syntax. Designed for scientific and numerical computing, Julia is transforming how we analyze and interpret complex clinical trial data. Its speed rivals that of C, while its ease of use is comparable to Python or R - making it ideal for handling the growing volume and complexity of data in the pharmaceutical industry. Beyond analysis, Julia also offers strong capabilities in data visualization through libraries like Makie and Gadfly, enabling the creation of interactive, publication-quality graphics that enhance data storytelling and insight generation. This presentation explores how Julia can accelerate statistical analyses, support real-time simulations, integrate with existing tools, and bring reproducibility and transparency to clinical workflows. With its rapidly growing ecosystem and adoption, Julia is set to play a key role in modernizing data science across pharma and shaping the future of clinical research.

DS_PPT_028: A Centralized Dashboard for Monitoring Continuous Glucose Monitor (CGM) Data Quality and Compliance in Clinical Trials

Priyanshu Argal

Eli Lilly

Background:

Continuous Glucose Monitor (CGM) devices generate large volumes of high-frequency data - a typical device records a reading every 5 minutes, producing approximately 288 readings per subject per day, or over 2,000 points per subject per week. In a multi-site trial enrolling several hundred subjects over a multi-month wear period, this can translate into millions of CGM records requiring review across the study. Common issues such as missing CGM visits, sessions, duplicate records, out-of-window readings, and inconsistent subject compliance can be difficult and time-consuming to identify manually at this scale, often requiring 12-14 hours/week of manual reconciliation per month among 5-6 each study. These challenges limit the ability of study teams to proactively monitor data quality and site performance in near real-time.

Objective:

To evaluate the impact of a centralized CGM dashboard on reducing review time and improving the timeliness of data issue detection for the Data Management team.

Methods:

CGM data are transferred monthly from the CGM vendor using a standardized SAS dataset specification and loaded into the clinical database. The dashboard was designed as a standardized, reusable framework and has been deployed across multiple concurrent studies, applying a consistent set of automated

quality checks to CGM values, header records, and event data. These checks identify missing records or incomplete sessions, detect duplicate entries (e.g., arising from daylight-saving time transitions or backfilled transmitter records), and flag CGM readings outside protocol-defined visit windows. Compliance is calculated at the daily, session, and subject levels: a subject-day is classified as compliant if the proportion of expected CGM readings captured (of the expected 288 readings per day, based on 5-minute sampling) meets or exceeds a pre-specified threshold [e.g., $\geq 70\%$]; session- and subject-level compliance are derived by aggregating daily compliance across the corresponding visit window or full monitoring period. Visit and dosing information from the Electronic Data Capture (EDC) system is integrated to provide protocol-specific context for each CGM recording, enabling near real-time identification of data quality issues in place of manual, end-of-cycle reconciliation.

Results:

The dashboard is now in use across five to six locked and ongoing studies, giving study teams a clear, visual way to track CGM data quality by country, site, subject, and visit.

As one example, in a study with over 21,000 expected CGM records, the dashboard caught more than 6,400 missing readings, over 3,100 missing sessions, and about 100 duplicate entries — issues that would otherwise take considerable manual effort to find. It also showed that roughly half of subjects were falling below expected data-availability levels, making it easy to see exactly which sites and subjects needed follow-up.

In another study, the dashboard flagged over 52,000 non-compliant records, with a simple site-by-site breakdown that made it obvious where the biggest gaps were and which sites needed outreach first.

Across all studies, common issues like daylight-saving-time duplicates, backfilled records, and manual calibration flags are now caught automatically as soon as data comes in — instead of being found much later during manual reconciliation.

Conclusion:

Implementing a standardized, program-level CGM dashboard has the potential to reduce manual data review effort, improve the timeliness of issue detection, and provide greater visibility into cross-study trends. Early experience across multiple studies suggests this centralized approach can strengthen oversight of CGM data quality and support more efficient, consistent data management throughout the clinical trial lifecycle. Next steps include formally measuring review-time savings and issue-detection turnaround, with the goal of scaling this dashboard framework to additional studies and other high-frequency data sources.

DS_PPT_032: Generative AI in Clinical Data Management: A Prototype for AI-Assisted Data Issue Identification

Pallavi Kharabe

Inference

Clinical Data Managers routinely review large volumes of study data to identify discrepancies, protocol deviations, missing assessments, and inconsistencies across related case report forms. Although rule-based edit checks identify predefined issues, many review activities still require manual interpretation of clinical context.

This presentation explores how Generative Artificial Intelligence (GenAI) can support, rather than replace, the clinical data review process. A prototype AI-assisted data review and report generation tool will be presented to demonstrate how GenAI can interpret structured clinical data, summarize potential issues, and prioritize records requiring human review. In this session will also discuss the case study, practical challenges, including data privacy, model reliability, regulatory expectations, and the importance of human oversight in regulated clinical research.

Attendees will gain an understanding of the opportunities, limitations, and practical considerations for incorporating GenAI into Clinical Data Management workflows while maintaining human decision-making and a risk-based approach to data review.

DS_PPT_033: From Protocol to Programming: USDM and Biomedical Concepts as the Backbone of Trustworthy Agentic AI

Chaitanya Repaka

Ailens

SDTM mapping remains one of the most labor intensive stages of clinical trial analysis. Programmers manually assemble mappings from scattered inputs: free text protocols, annotated CRFs, raw EDC and vendor datasets, SDTMIG, Controlled Terminology, and sponsor standards, with discrepancies surfacing late, often only during QC. Agentic AI is often demoed but rarely trusted with this work, since agents fed the same unstructured documents inherit the same ambiguity.

This session shares lessons learned from building and fine tuning AI LENS production agentic AI, which reasons over the connected CDISC standards chain instead: USDM for study design, Biomedical Concepts and SDTM Dataset Specializations for concept to domain traceability, and SDTMIG and Controlled Terminology for variable level rules. In production, this has reduced mapping specification turnaround by roughly 30 to 40 percent and mapping discrepancies by approximately 20 to 30 percent versus the manual baseline.

DS_PPT_035: Beyond Hematology: AI-Driven Risk Prediction and Biomarker Discovery for Tumor Lysis Syndrome in Solid Tumors and Novel Therapies - A Precision Medicine Framework.

Kirankumar Vyas

Veramed

Tumor Lysis Syndrome (TLS) is an oncologic emergency historically confined to hematologic malignancies but increasingly reported in solid tumours and patients receiving novel immunotherapies, where risk remains poorly characterized. We present a precision oncology framework integrating artificial intelligence (AI) and multi-omic biomarkers to enable early TLS risk prediction in these underrepresented populations. A multi-modal machine learning pipeline incorporating clinical variables, longitudinal laboratory trends, imaging-derived features, and genomic data (including circulating tumor DNA) is proposed to generate individualized risk scores. Advanced feature engineering, ensemble modelling, and rigorous validation strategies, including temporal and external cohort validation, are outlined. Complementary biomarker discovery using metabolomics and cfDNA fragmentation profiles is explored for preclinical detection. Through representative case scenarios and a proposed multicentre registry, this work demonstrates the clinical and economic potential of AI-guided prophylaxis. This approach aims to shift TLS management from reactive intervention to proactive, personalized risk mitigation across diverse oncology settings.

DS_PPT_036: Context-Aware AI Assistance within a Unified Clinical Programming Workbench

Rohan Vangari and Shashidhar Berelly

Vaics Consulting

Clinical programming teams often work across fragmented development environments, manual data-sharing processes, and disconnected knowledge sources, creating version-control challenges, inefficient collaboration, and inconsistent workflows. These issues become increasingly significant as study complexity grows.

This presentation introduces an integrated clinical programming workbench that unifies development, data management, and AI-assisted programming within a cloud-based environment. Users can provision RStudio, Jupyter Notebook, or Visual Studio Code workspaces, develop or synchronize code through Git, generate and validate SDTM, ADaM, and TFL outputs, and commit validated programs through structured version-controlled workflows. A shared network-based workspace provides secure, controlled access to common study datasets across all workbenches within a reporting area, with dataset locking and automated synchronization to centralized storage. The platform also embeds an AI programming assistant that leverages implementation guides, controlled terminology, metadata, MedDRA, and study-specific context to answer programming questions and assist with code debugging.

DS_PPT_038: Reimagining Double Programming: Agentic AI as an Independent QC Automation Layer for ADaM and TFLs.

Mahesh Divakaran

Ephicacy Lifescience Analytics

Double programming is our default safety net for TLF quality control, but is it truly independent? In practice, QC programmers share the same training, borrow similar logic, and sometimes glance at production code — quietly undermining the whole point. We tried something different: a multi-agent LLM pipeline that reads only the SAP, shells, and ADaM specs, then writes and runs its own Python analysis code, while production stays in R. No access to production programs, ever. A comparison agent reconciles outputs cell by cell and flags each mismatch as rounding, logic, or — surprisingly often — ambiguous specification wording.

This paper walks through the pilot results across demographic, AE, and efficacy tables; what the agent caught that human QC missed; where it stumbled; and how sign-off gates and full decision logs keep the approach defensible under GxP. As AI governance frameworks mature, this approach has the potential to evolve into a GxP-compliant QC model through validated workflows, complete traceability, and sustained human oversight.

DS_PPT_040: AI/ML-Enabled Direct Capture of Imaging-Derived Tumor Response Data into EDC: A Strategy to Minimize Transcription Error and Standardize Response Interpretation

Nithyashree Shivakumar

Eli Lilly

Manual transcription of radiographic tumor response data—lesion diameters, sum-of-diameters calculations, and categorical response assignments (RECIST 1.1, iRECIST) into Electronic Data Capture (EDC) systems remains error-prone, contributing to inconsistent readings and delayed data cleaning. This concept proposes an AI/ML-enabled workflow that captures imaging data directly from radiology source systems, automating tumor response calculations to reduce manual transcription and subjective interpretation. An automated pipeline would extract structured lesion measurements from radiology and PACS (Picture Archiving and Communication System) outputs, applying validated criteria to derive response categories (complete response, partial response, stable disease, progressive disease). Natural language processing would handle unstructured reports lacking structured fields, flagging discrepancies for human-in-the-loop review before validated transfer into the clinical database. By minimizing manual entry steps, this approach is expected to reduce transcription-related queries and improve cross-site, inter-reader concordance relative to manual workflows offering a scalable method to strengthen data integrity and audit-readiness in oncology trials.

DS_PPT_041: Enterprise Code Migration : An AI-Assisted approach to Legacy Modernization

Mirnalini Amirthalingam

ACL Digital

Migrating legacy applications from proprietary platforms to open-source technologies is essential for improving scalability, reducing licensing costs, and enabling AI-driven innovation. However, conventional code migration is often complex, time-consuming, and prone to functional inconsistencies. To address these challenges, we have developed an AI-assisted, in-house code migration solution. These efforts strengthen the migration ecosystem by combining semantic code understanding, static code analysis, intermediate representations, reusable transformation rules, large language models, and automated validation to generate functionally equivalent open-source code. A metadata-driven approach and reconciliation framework further ensure migration accuracy, consistency, and compliance. This collaborative approach reduces manual effort, accelerates modernization, improves code quality, and provides organizations with a scalable, reliable pathway for adopting open-source technologies.

DS_PPT_042: From Data to Decision: An Interactive R Shiny Platform for Lactation Pharmacokinetic Evaluation and Infant Risk Assessment

Dheeraj Gopu, Chiranjeevi Narasimharaju and Rachakonda Revanth Kumar

Bristol Myers Squibb

Evaluation of drug transfer into breast milk is a critical component of lactation safety assessments and regulatory submissions. The increasing adoption of Model-Informed Drug Development (MIDD) creates opportunities to build standardized, automated analytical frameworks that integrate lactation pharmacokinetic data with exposure-based risk assessment. A modular dashboard was built using R Shiny. Maternal plasma and breast milk data are imported to automatically calculate average and maximum milk concentrations, milk-to-plasma (M:P) ratios, Daily Infant Dose (DID), Relative Infant Dose (RID), and combined infant exposure from parent and metabolites. Subjects are classified by RID (<5%, 5–10%, ≥10%). Interactive plots summarize concentration–time profiles, M:P trends, RID distributions, and risk. A simulation module evaluates alternative milk concentrations, maternal dose, infant weight, and milk intake. This enabled automated lactation PK analysis, infant exposure estimation, simulation, risk stratification, and visualization, improving efficiency and reproducibility while supporting clinical pharmacology, pharmacometrics, and regulatory decision-making.

DS_PPT_043: From AI Output to Evidence: Trust and Governance in Clinical Research

Sudharsan Dhanavel and Amit Satra

MARS

Artificial Intelligence is becoming an integral part of clinical trial operations, supporting activities such as drafting site queries, adverse event coding, protocol deviation review, and study documentation. While these capabilities improve efficiency, they also raise an important question: how can organizations trust AI-generated outputs in regulated clinical research? This presentation explores responsible AI from the perspectives of clinical data management and biostatistics. It discusses why AI-generated content must be supported by human oversight, traceable evidence, and governance to ensure alignment with study protocols, source data, and regulatory expectations.

The session will highlight common risks, including hallucinations, outdated context, and lack of transparency, and present practical approaches for evaluating AI-generated outputs within existing quality processes. Attendees will leave with actionable recommendations for adopting AI responsibly while maintaining data integrity, regulatory compliance, and confidence in clinical decision-making.



EXPERTISE LEADS | INTELLIGENCE AMPLIFIES

One accountable **PARTNER** across every stage of your program

About SENSAN

Grounded in Science | Powered by Intelligence.

SENSAN started with a strong foundation in Biostatistics, Clinical Data, and Statistical Programming—and that scientific discipline remains at the heart of who we are. We have evolved beyond traditional delivery models to bring together clinical expertise, data sciences, consulting, and intelligent solutions under one ecosystem.

SENSAN's focus is simple: turn complex life sciences data into clearer insights, faster decisions, and greater impact.

SENSAN's vision is to transform clinical data review from **Reactive** processes into **Proactive** intelligence.



Our practice is built for your Business

Clinical & Regulatory Strategy

Biostatistics & Data Sciences

Digital Health & Real-World Evidence

Quality Compliance (GxP)

Talent & Resourcing Solutions

Agentic Intelligence & Innovation

STRATEGIC ADVISORY

Advisory Engagements

Project-based strategic consulting, scoped to a specific decision or milestone.

Retainer Advisory

Ongoing strategic support on call — not tied to any single project.

Outcomes-Based Partnerships

Engagement structured around milestones or results, not hours logged.

DELIVERY & STAFFING

Embedded Teams

SENSAN specialists placed directly inside your existing team and systems.

Managed Programs

We own a workstream end to end, against agreed timelines and deliverables.

On-Demand Talent

Vetted specialists you can scale up or down as the program shifts.



Contact us today!

609-955-8842



USA | INDIA | EUROPE



www.sensanbio.com



contact@sensanbio.com



Management Talks

Ind_MGT_PPT_002: Human + AI Leadership in Clinical Data Science

Vijayakumar Radhakrishnan

Precision For Medicine

The rapid advancement of artificial intelligence (AI) is transforming clinical data science, creating new opportunities to improve efficiency, accuracy, and innovation. However, the true value lies in effectively integrating AI with human expertise. This session explores the concept of Human + AI leadership and its role in shaping the future of clinical development.

Clinical data science is at an inflection point, where traditional statistical programming, data management, and biostatistics are increasingly augmented by machine learning, generative AI, and automation. While AI enhances data processing and predictive insights, human leadership remains critical for contextual interpretation, ethical governance, and regulatory alignment.

Through practical use cases across the clinical trial lifecycle ranging from protocol optimization to AI-assisted submissions, this presentation highlights the evolving role of clinical data scientists and the need for hybrid skillsets. Attendees will gain insights into building future ready teams, balancing automation with human judgment, and implementing responsible AI to drive faster, more reliable clinical outcomes.

Ind_MGT_PPT_006: Long duration trials in a Fast-Paced Era: Playing a test match with T20 expectations – Opportunities & Challenges

Raghava Regulla

Janssen (Johnson & Johnson)

It has become quite common that studies spanning more than 4-5 years. Working in long duration studies brings in unique opportunities (more time for preparation, evolving analysis needs, reappraisal of past decisions made) and challenges (data availability, on/off boarding of stakeholders during the study course, continuity of historical knowledge, ensuring quality & compliance over a long period, overlapping assignments/priorities etc).

Success requires the implementation of adaptive leadership strategies, leveraging waterfall, agile, and hybrid methodologies to effectively manage active and passive programming cycles, establishment of a Decision Register, structured backlog management, and periodic SWOT analyses to drive strategic planning, proactive risk management, and operational excellence.

This presentation explores how programming leads, as stewards of analysis, can sustain a long-term vision while delivering short-term milestones throughout the study lifecycle. It presents some practical strategies for the successful planning and execution of long-duration clinical trials.

Ind_MGT_PPT_007: Compliance-First AI Framework: Prompt-as-a-Package

Sarita Singh

Veramed

Background: Generative AI has the potential to transform clinical programming by accelerating code development and improving operational efficiency. However, the probabilistic nature of large language models introduces challenges related to reproducibility, traceability, auditability, and regulatory compliance. As organizations prepare for emerging Good AI Practice (GAIP) expectations and evolving regulatory frameworks such as the EU AI Act and GDPR, practical implementation models are needed to ensure AI-assisted processes remain trustworthy, auditable, and fit for use in regulated environments.

Approach: This presentation introduces a Compliance-First AI Framework centered on the concept of “Prompt-as-a-Package.” Rather than treating prompts as informal interactions, prompts are managed as version-controlled assets within the software development lifecycle. The framework translates GAIP principles into operational controls through structured YAML prompt templates, version management, automated validation gates, comprehensive audit logging, and human-in-the-loop oversight. AI-generated code is treated as a draft artifact that must successfully pass predefined deterministic test suites before progression into regulated workflows.

Expected Impact: A proof of concept demonstrates how prompt standardization, automated validation, and traceability can create an audit-ready AI workflow for ADaM programming activities. The proposed framework enables scalable AI adoption while preserving the quality. By operationalizing GAIP through governed automation, organizations can transition from experimental AI usage to sustainable and inspection-ready implementation.

Ind_MGT_PPT_010:From IQ to QI: The Next Evolution of Quality Delivery

Sugumaran Muthuraj

Precision For Medicine

The pharmaceutical and CRO industry has long relied on technical expertise, robust processes, and established quality frameworks to drive successful outcomes. As workplaces become increasingly complex, however, delivery excellence is being shaped by factors that extend beyond individual competence alone. Organizations have progressively embraced the importance of Intelligence Quotient (IQ), Emotional Quotient (EQ), and more recently Adaptability Quotient (AQ) to navigate rapid change. Yet an often-overlooked dimension remains critical: the collective psychology of teams. The ability to speak up, challenge assumptions, adapt to uncertainty, share accountability, and make sound decisions under pressure may be as influential to quality outcomes as technical capability itself. This presentation introduces Quality Intelligence (QI), a fresh perspective that brings together IQ, EQ, AQ, and Team Quotient (TQ) to explore how psychological awareness, collective adaptability, and team dynamics shape delivery excellence. By examining quality through a human-centered lens, the session offers a thought-provoking framework for understanding why equally capable teams can produce remarkably different outcomes, and what may truly distinguish high-performing organizations in the future. Perhaps the next frontier of quality is not Process Intelligence, but People Intelligence.

Ind_MGT_PPT_013: Beyond Delegation: Creating Accountability for Quality Excellence

Helen Jacob

ICON Clinical Research India Pvt. Ltd.

Accountability is a critical enabler of quality and delivery excellence within Statistical Programming. This session explores how clear ownership, structured governance, performance metrics, and continuous feedback can improve delivery predictability, reduce rework, and strengthen stakeholder confidence. Using real-world programming scenarios, the presentation demonstrates how accountability helps teams move from reactive issue management to proactive execution while maintaining high standards of quality and compliance. Attendees will learn practical approaches to building a culture of ownership that supports consistent, high-quality delivery and sustainable business outcomes.

Ind_MGT_PPT_014: From Individual Contributor to Force Multiplier: Accelerating Career Progression in the AI Era

Nishanth Nalan and Sakthi Thiruchengottaian

ACL Digital

Career growth in Biometrics has traditionally been anchored in technical expertise, domain knowledge, and years of experience. However, as AI increasingly absorbs routine work across Statistical Programming, Data Management, Validation and Reporting, technical proficiency alone does not suffice as a differentiator for career advancement. This shift raises an important question on how do some professionals accelerate while others despite comparable technical skills, plateau.

The answer lies in a distinct professional archetype, it is the Force Multiplier which amplifies not only the personal and professional capabilities of oneself but the capabilities of people, improves processes, accelerates innovation and delivers greater business impact. They move beyond completing tasks to enabling teams, mentoring colleagues, solving complex problems, embracing AI responsibly and driving meaningful change across the organization.

As technology evolves, careers are increasingly defined not by the work individuals accomplish themselves, but by the value they multiply for others. A force multiplier mindset expands influence because influence stems from the value created for people and organizations. While career growth is often measured by one's role the force multiplication is measured by one's impact.

A force multiplier mindset gives you more influence. Influence comes from the value you create for others. A Force Multiplier should not be a job level. It's a mindset and capability that can exist at every career level.

The future of biometrics will be shaped not only by experts but by Force Multipliers. The highest level of career progression is becoming a Force Multiplier someone whose impact continues to grow through the success of others.

Career aggression gets you noticed. Career progression gets you promoted.

Becoming a force multiplier makes you indispensable.

Ind_MGT_PPT_015: Problem First, AI Second: A Practical Framework for Value-Driven Innovation

Tanishq Lodha

Pfizer

Artificial intelligence offers significant opportunities to improve how organizations work, but successful implementation requires more than selecting the right technology. It begins with clearly understanding the business problem, identifying where AI can create meaningful value, and prioritizing opportunities based on impact, feasibility, and user needs. This presentation introduces a practical approach to managing AI and innovation projects, covering problem definition, opportunity prioritization, solution design, stakeholder alignment, and value articulation. It also highlights the importance of testing solutions with users, measuring outcomes, and adapting the approach as new insights emerge. Through practical examples, the presentation demonstrates how project teams can move from broad ideas to focused, implementable solutions while avoiding technology-first decision-making. The approach enables organizations to invest in the right opportunities, manage uncertainty, and deliver solutions that address real business needs and generate measurable value.

Ind_MGT_PPT_017: Leadership in the age of automation – The Human art

Vijay Dharmaraj

ACL Digital

In an era defined by automation, machine learning, and intelligent systems, the irreplaceable quality of human leadership stands in sharper relief than ever before. Leadership is not a formula that can be encoded, nor a strategy that can be scaled through algorithms, it is a profoundly human practice, shaped by empathy, adaptability, humility, and the courage to make hard decisions.

Effective leaders read the room and adapt their style; they praise publicly and correct privately, preserving dignity while driving growth. They hold three truths at once: humble yet assertive, open-minded yet decisive, unbiased yet emotionally attuned.

AI can widen a leader's reach, surfacing patterns, drafting first passes, clearing administrative load. But it can only support the practice of leadership, it cannot originate the trust, meaning, and judgment that define it. Now more than ever leadership is going to play an important role with changing landscapes.

inference

BIOMETRICS AT ITS BEST

EXPERIENCE. TRUST. SOLUTIONS

Delivering full-spectrum biometrics support—from trial design through BLA and NDA submissions. Trusted by sponsors for our DMC expertise and regulatory-ready analyses that drive confident, timely decisions.

Our Services

- Biostatistics
- Trial Design Optimization
- Clinical Data Management
- Statistical Programming
- Epidemiology
- Medical Writing



For all of our services and more information on Inference scan this QR code

Industry Posters

Ind_BIOS_PSTR_001: Transforming Adverse Event Data into Comparative Safety Insights

Harshita Kushwaha

AstraZeneca

Background: Traditional clinical safety reviews rely heavily on extensive tabular outputs, making comparative interpretation of adverse event (AE) profiles challenging

Objective: To demonstrate how graphical visualization techniques can improve comparative safety assessment and interpretation of AE data

Methods: A simulated renal–metabolism clinical trial dataset comprising 300 randomized patients and 1,159 AE episodes was analyzed. Visualizations were developed to evaluate treatment differences across AE frequency, severity, patient burden, timing, duration, and clinical impact

Results: Graphical approaches enabled rapid identification of treatment-specific safety patterns, including differences in AE frequency, severity distributions, patient-level burden, temporal occurrence, and high-priority safety signals. Multiple complementary visualizations provided a more comprehensive understanding of comparative safety profiles than traditional tabular summaries alone

Conclusions: Visual analytics transformed complex AE data into intuitive comparative safety insights, facilitating faster interpretation, improved communication of safety findings, and more efficient comparative safety assessment.

Ind_BIOS_PSTR_003: Can You Trust Your Survival Analysis? A Practical Guide to Checking Non-Informative Censoring and Proportional Hazards Assumptions in Clinical Trials

Rahul Patil

Phastar

Survival analysis methods such as Kaplan–Meier estimation and Cox proportional hazards model are widely used in clinical trials to evaluate time-to-event endpoints. However, the validity of their results depends on key underlying assumptions that are not always routinely assessed. Failure to evaluate these assumptions may lead to biased estimates and misleading clinical conclusions.

This poster focuses on two critical assumptions: non-informative censoring and proportional hazards. Using a practical, application-oriented framework, the poster explains why these assumptions matter, common approaches for assessing them, and their implications when violated. Diagnostic methods including reverse Kaplan–Meier curves, log-minus-log plots, Schoenfeld residuals, and formal proportional hazards tests are presented alongside their implementation in SAS. The poster also provides guidance on interpreting diagnostic results and selecting appropriate actions when assumptions are not met. The aim is to equip clinical trial statisticians with a practical checklist for routinely assessing and documenting survival analysis assumptions before drawing conclusions.

Ind_BIOS_PSTR_004: SPATIAL MODELING OF SELF-REPORTED THYROID DISORDER AMONG WOMEN IN INDIA - A SECONDARY DATA ANALYSIS ON NFHS-5 DATA

Ramya N, Maria Abraham and K.T. Harichandrakumar

Emmes services Pvt. Ltd.

INTRODUCTION: Thyroid disorders represent a substantial public health challenge in India, affecting 2.7% among women aged 15-49 years. Detailed spatial analysis of thyroid disorders among reproductive-aged women is scanty. This study develops and compares the accuracy of Spatial Lag (SL), Spatial Error (SE), and Ordinary Least Squares (OLS) models using National Family and Health Survey (NFHS-5) data.

METHODS: A cross-sectional study using NFHS-5 (2019–2021) data assessed spatial dependence of self-reported thyroid disorders through Moran's I, LISA, and bivariate LISA to identify predictors. OLS, SL and SE models were then fit and compared using log-likelihood, Akaike Information Criterion, and Schwartz Criterion.

RESULTS: Significant spatial autocorrelation was observed (Moran's I: 0.73; $P < 0.05$). The SE Model was identified as the most appropriate model.

CONCLUSION: Significant spatial clustering of self-reported thyroid disorders was observed across India, highlighting the need for geographically targeted interventions to reduce regional disparities and improve thyroid health outcomes among women.

Ind_BIOS_PSTR_005: Synthetic Data in Clinical Research: Evaluating the Consistency of Analytical Conclusions

Arunima Dass

Emmes services Pvt. Ltd.

Background: Clinical research often relies on small datasets, limiting the analysis that can be performed and making validation of findings challenging. Synthetic data have emerged as a potential approach to maximize the utility of available data, but their ability to reproduce conclusions from original analyses requires evaluation.

Method: Using a clinical dataset, synthetic datasets were generated through sequential conditional synthesis with Classification and Regression Tree (CART) models. Logistic regression models were fitted to both the original and synthetic datasets, and coefficient estimates and overall patterns of association were compared.

Results: Association between the variables remained in the same direction, although some differences in effect size were observed. Despite these differences, the overall conclusions were consistent across datasets.

Conclusion: These findings suggests that synthetic data can preserve key analytical results and may be a useful tool when working with small clinical datasets.

Ind_BIOS_PSTR_007: Bayesian Response Adaptive Randomization in Sequential Trials

Bharath A and Ashma Dorothy Monteiro

Manipal Academy of Higher Education (MAHE)

Adaptive clinical trial designs using Bayesian Response Adaptive Randomization (BRAR) provide a flexible framework for modifying trial procedures as data accumulates. However, unrestricted randomization tends to reduce the probability of correctly identifying the superior treatment, and constrained designs combining hard allocation bounds with sequential monitoring lack systematic evaluation. This study addresses that gap through a Monte Carlo simulation evaluating three constrained BRAR designs across flat, Jeffreys, and conjugate priors under small, medium, and large effect sizes for a binary outcome over 1000 replications using the BATSS package in R. Across all configurations, AR($n/2N$, 0.1) achieved the highest power, peaking at 0.84 under the Jeffreys prior at the large effect size, while recording the lowest average sample number in every case. AR(1, 0.1) directed the greatest proportion of patients to the superior arm but yielded the lowest power, and all configurations maintained a type I error rate of exactly 0.05 through binary search calibration.

Ind_BIOS_PSTR_009: PFS vs RFS: Similar Methods, Different Questions

Jennifer Lobo

Others

Selecting an appropriate time-to-event endpoint is fundamental to the design and interpretation of oncology clinical trials. Although Progression-Free Survival (PFS) and Recurrence-Free Survival (RFS) are both analyzed using standard survival analysis methods,

they are applied in different clinical settings and answer different research questions. This educational poster provides a practical, high-level comparison of PFS and RFS, focusing on their intended use, event definitions, baseline disease status, tumour assessment approaches, and key censoring considerations. The poster aims to simplify these concepts through clear comparisons and visual summaries that are accessible to statisticians, programmers, and other clinical research professionals who are new to oncology. By highlighting the key differences between these commonly used endpoints, the poster seeks to support informed endpoint selection and improve the interpretation of oncology trial results in everyday clinical research practice.

Ind_BIOS_PSTR_010: Separating Signal from Heart Rate Noise: Advanced QT Correction for Cardiac Safety Analysis

Abhinav Srivastava

Bristol Myers Squibb

Accurate QT interval correction is essential in thorough QT studies to assess drug-induced effects on cardiac repolarization. When a drug causes meaningful heart rate changes, conventional correction methods such as Bazett's and Fridericia's formulas may inadequately characterize the QT-RR relationship, leading to overcorrection, under correction, or misinterpretation of cardiac safety signals. This presentation illustrates alternative QT correction approaches within the ICH E14 framework when substantial ($\Delta\Delta\text{HR} > 10$ bpm) heart rate effects are observed. Individualized correction methods, including QTcS and QTcI, were evaluated using subject-specific QT-RR relationships derived from baseline 24-hour Holter ECG data. Correction adequacy was assessed through residual QTc-RR dependence analyses, and treatment effects were evaluated using linear mixed-effects and concentration-QTc models. Simulated data were used to demonstrate method application. Overall, individualized QT correction supported by adequacy assessment provides a robust framework for interpreting QT/QTc effects when drug-induced heart rate changes are present.

Keywords: QT interval, correction, Heart rate, cardiac repolarization.

IND_BIOS_PPT_079: Agentic AI and Bayesian Synthetic Control Sites for Next-Generation RBQM

Mahesh Divakaran

Epicacy Lifescience Analytics

Central Statistical Monitoring frequently makes comparisons of KRI values from individual sites against naïve average study-wide values which may wrongly label the sites for discrepancies arising from the case mix differences rather than the quality issues. As a solution, a Bayesian hierarchical modeling approach is proposed which would generate case mix adjusted expected profiles for KRI values from individual sites making them the synthetic controls for better comparisons. The disease severity, demographic covariates, and time of recruitment will be accounted for in the modeling. Further extending this approach, a multi-agent large language model (LLM) system monitors the performance of individual sites and performs automated risk assessment by employing a

detector agent, a classifier agent, and a narrator agent for deviation detection, classification, and natural language generation of evidence-based risks respectively. Evaluation is performed on a simulated multicenter oncology trial data set.

IND_PSTR_SP_005: Smarter Stats, Smarter Antibodies, Safer Vaccines

Trina Sen, Srinivas Vankadaru

Sanofi Healthcare

Background: The exponential growth of genomic, immunological, and clinical trial data has transformed vaccine research, requiring statistical programmers to adopt AI-driven methodologies alongside traditional biostatistical rigor.

Methods: This presentation examines how statistical programmers leverage machine learning algorithms, multi-language pipelines (SAS, R, Python), and cloud-based infrastructure to process high-dimensional immunogenicity data, automate safety signal detection, and support adaptive clinical trial designs. Validation frameworks ensuring regulatory compliance in GxP environments are highlighted.

Results: Case studies from respiratory vaccine programs (COVID-19, RSV, influenza) demonstrate reduced sample sizes through predictive immunogenicity modelling, accelerated variant surveillance via genomic analysis pipelines, and enhanced pharmacovigilance through NLP-based case processing.

Conclusions: Statistical programmers must evolve into hybrid biostatistician-data scientists, mastering AI/ML tools while maintaining regulatory rigor to accelerate development of life-saving vaccines against respiratory pathogens.

IND_PSTR_SP_006: Bridging Estimand Strategies and ADaM Implementation: A Standardized Approach for Handling Intercurrent Events

Vaishnavi Dhara

Novartis Healthcare Pvt. Ltd.

Implementing ICH E9(R1) estimands requires explicit, analysis-ready handling of intercurrent events (ICEs). The proposed ADaM framework uses a dedicated ADICE dataset to capture each subject's ICEs and estimand-specific handling strategies. ADICE links protocol/SAP definitions to ADEFF/ADEFFM derivations and analysis outputs, improving traceability, consistency, and interpretability across primary and supplementary estimands

IND_PSTR_SP_007: Digital Twins and Dynamic Reporting: Hepatic Impairment Pharmacokinetic Assessments in R

Francis Kevin Raj S, Susmitha Patel Thota

Pfizer

Evaluating hepatic impairment on drug exposure usually involves comparing clinical data against simulated healthy cohorts (*in silico* controls). However, the current methods for creating these simulated groups are mostly manual and slow. To fix this, we created an automated system in R. It uses past pharmacokinetic (PK) data and population models to generate virtual healthy patients, which we call "Digital Twins."

Using Quarto in R, our tool seamlessly combines the simulated healthy data with real clinical trial data from patients with hepatic impairment. The system automatically processes the key PK measurements (C_{max} and AUC) by applying a log-transformation and running an ANOVA. It calculates the necessary geometric mean ratios and 90% confidence intervals, directly outputting clean, ready-to-use tables.

By connecting the entire process, from simulation to statistical analysis into one automated workflow, we cut out manual data transfers and save a significant amount of reporting time. The system also includes interactive charts, which allows to explore the data visually. Ultimately, this R-based solution makes PK evaluations faster, more reliable, and modernizes clinical data storytelling.

IND_PSTR_SP_008: Validating Clinical Tables Using ARDs

Dinesh Bhamare, Dinesh Bhamare, Linga Reddy Aenugu

Syneos

Quality control (QC) of clinical reporting tables is important to ensure that reported results are accurate, consistent, and meet regulatory requirements. As reporting tools such as {gtsummary}, {crane}, and {citril} create increasingly complex tables, manual QC becomes time-consuming, prone to errors, and difficult to repeat consistently.

An automated QC approach using Analysis Result Datasets (ARDs), a standard dataset format for storing analysis results such as summary statistics, counts, and percentages, provides a more efficient and transparent way to validate these tables. In this workflow, results from the final tables are converted into ARDs and compared with ARDs independently created from the source data.

Datasets are matched using common elements such as treatment groups, variables, and statistic names, then compared row by row with the {difdf} package to detect differences.

This provides a reliable, reproducible, and scalable validation approach, reducing manual effort while increasing confidence in clinical reporting outputs.

IND_PSTR_SP_010: Handling Multiple Baselines in ADaM: A Practical Approach

Hemanth Kumar Lingala, Vikasgoud Rangol

Oryixon

Multiple baseline selection is a recurring challenge in ADaM implementation, especially in crossover, re-baselining after washout, treatment re-initiation, and rescue/add-on therapy studies. Inconsistent baseline handling can affect change-from-baseline analyses, interpretation, and regulatory review.

This presentation describes a rule-based, sponsor-agnostic approach for defining and implementing multiple baselines using CDISC-standard ADaM conventions, particularly BASETYPE, ABLFL, BASE, CHG, and PCHG. Practical scenarios are mapped to relevant estimand strategies and translated into reproducible SAS programming logic.

A 2-period crossover example demonstrates how baseline candidates are identified, selected, propagated, and validated within BDS. Key programming and QC considerations are discussed to support transparent, traceable, and submission-ready ADaM datasets.

IND_PSTR_SP_012: One Macro to Rule Them All: A Metadata-Driven SAS® Framework for Dynamic ADaM Unit Conversion

Naveen Prasath

Sensan Biosciences

In clinical trial reporting, the Analysis Data Model (ADaM) demands rigorous unit standardization for laboratory (ADLB) and vital signs (ADVS) domains. Conventionally, programmers embed conversion factors directly inside DATA step logic, an approach that is inefficient, error-prone, and difficult to maintain as sponsor specifications or standard dictionaries evolve. This paper introduces a metadata-driven SAS® macro framework that externalizes unit-conversion rules from the codebase and applies them dynamically at runtime using SAS Hash objects.

Beyond simple linear scaling, such as pounds to kilograms, the framework is extended to resolve non-linear, formula-based conversions, such as Fahrenheit-to-Celsius temperature conversion, by combining CALL SYMPUTX and the RESOLVE function to evaluate arbitrary arithmetic expressions stored as metadata text at runtime. This dual-mode design allows a single reusable macro to handle both constant multipliers and algebraic formulas without any change to the underlying code, improving data accuracy, simplifying QC, and reinforcing adherence to CDISC standards.

IND_PSTR_SP_015: eDISH (Evaluation of Drug-Induced Serious Hepatotoxicity): A Visual Approach to Liver Safety Assessment in Clinical Trials.

PAVANI DAYANIDHI, Mamatha Panchika, venkatesh Pinaboina

Panakeia

Drug-Induced Liver Injury (DILI) is one of the biggest concerns in clinical drug development and needs to be identified in time in order to ensure patient safety. eDISH (Evaluation of Drug-Induced Serious Hepatotoxicity) is an interactive graphic which evaluates the correlation between liver enzyme values and total bilirubin, and can help identify subjects who may need further investigation.

This presentation shows how eDISH can be used for liver safety evaluation, providing explanation on how to interpret eDISH's four quadrants and how Hy's Law can be helpful. It also illustrates how clinical programmers can contribute to eDISH preparation by creating CDISC compliant SDTM/ADaM datasets and generating eDISH outputs. Using eDISH helps to quickly visualize liver laboratory parameters and detect hepatotoxicity. A good eDISH process helps to perform liver safety evaluations and deliver necessary information for regulatory submission.

IND_PSTR_SP_017: Clinical Review Dashboard Using R Shiny

NIKHIL KASHABOINA, Shiva Vishwanatham, Praneeth Reddy Dasari

Panakeia

Traditional approaches to clinical data review involve preparing the same tables, listings, and graphs every time there is a change in the dataset. If the reviewer or the data manager wants another way to look at the data, more ad hoc reports are requested, resulting in delays and extra programming work.

This talk will introduce an easy-to-use interactive dashboard implemented using R Shiny which allows reviewing SDTM and ADaM datasets in real-time. The user can filter data based on domains, subjects, visits, or laboratory parameters without any additional report requests.

It includes visualization of lab parameters over time, adverse events, and individual subject data. It is connected to the CDISC datasets directly which makes it easier to spot the strange data even for non-programmers. This talk will show how to implement such an application, including scheduling data refreshment and optimizing it for large datasets.

IND_PSTR_SP_018: Decision-Driven Clinical Programming: From Protocol to Submission

Aparna Kurella, Kameswari Pindiprolu

Oryxion

Regulatory submissions rarely fail because of programming errors alone; they fail when critical decisions remain undocumented or unexamined. Between specification and submission lies a series of critical decisions that determine a study's regulatory fate. This poster highlights a decision-focused framework for evaluating programming activities, enabling programmers to identify high-impact risks before they affect regulatory submissions.

A phase-wise decision framework spanning the programming lifecycle - from protocol review through submission -with each checkpoint evaluated against likelihood and impact using a risk-based matrix. Errors feared most, such as incorrect domain mapping, proved rare but high-impact, while issues causing actual regulatory delay, such as missing Reviewer's Guides or incomplete traceability, were common and largely preventable. A representative case demonstrates how a single undocumented cross-reference extended the regulatory review timeline.

Standards-aligned, gate-based decision-making improves traceability, reduces rework, and strengthens submission readiness, providing a practical framework for consistent, inspection-ready clinical programming across studies and sponsors

IND_PSTR_SP_019: Modernizing Programming Best Practices for Regulatory Submissions: Balancing Efficiency with Compliance

Gali Srilatha

Sensan Biosciences

Achieving successful regulatory submissions requires statistical programming that is accurate, compliant, scalable, and efficient. As clinical trials become increasingly complex with adaptive study designs, decentralized trials, expanding data volumes, and evolving regulatory expectations, traditional programming approaches are no longer sufficient.

This presentation explores practical strategies for modernizing statistical programming throughout the regulatory submission lifecycle while maintaining compliance.

Key approaches include metadata-driven programming, reusable programming frameworks, standardized macros, version control, risk-based quality control, automation of routine validation, and comprehensive documentation to improve traceability and audit readiness.

Real-world examples demonstrate how these practices reduce manual effort, minimize programming errors, improve code reusability, strengthen collaboration, and accelerate submission readiness.

Emerging trends, including AI-assisted code development and intelligent validation support, are also discussed as opportunities to enhance efficiency while ensuring appropriate human oversight.

Attendees will gain practical strategies to improve productivity, reproducibility, data integrity, and deliver high-quality regulatory submissions aligned with global health authority expectations.

IND_PSTR_SP_020: CRAFT- A Metadata-Driven Fit & Shape Matrix automation for End-to-End Statistical programming

Ramya Sengodan, Soundarya Palanisamy

ACL Digital

Clinical data programming teams manually rebuild CRF annotations, SDTM datasets, ADaM datasets, and TLF outputs for every protocol or CRF revision, slowing turnaround and introducing mapping errors. We present a Fit & Shape Matrix: a reusable, metadata-driven template that maps source variables to the structure each downstream deliverable requires. Two pathways apply it: aCRF-to-SDTM, driven by CRF and source-dataset metadata, and ADaM-to-TLF, driven by SDTM metadata and SAP-defined logic. Mapped across the full protocol-to-CSR pipeline, the framework removes revision-by-revision manual rebuild at both handoffs. Process mapping surfaced recurring challenges, including manual PDF annotation, inconsistent source imports, and validation gaps beyond Pinnacle 21. Applying the matrix template shortens CRF-to-SDTM turnaround, improve mapping consistency, make templates reusable across studies, and strengthen the audit trail. We share the process map, the two automation pathways, and next steps for piloting the template on an active study.

IND_PSTR_SP_021: Automated Dataset Validation and Accuracy Analyzer

Samyuktha Jagadeesawaran

ACL Digital

Accurate dataset validation is essential for ensuring high-quality clinical programming deliverables. Conventional comparison methods rely on identical observation counts and record alignment, making them ineffective when datasets contain missing, additional, or unmatched records. This often results in extensive manual reconciliation and increased quality control effort.

The Automated Dataset Validation and Accuracy Analyzer is an R-based tool designed to perform comprehensive dataset comparison regardless of observation matching. For aligned datasets, it calculates variable-level accuracy, identifies discrepancies, generates mismatch reports. For unmatched datasets, it performs subject-level record count reconciliation, variable value frequency comparison, duplicate detection, and detailed mismatch reporting,

enabling meaningful validation even for repeated-measure domains such as Laboratory and Vital Signs.

By combining accuracy assessment with intelligent reconciliation, the framework streamlines quality control, reduces manual effort, accelerates discrepancy analysis, and provides a scalable solution for robust clinical dataset validation.

Ind_DS_PSTR_001: From Detection to Prediction: An AI-Augmented Screening Dashboard for Preventing Inadvertent Enrollment

Violina Kalika

Eli Lilly

Inadvertent Enrollment (IE) - the randomization of participants with ineligible or incomplete eligibility data — poses a significant risk to trial integrity, participant safety, and data quality. We designed and implemented a comprehensive Screening Dashboard integrating multiple operational data sources to surface eligibility risk in near real-time. A Clinical Data Representative-led review model, supported by a centralized backend tracker, enabled structured triage of screening anomalies across Data Management, Medical, Clinical Development, and Monitoring functions. Defined query timelines, status-based actions, and a tiered escalation framework drove early detection and rapid management of potential IEs. Building on this foundation, I propose an AI layer: a predictive IE risk-scoring model, unsupervised anomaly detection to flag outlier sites objectively, and NLP-assisted triage of the backend tracker — shifting the process from detection to prediction. This presentation shares the implemented framework, measurable outcomes, and a human-in-the-loop roadmap toward predictive, risk-based eligibility management.

Ind_DS_PSTR_002: Bidirectional Product Complaint Reconciliation in Device Trials Using Power BI

Srilakshmi R

Eli Lilly

Background: In device trials, the Global Product Complaint (PC) listing is reconciled against the adverse event (AE) dataset in Electronic Data Capture (EDC) across multiple rules. The check most teams miss is the reverse - a complaint in the PC listing where the AE was never marked device-related in EDC.

Objective: To describe a Power BI dashboard automating bidirectional PC reconciliation using the Unique Tracking Number (UTN) - encoding seriousness (S or N) and AE sequence number — as the matching key.

Approach: Connecting both sources in Power BI decodes UTN - running forward, reverse, and Serious Adverse Event (SAE) checks. Without automated flagging records go undetected. Transfers to systems with a common structured field.

Conclusion: Bidirectional reconciliation removes the reverse gap most teams miss. A complaint never captured in EDC stays invisible until someone digs for it. A common field drives the matching - no EDC change, no new tools.

Ind_DS_PSTR_003: Transforming Clinical Programming with LLMs: Automated Code Review

Deepika S

Pfizer

The integration of AI agents into statistical programming workflows is transforming traditional processes such as code review, documentation, and quality assurance. Traditional manual reviews of R programs are time-intensive and subject to reviewer variability, particularly when assessing compliance with organizational coding standards, runtime efficiency, and code maintainability. This work presents a Streamlit-based AI agent that automates R code review and optimization using large language models accessed through GitHub Models. The agent accepts R code via paste, file upload, SFTP or interactive chat, and evaluates it against predefined coding standards and best practices through structured prompt engineering. It identifies inefficiencies, recommends performance improvements, enforces consistent use of assignment operators, pipe conventions, NA handling, explicit package referencing, and naming standards, and generates complete refactored code with inline annotations.

The agent delivers actionable, developer-friendly feedback with severity-graded suggestions while preserving the original code's logical functionality and consistency of R code review processes.

Ind_DS_PSTR_004: Trust, but Verify: A Practitioner's Experience Using a Large Language Model in Clinical Data Management

Saachi Mehta

Eli Lilly

Discussion of AI in clinical data management (CDM) is dominated by optimism and vendor demonstrations, with few honest accounts of what happens when a large language model (LLM) is applied to everyday tasks. This presentation is a practitioner's field report on using an LLM across real CDM activities, with equal attention to where it added value and where it fell short.

It covers both operational tasks, such as team tooling and cross-functional communication, and data-focused tasks, such as reconciliation, discrepancy detection, and consistency checks. In one coding check across 10,871 verbatim terms, the model flagged 9 potential verbatim-versus-decode mismatches; on review, only 1 was genuine and 8 were false flags requiring correction. Worked examples show how an LLM (Claude & Microsoft Copilot) can be confidently wrong

when it misses a domain relationship, and how output quality depends on prompt specificity. The core message: LLMs can accelerate CDM work, but human judgment remains essential.

Ind_DS_PSTR_005: Beyond Prompt Engineering: AI-Assisted Analysis Metadata Authoring with CDISC ARS

Surendar Sekar

ICON Clinical Research India Pvt. Ltd.

The rapid adoption of Generative AI in clinical programming has accelerated code generation but introduced variability, limited traceability, and inconsistent interpretation of Statistical Analysis Plans (SAPs). This presentation proposes a metadata-first approach in which an AI agent transforms SAP requirements into CDISC Analysis Results Standard (ARS)-compliant metadata using standardized AnalysisSet, GroupingFactor, AnalysisMethod, and ReportingEvent definitions from ARS-v1.

Rather than generating analysis code directly, the AI assists in authoring and validating machine-readable analysis metadata - demonstrated through representative SAP requirements, including an adversarial case correctly flagged by the validator. The resulting ARS metadata establishes end-to-end traceability from protocol and SAP through analysis implementation, enabling consistent, explainable, and reusable study reporting.

By positioning ARS as the governance layer between human-defined science and AI-enabled automation, this framework offers a foundation for reproducible analyses, metadata reuse, and agent-driven clinical reporting under oversight.

This vision shifts from AI-coding assistance to governed standard execution in clinical-development.

Ind_DS_PSTR_007: Healthcare Data Analytics and Wearable Devices: Transforming Modern Healthcare

Tamilselvi Narayanasamy and Jayadevi Jeevanantham

Sensan Biosciences

Wearable devices are revolutionizing modern healthcare by enabling continuous monitoring of patients through smartwatches, fitness trackers, wearable ECGs, and glucose monitoring systems. These technologies collect real-time health data such as heart rate, physical activity, sleep patterns, and vital signs, providing valuable insights into an individual's health status. Healthcare data analytics integrates artificial intelligence (AI), machine learning, and big data technologies to transform raw wearable data into actionable clinical insights. By analyzing continuous patient-generated data, healthcare providers can detect early signs of disease progression, monitor chronic conditions remotely, predict health risks, and improve treatment outcomes. Wearable technologies have become particularly valuable in cardiovascular disease

management, diabetes care, fitness monitoring, and post-treatment follow-up, enabling timely interventions and reducing hospital admissions.

This poster highlights recent advancements in wearable devices, explores the role of healthcare data analytics in improving clinical decision-making, and discusses future opportunities for personalized medicine and patient-centered healthcare.

Ind_DS_PSTR_010: Secure Data Exchange with Intelligent Validation

Samyuktha Jagadeeswaran

ACL Digital

Organizations have many secure file transfer solutions available today. However, validation of the transferred data remains a largely manual process. Teams often perform metadata checks, dataset comparisons, or custom validation scripts after files are received, increasing effort and delaying downstream activities

The proposed solution extends a Secure Data Exchange Platform that provide solution for secure file transfer, automated integrity verification, audit logging, monitoring, and reporting throughout the file exchange life cycle

The initial phase of the project focuses on automating the validation process that is typically performed manually after files are exchanged between organizations. Rather than replacing existing file transfer solutions, the platform integrates with them, provides automated verification and the other downstream processes.

Ind_DS_PSTR_012: QInsight - Empowering Clinical Programming through Interactive Data Visualization

Sugunesh Sivalingam

ACL Digital

Clinical trials generate large volumes of graphical outputs throughout the study lifecycle, making efficient access, review, and interpretation increasingly challenging. Conventional approaches rely on static reports and dispersed graph files, resulting in duplicated effort, delayed review cycles, and limited data exploration. To address these challenges, we developed an in-house Data Visualization Tool (DVT) that serves as a centralized graph repository for standardized clinical trial visualizations. The platform integrates SDTM and ADaM datasets to generate interactive dashboards that enable seamless access to patient profiles, safety trends, efficacy visualizations, and data quality metrics. Users can efficiently search, filter, compare, and interact with graphical outputs through a secure, user-friendly interface, supporting patient-level exploration and study-level trend analysis. By transforming isolated static graphs into an organized and interactive visualization repository, the tool enhances data accessibility, reduces manual effort, improves consistency in data review, and promotes collaboration across clinical programming, biostatistics, data management, and clinical teams. This solution demonstrates

how an integrated graph repository can streamline clinical data review, improve operational efficiency, and facilitate informed decision-making while supporting regulatory compliance and data transparency.

DS_PPT_004: Transforming Clinical Data Management with Next Generation Artificial Intelligence and Data Science

Vara Prasad Mukkara

EPDA

Clinical Data Management (CDM) plays a fundamental role in ensuring the integrity, quality, and reliability of clinical trial data. The rapid growth of digital healthcare technologies, decentralized clinical trials, wearable devices, and real-world data has significantly increased the volume and complexity of clinical data. Traditional CDM processes are often resource-intensive, time-consuming, and susceptible to human error. Next Generation Artificial Intelligence (AI), including Machine Learning (ML), Natural Language Processing (NLP), Agentic AI, and Predictive Analytics, is transforming conventional data management into an intelligent, automated, and risk-based ecosystem.

This paper explores how AI enhances every stage of the CDM lifecycle, including protocol review, electronic Case Report Form (eCRF) design, edit check generation, automated query management, system driven SAE triggers, medical coding assistance, data quality surveillance, database lock readiness, and regulatory compliance including customized reports. AI-driven technologies improve operational efficiency, reduce manual effort with accuracy, accelerate study timelines, and strengthen compliance with ICH-GCP, 21 CFR Part 11, CDASH, CDISC, and ALCOA+ principles. Although AI provides substantial benefits, human oversight remains essential for clinical judgment, ethical governance, and regulatory accountability. The paper concludes that Next Generation AI represents a paradigm shift in clinical data management and will become a key enabler of future digital clinical research.

Ind_MGT_PSTR_001:Corporate Stories That Shape Success Corporate Values That Create Careers

Damini Sharma

ACL Digital

Corporate stories are more than experiences. They are reflections of the values we live by every day. In today's dynamic world, technical skills may open doors, but it is our values that build trust, inspire people, and create lasting impact.

Positivity fuels resilience. Accountability builds credibility. Leadership empowers others. Time Management drives efficiency. Integrity earns trust. Communication connects ideas. Adaptability embraces change. Teamwork multiplies success.

Together, these values form the foundation of Professionalism and are the Stepping Stones to greater heights in our careers. When we embody these principles, we don't just grow individually but we elevate teams, organizations, and the world around us.

Let every story we create be a testament to the values we uphold and the legacy we aspire to leave behind.

"Coming together is a beginning; keeping together is progress; working together is success."

~ Henry Ford

The workplace is evolving rapidly with the growing presence of Generation Z and the emergence of Generation Alpha, bringing new perspectives, strengths, and expectations to organizational communication.

As digital natives, these generations contribute technological fluency, adaptability, collaborative thinking, creativity, and a strong inclination toward continuous learning, creating new opportunities for innovation and workforce engagement.

At the same time, organizations are navigating evolving communication challenges, including diverse communication preferences, maintaining meaningful engagement in hybrid work environments, balancing technology with authentic human interaction, fostering collaboration across diverse teams, and aligning communication with the expectations of an increasingly digital workforce.

As communication becomes a strategic driver of employee experience and organizational success, there is a growing need to understand how communication practices are evolving and how organizations can adapt to these changes.

This abstract explores the evolving communication landscape, examines the communication challenges associated with engaging Generation Z and Generation Alpha, and highlights modern workforce engagement strategies, the role of technology in enabling human-centred communication, approaches to building a culture of connection and collaboration, methods for bridging communication expectations across a digital workforce, and key considerations for preparing organizations for the future of work.

Together, these perspectives provide a holistic view of how organizations can strengthen communication in an increasingly connected and evolving workplace.

**MAKING MEDICINE THAT'S
NEVER BEEN MADE
MEANS DOING WHAT'S
NEVER BEEN DONE.**

LILLY.COM



STD_PSTR_001: AN EPIDEMIOLOGICAL STUDY ON MALNUTRITION AMONG SCHOOL GOING ADOLESCENTS OF BLOCK FARIDABAD, HARYANA

Dhairya Aggarwal

ESIC Medical College & Hospital, Ludhiana

Malnutrition remains a major public health concern among adolescents in India. This cross-sectional study assessed the burden and determinants of malnutrition among school-going adolescents in Block Faridabad, Haryana.

A total of 800 students from classes 6th–12th studying in government and private schools from urban and rural areas were selected using a two-stage sampling technique.

Data were collected using a validated structured questionnaire and analyzed using STATA 17.

The overall prevalence of malnutrition was 52.9%, comprising 45.3% underweight and 7.6% overweight/obesity. Underweight was more prevalent among females (49.2%) than males (42.9%), whereas overweight/obesity was higher among males (11.2%).

Significant associations were observed between malnutrition and age, physical activity, and meal-skipping habits ($p < 0.05$). Multivariable logistic regression identified age, gender, physical activity, and skipping meals as independent determinants of malnutrition.

The findings highlight the urgent need for targeted adolescent nutrition and lifestyle interventions.

STD_PSTR_002: Association between neutrophil-to-lymphocyte ratio and all-cause mortality in adults: A Systematic Review and Meta-Analysis

Harsh Kumar

IMS, BHU

Background: Elevated baseline neutrophil-to-lymphocyte ratio (NLR) has emerged as a potential inflammatory prognostic biomarker across chronic diseases.

Objective: To systematically evaluate the association between baseline NLR and all-cause mortality.

Methods: PRISMA 2020 compliant systematic review and REML random-effects meta-analysis of cohort studies.

Results: Twelve cohort studies involving 20,450 participants and 4,555 mortality events were included. Elevated baseline NLR was significantly associated with increased mortality risk ($HR = 1.49 [1.32–1.68]$, $p < 0.001$).

Conclusion: Baseline NLR is an independent and clinically accessible prognostic marker for mortality risk stratification.

STD_PSTR_003: Association between neutrophil-to-lymphocyte ratio and all-cause mortality in adults: A Systematic Review and Meta-Analysis

Harsh Kumar

IMS, BHU

Climate change and rising temperatures pose growing risks not only to physical health but also to mental well-being, particularly in urban settings of low- and middle-income regions. This study explores the association between self-reported heat exposure and mental health outcomes among young adults (18–35 years) in urban areas of India and sub-Saharan Africa. Using a cross-sectional survey and standardized mental health measures, the study assesses how perceived heat exposure relates to levels of depression, anxiety, and stress. The findings indicate a clear and consistent positive association between higher perceived heat exposure and poorer mental health outcomes. Young adults who reported experiencing more intense heat also tended to report higher levels of psychological distress across all three dimensions. Additionally, notable differences emerged between study locations, suggesting that contextual and environmental factors may influence the extent of these impacts.

Overall, the study highlights that heat exposure is not only an environmental or physical health concern but also a significant psychological stressor. These findings underscore the need to integrate mental health considerations into climate adaptation and urban public health strategies, particularly for vulnerable young adult populations in rapidly urbanizing regions.

STD_PSTR_004: One-Step TMLE vs. Cox Regression for Causal Survival Analysis in Oncology: A Simulation Study

Vaishnavi Deore

Manipal Academy of Higher Education

Cancer survival studies are complicated by censoring, dropout, and time-varying treatment, which can bias standard methods. This study evaluates one-step Targeted Maximum Likelihood Estimation (TMLE) for estimating causal survival effects in simulated oncology time-to-event data, benchmarked against Cox regression. Fifteen simulation scenarios varied three factors: event rate (high, moderate, low), confounding strength (weak $\beta=0.2$, strong $\beta=0.8$), and sample size ($n=100, 500, 1000$). Performance was assessed using bias, Monte Carlo standard deviation, RMSE, and 95% CI coverage. TMLE performed best under strong confounding with adequate covariate overlap, doubly-robust correction outperformed Cox on every metric, and it widened as sample size increased. Cox was more reliable under heavy censoring and sparse events, where TMLE's hazard-based components became unstable. Cox's advantage grew with n .

STD_PSTR_005: A Canonical Correlation Analysis of Metabolic Dysregulation and Systemic Inflammation

Urmika Mandal

IMS, BHU

Background: Metabolic dysregulation and systemic inflammation are closely interconnected processes that contribute to the development and progression of type 2 diabetes mellitus (T2DM) and cardiovascular disease. Conventional univariate approaches fail to capture these complex multidimensional relationships. This study aimed to evaluate the association between metabolic and inflammatory profiles using Canonical Correlation

Analysis (CCA).

Methods: A cross-sectional analysis was conducted using NHANES 2017–2018 data comprising 909 individuals aged ≤ 30 years. The metabolic variable set included BMI, waist circumference, HbA1c, HOMA-IR, HDL, total cholesterol, triglycerides, and LDL, while the inflammatory set comprised HS-CRP, NLR, SII, SIRI, PIV, PLR, NPR, LMR, PAR, albumin, and ferritin. Data preprocessing involved imputation, winsorization, log transformation, and standardization. CCA was performed, and Wilks' Lambda, canonical loadings, and redundancy indices were used for interpretation.

Results: Eight canonical functions were extracted, with the first two demonstrating very strong associations ($R_c = 0.990$ and 0.981 ; $p < 0.001$). Triglycerides, HOMA-IR, waist circumference, BMI, and HDL were the major contributors to the metabolic variate, while PAR, LMR, NPR, and SIRI contributed most to the inflammatory variate. The first canonical function explained the greatest proportion of shared variance, indicating substantial overlap between metabolic dysfunction and systemic inflammation.

Conclusion: Canonical Correlation Analysis effectively identified a strong multivariate relationship between metabolic abnormalities and inflammatory markers. The findings highlight triglycerides as a key metabolic determinant linked with systemic inflammation and support the integration of metabolic and inflammatory biomarkers to improve cardiometabolic risk assessment.

STD_PSTR_008: AI-Driven Digital Twins for Next-Generation Clinical Trial Simulation

Prabhuraj K

Bharathidasan University

Clinical trials are essential for evaluating the safety and efficacy of new drugs, but they are often expensive, time-consuming, and prone to failure due to poor patient recruitment and population variability. This study proposes an AI-driven digital twin framework that creates virtual patients using anonymized electronic health records, previous clinical trial data, genomic information, and real-world datasets collected from diverse populations worldwide. These digital twins simulate how patients are likely to respond to different treatments before conducting actual clinical trials. The framework integrates advanced biostatistical methods, including descriptive statistics, propensity score matching, Bayesian hierarchical modeling, mixed-effects regression, Kaplan–Meier survival analysis, Cox proportional hazards regression, Monte Carlo simulation, bootstrap resampling, and adaptive Bayesian trial design to ensure reliable prediction and statistical validation. Machine learning models such as Random Forest, XGBoost, and deep neural networks are used to predict treatment efficacy, adverse drug reactions, and patient dropout risk. Statistical programming using R, Python, and SAS supports data preprocessing, model development, simulation, and validation, while explainable AI techniques improve model transparency. The proposed framework is expected to optimize patient selection, reduce trial duration and cost, improve statistical power, and support better decision-making before human recruitment. This AI-enabled in silico clinical trial approach has the potential to complement conventional clinical trials and accelerate precision drug development while maintaining scientific rigor and regulatory confidence.

STD_PSTR_010: Privacy-Preserving Federated Learning for Multi-Hospital Digital Pathology Analysis

Yogita K

MAM college

Digital pathology has transformed histopathological diagnosis by enabling high-resolution whole-slide image analysis using artificial intelligence (AI). However, developing robust AI models requires large, diverse datasets that are often restricted by patient privacy regulations and institutional data-sharing policies. This study proposes a privacy-preserving federated learning (FL) framework for collaborative digital pathology analysis across multiple hospitals without transferring sensitive patient data. In the proposed framework, each participating institution trains a local deep learning model on its own histopathology images, while only encrypted model parameters are shared with a central aggregation server. Differential privacy and secure aggregation techniques are incorporated to minimize the risk of patient data leakage while maintaining model performance. The framework is designed to improve tumor classification, cancer subtype prediction, and histopathological image segmentation by leveraging diverse datasets from geographically distributed healthcare centers. Performance will be evaluated using accuracy, sensitivity, specificity, F1-score, and area under the receiver operating characteristic curve (AUC). Compared with conventional centralized learning, the proposed federated framework aims to achieve comparable diagnostic accuracy while ensuring data privacy, regulatory compliance, and enhanced model generalizability. This approach has the potential to accelerate trustworthy AI adoption in precision pathology and collaborative cancer diagnostics across healthcare institutions.

STD_PSTR_013:

Darshan R

St Josephs institute of technology

Video surveillance systems are widely used to monitor public and private spaces for security purposes. However, manual monitoring of large volumes of surveillance footage is inefficient and prone to human error. This project proposes an intelligent video surveillance system that automatically detects suspicious or criminal activities using deep learning techniques.

The system analyzes video frames extracted from surveillance footage and identifies abnormal activities such as robbery, assault, or vandalism. By using a CNN-based model combined with machine learning classification, the system can detect anomalies and alert users in real time.

STD_PSTR_014: Effects of Acute Micro/Nano Plastics Exposure on Oxidative Stress Biomarkers – A Comprehensive Meta-Analysis of In Vitro and In Vivo Studies using RStudio.

Madhumitha Thirumalaisamy

Bharathiar university

Micro/Nanoplastics (MNP) have emerged as a serious threat to the environment and human health. Although oxidative stress is recognized as a key mechanism underlying MNPs-induced toxicity, the effects of acute MNP exposure on oxidative stress biomarkers remains inconclusive. This work aimed to quantitatively evaluate this gap through a comprehensive meta-analysis. A systematic literature search was conducted across PubMed, ScienceDirect and Web of Science between 2019 and 2024. Out of 7679 retrieved studies, 14 in vitro and 20 in vivo studies were selected after screening and quality assessment. Random-effects meta-analyses were performed in R studio to assess changes in reactive oxygen species (ROS) and catalase (CAT) levels, using standardized mean differences (SMDs) as effect size. Acute MNPs exposure significantly increased ROS, while antioxidant responses differed between experimental systems, with decreased CAT activity in vitro, while increased CAT activity in vivo.

STD_PSTR_016: StatAssist: A Hybrid AI Assistant for Statistical Programming, Code Evaluation and Data Analytics in SAS and R

Niranjana S Variar

University of Calicut

Problem Statement:

Statistical programming in SAS and R requires significant programming expertise. Existing platforms lack integrated natural language interaction and beginner guidance. Users often switch between multiple tools for code generation, validation, visualization, and reporting.

Learning statistical programming remains challenging for students and researchers.

Motivation:

Develop a unified intelligent platform that simplifies statistical programming by integrating natural language processing, automated code generation, validation, visualization, and AI-assisted analytical support.

STD_PSTR_018: MediScribe AI- AI Assisted Automation for Regulatory Writing

Berwin N

St.Joesph's Institute of Technology

Clinical Study Report (CSR) generation remains a major bottleneck in biopharmaceutical development, often requiring 6 to 12 weeks of manual effort and carrying significant risks of transposition errors across complex clinical trial documents. This project introduces a secure, highly reliable AI Copilot designed for maximal human-in-the-loop augmentation rather than total replacement. Built entirely on an open-source, locally deployed infrastructure, the system eliminates reliance on centralized cloud platforms to guarantee

absolute data privacy, intellectual property protection, and zero PHI leakage. The architecture integrates locally fine-tuned open-weight language models, a Hybrid GraphRAG pipeline for cell-level mathematical traceability back to statistical tables, and an integrated version-aware regulatory learning assistant. Enforced through 21 CFR Part 11 compliant audit logging and multi-stage human approval gates, this framework reduces first-draft timelines from weeks to 2–3 days while maintaining strict compliance with ICH E3 and regional health authority guidelines.

STD_PSTR_021: AI assisted Dependency Mapping and Change Impact Analysis for Clinical Programming using NLP

Jerlin J

St.Joesph's Institute of Technology

Clinical programming teams often spend considerable time manually identifying how changes in ADaM datasets impact Tables, Listings, and Figures (TLFs), making the process time-consuming, error-prone, and dependent on programmer expertise.

This work proposes an NLP-powered ADaM–TLF Dependency Mapping framework that automatically analyzes SAS programs to identify relationships between ADaM variables and TLF outputs. The framework parses SAS code, extracts datasets and variables, constructs a dependency graph, and performs automated impact analysis whenever a variable is modified. It generates comprehensive impact reports that help programmers quickly identify affected outputs, reducing manual effort, improving traceability, and enhancing the efficiency, accuracy, and reliability of clinical programming workflows.

STD_PSTR_022: SentiBot AI - Enhancing Clinical Trial Compliance Through Automated Protocol Deviation Detection

Paul Ibzhan JB

St.Joesph's Institute of Technology

Protocol deviations are a major source of increased monitoring effort and regulatory findings in clinical trials. While Electronic Data Capture (EDC) systems store structured data and free-text comments, potential deviations are often hidden in narrative entries and require time-consuming manual review.

SentiBot is an AI-powered Natural Language Processing (NLP) solution that analyzes free-text data to identify protocol deviations such as missed visits, missed doses, medication non-compliance, incorrect dosing, and prohibited concomitant medication use. It cross-validates detected events with structured EDC datasets, including visit schedules, exposure records, and concomitant medications, to improve accuracy and reduce false positives. Rather than replacing human review, SentiBot generates risk-ranked alerts for Clinical Operations, Data Management, or Medical Monitors to verify. This decision-support approach reduces manual effort, improves protocol compliance, enhances inspection readiness, and aligns with ICH-GCP, ALCOA+, and 21 CFR Part 11 principles.

STD_PSTR_024: AI Assisted SAP Driven Mock Shell Generation for Clinical Trial Tables, Listings and Figures (TLFs)

Panimalar Ezhilan

VIT Vellore

Developing mock shells for Tables, Listings, and Figures (TLFs) from the Statistical Analysis Plan (SAP) is a critical step in clinical study reporting, yet it remains predominantly manual, repetitive, and time-intensive. This traditional workflow demands substantial biostatistical effort to interpret SAP specifications, translate them into TLF metadata, and ensure consistency and traceability across study outputs.

To address these challenges, this project proposes an AI-assisted, SAP-driven tool to transform TLF mock shell development from a repetitive manual task into an intelligent, efficient, and review-driven workflow. Built in Python, the tool combines automated SAP keyword extraction with a structured metadata-entry interface to interpret SAP content and intelligently pre-populate TLF mock shell specifications.

The extracted information is presented through a reviewable and editable interface, keeping the biostatistician at the center of validation and decision-making. Each metadata element can be linked to its corresponding SAP source through annotations, providing field-level traceability, faster review, and greater confidence in the generated output. Validated metadata can then be exported through standardized Excel outputs for downstream TLF programming and reusable generation of related mock shells. By integrating AI-assisted automation, standardization, traceability, reusability, and human oversight into a unified workflow, the solution aims to reduce repetitive transcription and formatting effort while preserving biostatistical expertise and judgment.

STD_PSTR_025: StatOS: An Agentic AI Framework for Modernizing Enterprise Statistical Programming

Sabinesh S

Stella Mary's College of Engineering

StatOS is an Agentic AI framework designed to modernize enterprise statistical programming, where large legacy analytical codebases are often difficult to understand, maintain, validate, and evolve due to complex dependencies and limited documentation.

Existing modernization approaches primarily focus on code migration, providing limited support for program understanding, dependency intelligence, validation, and documentation. StatOS addresses these challenges through specialized AI agents that collaboratively perform program analysis, dependency mapping, documentation generation, validation, and modernization support using deterministic analysis, Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), and a Knowledge Graph.

Modernization in StatOS includes automated program understanding, dependency analysis, documentation generation, validation assistance, and migration recommendations. The framework delivers an AI-assisted enterprise platform that enables explainable, context-aware, and scalable statistical programming modernization, improving developer productivity, code quality, and maintainability in regulated analytical environments.

STD_PSTR_026: aCRF Builder - Streamlining Clinical Trial Metadata Management Through Automated Annotated CRF Generation

Shalini R

SRM, Trichy

This project introduces an automated, CDISC SDTM-MSG v2.0-compliant Annotated Case Report Form (aCRF) generation system.

Bridging raw PDF CRFs and Excel mapping sheets, the engine leverages a Tiered Smart Search algorithm for exact visual label anchoring, intelligent master-page referencing for repeating clinical forms, and robust cell deduplication.

Featuring an interactive UI for real-time drag-and-drop box adjustments, automated 4-color palette enforcement, and a hierarchical Domain-Form-Visit bookmarking structure, the solution eliminates redundant manual annotation efforts.

Built-in compliance checks validate styling, text patterns, and layout constraints before final vector PDF compilation.

By replacing labor-intensive manual PDF markup with rule-based precision and dynamic layout decollision, this platform accelerates clinical regulatory submission workflows, ensures strict adherence to CDISC guidelines, and establishes a foundation for future enterprise-grade, privacy-compliant AI integrations.

STD_PSTR_029: AI-powered R-to-SAS Code Translation with Clinical Trial Intelligence

Lakshmi Prabha D

Bharathiar University

Clinical trial data management requires datasets to comply with global standards such as CDISC and the Study Data Tabulation Model (SDTM). Traditionally, SAS has been the primary programming language for creating these standardized datasets, while modern data scientists increasingly use R for data preparation and analysis.

This difference often requires manual conversion of R scripts into SAS code, making the process time-consuming and prone to errors. This project presents an automated offline R/Natural Language-to-SAS Code Converter that simplifies this workflow. The framework combines Python Regular Expressions (Regex) for fast translation of simple syntax with a locally deployed Qwen 2.5 Coder 7B Instruct model through LM Studio for handling complex code transformations and natural language instructions.

It also includes a clinical domain classifier to organize variables into SDTM domains such as Demographics (DM) and Adverse Events (AE), along with a self-healing validation module that detects and corrects common SAS coding errors. Operating entirely offline, the system ensures patient data security and regulatory compliance.

STD_PSTR_031: Ohwia caudata safeguards hyperglycemia-induced renal cell damage: An invitro dose-dependent antifibrotic study

Sakthishree N

Bharathiar University

Diabetic kidney disease (DKD) is a leading cause of end-stage renal disease, affecting nearly 25% of individuals with diabetes worldwide. It is characterized by hyperglycemia-induced tubular epithelial-myofibroblast transdifferentiation (TEMt) and renal fibrosis. This study evaluated the protective effects of Ohwia caudata (OC) against high glucose-induced injury in HK-2 (Human kidney) cells treated with 45 mM glucose and OC (10, 50, and 100 µg/mL). A completely randomized experimental design comprising five treatment groups was considered, and biomarker responses were summarized using mean ± standard deviation. Group differences were assessed using one-way analysis of variance (ANOVA), followed by Tukey's post hoc multiple comparison test, with statistical significance set at $P < 0.05$. Dose-response trends and fibrosis markers were evaluated using correlation and linear regression analyses. Effects of Ohwia caudata extract on cell viability of high glucose challenged HK2 proximal tubular cell were analyzed using MTT assay.

OC treatment demonstrated a significant, dose-dependent reduction in TEMt and fibrosis-associated biomarkers, including α -SMA and vimentin, compared with the high-glucose group ($P < 0.05$). Immunofluorescence analysis indicated restoration of MFN1 expression and suppression of p-Drp1 (Ser616), suggesting inhibition of mitochondrial fission. The statistical model indicated significant treatment effects across experimental groups, supporting a strong association between improved mitochondrial homeostasis and reduced fibrotic response. These findings suggest that Ohwia caudata attenuates diabetic renal fibrosis and highlights statistical modelling in quantifying treatment efficacy and identifying potential therapeutic targets for diabetic kidney disease.

Thank you