IASCT
Indian Association for Statistics in Clinical Trials

ConSPIC TRIVANDRUM 2025

ConSPIC

TRIVANDRUM 2025

September 12 - 13
Hyatt Regency, Trivandrum

e-Brochure

Conference for Statistics and Programming in Clinical Research

Organized by

Indian Association for Statistics in Clinical Trials (IASCT)

Indian Association for Statistics in Clinical Trials

Contents

Welcome Note from the President

Dear Friends,

On behalf of the **IASCT Executive Committee**, it is my pleasure to welcome you all to the 14th edition of our flagship event **"Conference for Statistics and Programming in Clinical Research (ConSPIC 2025)".**

The preparations for ConSPIC 2025 (September 12th and 13th) are in full swing. We are very excited, as this year, ConSPIC is coming to the beautiful, vibrant and culturally rich city of **Trivandrum**.

We heard your feedback last year and therefore this time ConSPIC is packaged as a smart 2-day event, plus a pre-conference workshop (11th September) for industry and students respectively, instead of the traditional 3 daylong event. Though the duration is reduced, the essence of ConSPIC is retained i.e. learning, science, building connections and includes the scintillating ConSPIC night on 12th Sep.

I would like to thank all industry colleagues and students, who have sent their abstracts this year for ConSPIC. Thanks to the Program Scientific Committee (SC), who have meticulously and painstakingly gone through all the abstracts (300+) and carefully selected 100+ abstracts that will be presented over the 2 days of ConSPIC as talks and posters across 4 streams - Biostatistics, Statistical Programming, Data Sciences and Management. Thanks to all the presenters who have worked hard on their presentations. Without them, ConSPIC wouldn't be the same.

As always, this year too we have keynote/plenary talks, a leadership session and a panel discussion with eminent industry leaders and experts.

Kudos to the Organizing Committee (OC) and our team of volunteers, who are working tirelessly and leaving no stone unturned, to put together what promises to be a remarkable and memorable event.

A special note of thanks to the **Program Chair**, **Dr. Pooja Shinde f**or her exceptional leadership and support to IASCT over the years**.** Pooja is a seasoned IASCT stalwart with many years of experience in organizing IASCT events, including previous editions of ConSPIC.

IASCT has been a learning platform since 2007 and 2024-25 was no different. It was a busy year, with multiple events/workshops and industry/academia connect events and ConSPIC

2024 in Pune. It was very successful and saw over 300 participants, 100+ talks and 40+ posters.

I would like to express my heartfelt gratitude to all IASCT committee members (Events, Admin, social media team) and volunteers. Their dedication and deep commitment towards IASCT is what makes IASCT events unique and special.

We are immensely thankful to the esteemed IASCT sponsors for 2024-25 (19 sponsors), for their continuous encouragement, generous and unwavering support which has helped us to organize several events. We are glad to be associated with organizations who share the same passion for learning and growth, as we do at IASCT.

This year (2025-26) we have 13 sponsors, with some new additions. We are truly grateful for their partnership and support which will help IASCT to continue organizing high quality learning events and bringing experienced industry experts and leaders to many of these events/workshops.

I would like to extend my gratitude to all IASCT members for their participation, constant support, encouragement, enthusiasm and passion for clinical research and learning.

I would also like to thank all my fellow Executive Committee members, Pooja Shinde, Rashmi Jain, Sucharita Pilli, Nishanth Nalan, Jagannatha P.S., Surendra Yajamanam, for their constant support and contributions to IASCT and the clinical research community.

We are delighted to invite you this September to ConSPIC 2025 in Trivandrum. We look forward to insightful learning sessions, reconnecting with colleagues, and building new friendships

Warm regards,

Ashish Charles

President, IASCT

From the Desk: Program Chair

Dear Colleagues and Participants,

We are delighted to welcome you to the Conference for Statistics and Programming in Clinical Research (ConSPIC) 2025, to be held in Trivandrum on Sept 12th and 13th. This year's two-day event promises to be both enriching and inspiring, bringing together a vibrant mix of industry professionals and academic researchers.

Key highlights of ConSPIC 2025 include:

- Multiple thematic tracks featuring selected talks and poster presentations by experts from the industry and emerging student researchers
- A compelling keynote address, along with plenary sessions and an energetic panel discussion
- Opportunities to engage with thought leaders who are shaping the future of statistics and programming in clinical research

We believe ConSPIC 2025 will serve as an excellent platform for knowledge exchange, networking, and collaboration. Your active participation will be instrumental in making this conference a memorable and impactful experience.

Before I conclude, I'd like to extend a heartfelt thank you to our sponsors for their unwavering support—ConSPIC and other IASCT events would not be possible without your generous contributions.

I also would like to express my sincere gratitude to the Executive Committee for their valuable guidance, the Organizing Committee for their dedication and tireless efforts in bringing this event to life, and the Events Management and Admin Committee for their continued support across all IASCT initiatives.

Together, your commitment and collaboration have been instrumental in shaping the success of ConSPIC 2025.

We look forward to seeing you there!

Regards

Pooja Shinde

Program Chair

Our Sponsors

Agenda

https://iasct.net/conspic2025/agenda

Pre-conference Workshop

Dr. Mahesh Iyer, Ph.D.
Head of Biometrics and Data Sciences, BMS India

Mahesh is currently Head of Biometrics and Data Sciences at Bristol-Myers Squibb, India. Mahesh Iyer has over 25 years of experience in research and development in healthcare and has been responsible for guiding many products through the development life cycle throughout his career at Pfizer, Novartis, Bristol Myers Squibb and Boehringer-Ingelheim.

Prior to his current role at BMS, Mahesh was Head of Stats, AI/ML, Quantitative and Digital Sciences, and Innovation, India-Philippines at Pfizer. Prior to his Pfizer role, Mahesh was VP at Parexel, responsible for heading the Innovation and Technology function for Parexel and was also India Head for Global Data Operations. Mahesh has also been a co-founder of Sineflex Solutions LLP, a consulting firm focused on enabling and accelerating innovation in the healthcare space. He was also the head of the BIRAC-funded med-tech accelerator at the Centre for Innovation and Entrepreneurship, IIIT Hyderabad. Mahesh continues to be responsible for mentoring startups in the med-tech space and helping them scale their products and solutions.

Mahesh brings a strong analytical mind-set, deep insights into healthcare development and a proven record of implementing innovative solutions in the healthcare domain. Mahesh is passionate about enhancing industry academia collaboration; he set up one of the first part-time Ph.D. programs in Statistics for Novartis associates, teaches at several Indian universities and has chaired multiple conferences over the years. He is Past-President of the Indian Association for Statistics in Clinical Trials and Past-President of the International Indian Statistical Association, India Chapter. He was recently featured in the Top-100 AI leaders in India.

Apart from his functional activities, Mahesh has led various organizational developmental activities and trainings. He is a certified coach in emotional intelligence, assessment centers, and psychometric evaluations. Mahesh has completed his Ph.D. in Statistics from Temple University, Philadelphia.

Keynote Speaker

Nataraj Kalyanaraman
Vice President and Head of drug development Hyderabad
Bristol Myers Squibb

Nataraj Kalyanaraman oversees the BMS Drug Development organization in Hyderabad, the largest Drug Development site outside the US. He previously held leadership roles at BMS in Switzerland and the U.S., contributing to portfolio strategy, strategic transformations, and integration for acquisitions like Celgene and Myokardia.

With over 20 years of experience in biopharma, healthcare, and global health, he brings deep expertise in biopharmaceutical R&D. Before joining BMS in 2017, he was a strategy consultant advising major pharma companies and global health organizations.

He holds degrees in pharmacy, medicinal chemistry, and an MBA, and began his career as a scientist in Preclinical Research.

Plenary Speakers

Suresh Ramu
Co-founder & CEO
Cytecare Hospitals

Suresh Ramu is the Co-founder & CEO of Cytecare Hospitals, a comprehensive cancer care hospital network. The first 150-bed organ-site focused cancer hospital that went live in 2016 in Bangalore. Ramu is an alumnus of IIT- Madras and IIM- Calcutta and he started his entrepreneurial journey in 2011 as Co-founder of Cytespace Research, a clinical research site solutions organization.

Ramu began his career with PricewaterhouseCoopers, later with a startup telemedicine company. He then was with Quintiles Translational for about 10 years leading various global and regional functions like Clinical Operatins, Biometrics and Pharmacovigilance.

Interactive Group Discussion

Surendra Yajamanam

Director Programming, *Department of Biostatistics*
GlaxoSmithKline Pharmaceuticals Ltd

Surendra Yajamanam is currently Director, Programming in the Department of Biostatistics in GlaxoSmithKline Pharmaceuticals Ltd.

He earned his bachelor's degree in mechanical engineering from Sri Krishnadevaraya University in India. Surendra has been working in GSK for the past 19 years, with over two decades of experience in clinical programming across a wide spectrum of therapeutic areas, establishing himself as a subject matter expert in the clinical domain. Under his leadership, the Oncology and HIV programming teams in India have witnessed remarkable expansion, establishing themselves as integral partners of the global team. Surendra has played a pivotal role in Clinical Programming of GSK for several years and is evident through his contributions to some of the key capability projects & initiatives at GSK.

Beyond his professional role, Surendra is deeply committed to advancing clinical research in India and actively participates in numerous professional endeavors, including chairing conferences and presenting at various industry forums. He also served as the president of the Indian Association for Statistics in Clinical Trials (IASCT) in the past.

Topic of his presention- “Transforming clinical Trials analysis and reporting - The agile way”. An interactive session outlining the ways of working using Agile methodology, with real life examples. Users will have the opportunity to learn how to apply these learnings to their everyday work and optimize analysis and reporting clinical trials.

Panel discussion

Venu Mallarapu
Executive Vice President
eClinical Solutions

Venu Mallarapu is a business and technology leader with over 26 years of experience in Life Sciences industry. He has helped organizations with business and IT advisory, strategic consulting, relationship, and delivery management. As eClinical Solutions' Executive Vice President of Global Operations, Venu is responsible for overseeing the full spectrum of operational ecosystem spanning global operations, customer success, support services, India operations, and operational excellence.

Venu is a subject matter expert in the Clinical, Regulatory, Quality and Safety & Pharmacovigilance functions. He has delivered strategy and transformation advisory consulting to top global pharma, biotech, vaccine, and medical devices clients. He speaks at various life sciences and tech industry events on transformation, innovation and next generation technologies and infrastructure. He is a recognized industry thought leader with published articles, blog posts, webinars, and seminars on a variety of topics that impact and drive Life Sciences R&D.

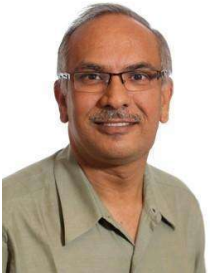
His life goal is to make a positive impact on human health & happiness.

Dr. Mahesh Iyer, Ph.D.
Head of Biometrics and Data Sciences,
BMS India

Mahesh is currently Head of Biometrics and Data Sciences at Bristol-Myers Squibb, India. Mahesh Iyer has over 25 years of experience in research and development in healthcare and has been responsible for guiding many products through the development life-cycle throughout his career at Pfizer, Novartis, Bristol Myers Squibb and Boehringer-Ingelheim. Prior to his current role at BMS, Mahesh was Head of Stats, AI/ML, Quantitative and Digital Sciences, and Innovation, India-Philippines at Pfizer. Prior to his Pfizer role, Mahesh was VP at Parexel, responsible for heading the

Innovation and Technology function for Parexel, and was also India Head for Global Data Operations. Mahesh has also been a co-founder of Sineflex Solutions LLP, a consulting firm focused on enabling and accelerating innovation in the healthcare space. He was also the head of the BIRAC funded med-tech accelerator located at the Centre for Innovation and Entrepreneurship, IIIT Hyderabad. Mahesh continues to be responsible for mentoring startups in the med-tech space, and helping them scale their products and solutions. Mahesh brings a strong analytical mind-set, deep insights into healthcare development and a proven record of implementing innovative solutions in the healthcare domain. Mahesh is passionate about enhancing industry academia collaboration; he set up one of the first part-time Ph.D. programs in Statistics for Novartis associates, teaches at a number of Indian universities and has chaired multiple conferences over the years. He currently leads the AI/ML working group for ISCR. He is Past-President of the Indian Association for Statistics in Clinical Trials and Past-President of the International Indian Statistical Association, India Chapter. He was recently featured in the Top-100 AI leaders in India, and the Top-20 Future of Work Icons for AI/ML. Apart from his functional activities, Mahesh has led various organizational developmental activities and trainings. He is a certified coach in the areas of emotional intelligence, assessment centers, and psychometric evaluations. Mahesh has completed his Ph.D. in Statistics from Temple University, Philadelphia.

Dr. Ashwini Mathur
Executive Director
Onesto Consulting



Ashwini Mathur is a leader in the clinical research area with nearly 30 years of experience in research, data science, and healthcare innovation. He holds a Ph.D. in Biostatistics from the University of California, Berkeley, and an Executive MBA from the Indian Institute of Management, Bangalore. His career spans leadership roles at GlaxoSmithKline in the US and India, and at Novartis across India, Ireland, and the UK, where he built and managed global teams in data science, AI, clinical operations, and technology innovation. Most recently, as Executive Director at Onesto Consulting, he focuses on leadership training, healthcare analytics, and strategic consulting. Dr. Mathur has been recognized with multiple industry awards and has contributed to academia as an Adjunct Professor at UCSF and RCSI. With expertise in AI, digital transformation, and team development, he

continues to mentor the next generation of leaders in clinical research and healthcare.

Katherine HUTCHINSON
EVP Alliance Management
Veramed

35+ years in the pharma industry across big pharma and biotech (15 years) and service sector (20+ years). Work globally, based in UK. Educated to Masters level in Statistics, have supported drug development as a statistician, project statistician and manager. Latterly, held senior management roles to grow and support teams of biometricians and executive oversight of client portfolios. Believer in people and quality driving success, navigating and utilising AI tools to achieve efficiency.Genuinely excited at the impact of AI that will be realised over the next 1-5 years in pharma to reduce the cost of getting drugs and vaccines to market.

IASCT Executive Committee

Ashish Charles, President

Director Statistical Programming, Clinical Data Operations, Novartis Healthcare

Ashish Charles is currently working as Director - Statistical Programming, Clinical Data Operations, at Novartis Healthcare Pvt. Ltd. Hyderabad. He has over 19 years of experience in the Pharma industry, leading and managing large teams across the globe, and leading analysis and reporting of clinical trials across various therapeutic/disease areas. He holds a master's degree in computer applications from Bangalore University.

Ashish has been associated with IASCT in various roles since a decade and has been instrumental in leading and organizing IASCT events across various locations within India. Until recently, Ashish served as the Secretary of IASCT. Ashish currently serves as the President of IASCT. He is very passionate about the clinical industry and the latest developments in medical science. On a personal front, Ashish loves to read and is fond of music.

Pooja Shinde, Vice President

Senior Manager Global Biometric Delivery Lead, Parexel

Pooja Shinde (Senior Manager Global Biometric Account Lead) currently works for Global Biometrics Oversight group in Parexel and responsible for strategically leading and managing client partnership(s) within Global Data Operations (GDO) to ensure Parexel delivers quality projects to its clients and achieves agreed operational targets.

Also, she is the Vice President of the Indian Association for Statistics in Clinical Trials (IASCT). She completed PGDBM from NMIMS, MSc in Clinical Research from SMU and holds a Graduate certificate in Neurorehabilitation from Brunel University. She has been an active participant at many industrial events like IASCT, ConSPIC, SCDM, PhUSE SDE, etc. She is a member of Indian Association of Physiotherapist (IAP) and likes travelling and sports.

Nishanth Nalan, Secretary

Director and Practice Head (Life Sciences), ACL Digital

Expert Practice Leader in Clinical Data Services with extensive experience in building, scaling and leading delivery capability centres in the US and India. Experienced in establishing CRO services delivery infrastructures (centers of delivery excellence) from scratch and led various infrastructure initiatives including QMS and key support functions.

Currently in a senior management role with both functional and business responsibilities overseeing the setup, leadership, culture and strategic direction of the practice to drive growth and revenue in alignment with the company's goals. Possesses strong business

acumen and developed expertise in integrated marketing, sales, PMO, RMG, sales operations, finance, talent acquisition and HR.

Experienced in diverse clinical therapeutic areas, including Oncology, Neurology, Gastrointestinal, Dermatology, Vaccine, Urology, Muscular-Skeletal, Cardiovascular, Infectious Diseases, Alzheimer's Disease, and Public Health Crisis Studies like COVID-19 interventions. Hands-on experience in delivering key submissions throughout the lifecycle of programming deliverables like DMC, IA, CSR, SCS, RMP, PSUR, DSUR.

With 19 years of industry experience in Life Sciences including 14 years managing major business accounts, contracts, and sponsors and 12 years leading clinical business groups/units (Clinical Data Services, Biometrics, Statistical Programming). Reported to the CEO/CXO for 12 years and to the VP - Global Head for 7 years.
Holds a master's in computer applications (MCA) and a Post Graduate Diploma in Human Resource Development (PGDHRD) from PSG CAS, Coimbatore. Also completed the Executive Education Program at the Indian Institute of Management (IIM), Bangalore. Known for a strong work ethic and values in the workplace with a reputation for building high-performing and inclusive organizations, business units, teams, and structures

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.

Rashmi Jain, Events Committee Head
Statistical Specialist, Novo Nordisk

Rashmi G Jain is a Statistician and has more than 13 years of experience in the field of Biostatistics. Currently, she is a Statistical Specialist for Biostatistics team in Novo Nordisk Bangalore. She has been driving various non-interventional study activities, clinical trials, and exploratory analysis of data across different disease areas.

Rashmi post-graduated with a statistics degree in 2010 from Pune University. Actively involved in external outreach programs. Presented in PhUSE, CDISC, IASCT, ConSPIC and ISCR conferences.

She is eager to indorse strategic innovation and experimentation among statisticians. She has a pharma experience by working with GSK vaccines and Novo Nordisk. She likes travelling and cooking.

Sucharita Pillai, Admin Committee Head
Statistical Specialist, Novo Nordisk

Sucharita is a Senior Statistical Programmer at Novo Nordisk. She has 9 plus years of experience in the clinical domain and is responsible for data analysis, system development, expert, and learner in building applications by open-source technologies and Power-applications and passionate in digitalization. She is a Volunteer in PHUSE, India and working as a member in PHUSE working group "Best Practices for Interactive Analysis for Decision Making Submissions". She writes poems and is active in social activities.

Surendra Yajamanam Ex-President
Director – Programming, GlaxoSmithKline Pharmaceuticals Ltd.

Surendra Yajamanam is currently Director, Programming in the Department of Biostatistics (India) in GlaxoSmithKline Pharmaceuticals Ltd. He holds a bachelor's degree in mechanical engineering from Sri Krishnadevaraya University, India. Surendra is associated with clinical programming for 20+ years with a broad range of experience across various therapeutic areas and is SME in clinical domain.
He is passionate about innovation and played a significant role in various efficiency improvement initiatives in GSK. He is associated with IASCT over a decade and is currently serving IASCT as President. Surendra is highly committed about clinical research growth in India and actively involved in various professional activities as chair of conferences and presented in various professional forums. ConSPIC means to me: A learning feast with lots of fun, vibrant ambience, and fabulous occasion to meet friends

Scientific Program Committee

Dr. Sachit Ganapathy

Principal Biostatistician, Pfizer

Dr. Sachit Ganapathy is the Principal Biostatistician at Pfizer Healthcare India Pvt. Ltd., where he leads the design, analysis, and reporting of biological and omics data for late-phase clinical oncology programs. He also develops tools and pipelines to empower scientists and statisticians. He is passionate about mentoring young statisticians and believes in the power of data to tell compelling stories and spark meaningful conversations.

ConSPIC to him: Networking, fun, invigorating

Abhishek Mittal

Associate Director, Statistical Programming (SP), Johnson & Johnson

Abhishek Mittal is Associate Director of Statistical Programming at Johnson & Johnson, where he leads the Oncology Programming Team in India and serves as Site Lead for the Bengaluru office. With over 20 years of experience across BMS, Novartis, and IQVIA, he brings expertise in therapeutic areas including Infectious Diseases, Neuroscience, Rare Diseases, and Oncology. An MBA graduate from Rutgers Business School, he is passionate about professional development and has contributed to numerous regulatory submissions while driving initiatives in business operations, vendor management, and change management.

ConSPIC to him Learn and Fun

Aduri Chinnappa

Associate Director, Bristol Myers Squibb

Aduri Chinnappa Reddy is an Associate Director at Bristol Myers Squibb Business Services India Pvt. Ltd., where he leads programming for a Cell Therapy compound. He holds an M.Sc. in Applied & Biostatistics from Hasselt University and began his as a Biostatistician at Business & Decision. He transitioned to statistical programming at Novartis Pharma in 2009 and joined BMS India in 2023. His extensive experience spans Oncology, Cardio, Renal, and Cell Therapy, with significant contributions to global regulatory submissions, direct interactions with health authorities, and data integration during mergers and acquisitions.

ConSPIC to him: Learn and Fun

Saumil Tripathi

Associate Director, Ephicacy Life Sciences

Saumil Tripathi is an Associate Director of Biostatistics & Programming at Ephicacy with 16 years of experience in clinical research and statistical programming. He holds a degree in Pharmacy and has worked at IQVIA and Cytel, gaining extensive expertise in global clinical research. Currently, as FSP Manager, he leads project and portfolio programming across diverse clinical research engagements, with a focus on Vaccines, Oncology, and Respiratory. His expertise spans PK/PD, safety, efficacy analysis, and submission programming. Outside of work, he enjoys cultural engagement, traveling, and watching movies.

ConSPIC to him: ConSPIC is a valuable data science platform fostering idea sharing, networking, and knowledge exchange through engaging sessions.

Ajay Sinha

Director, Statistical Programming – Oncology, AstraZeneca

Ajay Sinha serves as the Director of Statistical Programming – Oncology at AstraZeneca, bringing over 20 years of experience in the life sciences sector. As a forward-thinking leader in Biometrics, he has built and expanded high-performing teams in both India and the US, driving innovation, automation, and successful regulatory submissions. At AstraZeneca, he leads the Statistical Programming – Oncology team, advancing progress in oncology research. Beyond his leadership role, he contributes as a mentor, faculty member, and global thought leader, shaping the future of the Biometrics community.

ConSPIC to him: Motivation/Inspiration, Connection, Growth

Organizing Committee

Anurag Srivastav

Associate Director, Ephicacy Lifesciences

Anurag Srivastav, Associate Director at Ephicacy Lifesciences Analytics Pvt. Ltd., brings 17 years of experience in the clinical and pharmaceutical industry. With expertise in statistical programming and analysis, he manages teams and projects, resolves challenges, and mentors' statistical programmers.

He is passionate about making a difference in clinical research, contributing to patient outcomes by advancing trials and accelerating timelines to meet industry expectations.

ConSPIC to him: Conference for Students and Presenters to Ignite Catalyst

Ketan Kapoor

Senior Programmer,Veramed

Ketan serves as a Senior Programmer at Veramed and holds a master's degree in Statistics. Driven by his conviction that "work is worship" and that genuine efforts always yield results, he approaches his work with both dedication and enthusiasm. His career includes positions at Novartis and Syneos Health, experiences that have not only fostered his self-learning but also enriched his understanding of reliability and connectivity within the biostatistical programming domain.

ConSPIC to him: Being part of ConSPIC is an opportunity to engage, learn, grow, and share knowledge. For him, ConSPIC is best summed up as "productive networking."

Samrat Bollam

Business Developer, SCL IT Technologies

Samrat is an accomplished business development professional with 15+ years of cross-industry experience spanning biostatistics, CRO services, SAS programming, clinical staffing, and banking operations. At SCL IT Technologies, he specializes in client engagement, regulatory compliance, and banking risk and control management, with a strong background in due diligence and process management. He is committed to driving growth, building

partnerships, and delivering value across clinical operations and financial services.

ConSPIC to him: ConSPIC is a space to be inspired and challenged, where collaboration and fresh insights drive real-world impact and collective growth.

Satya Vyshnavi Thondapu

Associate Director, Eli Lilly and Company

Satya Vyshnavi Thondapu is an Associate Director – Statistical Programming at Eli Lilly and Company, with nearly 14 years of expertise in statistical programming. She also serves the community as a member of the IASCT Events Committee. Holding a Master's in Pharmacy and an Executive Program in Leadership and Management from IIM, she is passionate about building highly skilled teams, driving technical innovation, and fostering an inclusive workplace through strategic leadership.

ConSPIC to her: A space to be inspired, challenged, and energized – turning fresh insights into real-world impact.

Sitaram Sahoo

Manager, Bristol Myers Squibb

Sitaram Sahoo is a Manager at Bristol Myers Squibb, contributing as an individual contributor to key compounds in the Immunology therapeutic area. With over six years of experience, he brings expertise in SAS and R programming, ensuring high-quality outputs that support impactful clinical research. He holds a master's in bioinformatics from IBAB, and has a strong interest in competitive intelligence, leveraging data to gain deeper insights into the performance of competitor molecules and their shared objectives.

ConSPIC to him: A wonderful platform where like-minded professionals across the industry come together to learn, share, collaborate, and foster collective growth.

Sugumaran Muthuraj

Senior Manager, Precision for Medicine

Sugumaran Muthuraj is a Senior manager in statistical programming with over 14 years of experience in clinical research. He specializes in statistical programming, data analytics, and visualization, and has led regulatory submissions for biopharma, biotech, and CRO clients. Previously with ICON Clinical Research, he joined Precision for Medicine in 2021. He is passionate about mentoring, driving innovation in software solutions, and enhancing research outcomes. He holds an MBA in Pharmaceutical Management, a master's in bioinformatics, and is currently pursuing a PhD in Bioinformatics. Beyond his professional contributions, he is also actively involved in CSR initiatives.

ConSPIC to him: Conference for Students and Presenters to Ignite Catalyst

Vijay Dharmaraj

Senior Manager, Statistical Programming, Symbiance

Vijay Dharmaraj is a Senior Statistical Programmer at ACL Digital and a seasoned clinical research professional with over nine years of experience. He is recognized for his precision, teamwork, and mentorship. He has contributed across multiple therapeutic areas, with strong expertise in ISS/ISE analysis for oncology trials and hands-on involvement in regulatory submissions such as HAQ and DSUR. Known for his attention to detail and ability to translate complex data into meaningful insights, he also mentors colleagues and plays a key role in driving team success. Outside of work, he enjoys watching cricket and tennis, and playing chess.

ConSPIC to him: A place to connect with peers, exchange ideas, and learn from the best minds in the industry.

VinayKrishna Thatta

Senior Clinical Data Standards Specialist, Novartis

VinayKrishna Thatta is a Senior Clinical Data Standards Specialist at Novartis with nine years of experience in the pharmaceutical industry. With a background in Computer Science Engineering, he specializes in clinical data management, automation, and transformation, with a focus on cardiovascular therapies. He plays a key role in ensuring consistency and compliance across global studies, driving efficiency through innovation and adherence to standards. As a member of ConSPIC for the past two years, Vinay is actively engaged in advancing clinical data standards in clinical research.

ConSPIC to him: ConSPIC is a platform for networking, learning from industry peers, and staying updated with the latest trends.

Biostatistics Talks

BIOS_PPT_001: A joint latent-class Bayesian model with application to ALL maintenance studies

Damitri Kundu
Eli Lilly Services India Pvt. Ltd.

Acute Lymphocytic Leukemia (ALL) is a significant cause of childhood cancer deaths, particularly in developing nations. This presentation details a joint latent-class Bayesian analysis of clinical trial data from the Tata Translational Cancer Research Center, focusing on relapse rates among pediatric ALL patients treated with 6MP and MTx. The study measured lymphocyte, neutrophil, and platelet counts during weekly/bi-weekly clinical visits. A latent-class model was applied to lymphocyte counts, the primary biomarker associated with ALL, while linear mixed models were utilized for the other two biomarkers. Time-to-event was modeled using a semi-parametric proportional hazards model, which was linked to longitudinal submodels through Gaussian random effects.The proposed model identified two latent classes for lymphocyte counts, estimating class-specific average non-relapse probabilities at various time points, demonstrating a significant difference between the classes. An extensive simulation study validated the accuracy and practical utility of the proposed joint latent-class model compared to traditional models.

Keywords: Acute Lymphocytic Leukemia (ALL), Bayesian Hierarchical Models, Longitudinal outcomes, Latent class model, MCMC.

BIOS_PPT_016: Exploring Heterogeneity of Treatment Effects in Clinical Trials

Sagar Das
Eli Lilly Services India Pvt. Ltd.

In this presentation we will review statistical methods for evaluating the heterogeneity of treatment effects (HTE). The focus is on subgroup identification and the estimation of individualized treatment regimens from randomized clinical trials. We will categorize various approaches into principled methods and common practices, highlighting the advantages and disadvantages. Discussion will include the challenges of subgroup assessments, the importance of principled data-driven subgroup identification, and the integration of causal inference, machine learning, and multiple hypothesis testing. The topic also includes a simulation study and a case study to illustrate the application of these methods.

BIOS_PPT_021: Statistical designs for combination-drug dose escalation clinical trials: a methodological review

Malhar Khatu
Janssen (Johnson & Johnson)

Identifying the optimal dose combination in dual-drug trials is challenging due to complex drug interactions and uncertain dose-toxicity relationships. Traditional dose-escalation methods may not fully capture these dynamics, often leading to sub-optimal dosing decisions. In this work, we explore and compare three adaptive dose-finding designs—Combination-BOIN, TALE, and Bivariate-BLRM—within a simultaneous dose-escalation strategy, assessing their ability to estimate the maximum tolerated combination (MTC)

safely and accurately across clinically relevant hypothetical scenarios. We also evaluate stepwise sequential versus simultaneous escalation strategies using the Combination-BOIN design through extensive simulations, focusing on selection accuracy, patient safety, early stopping, and trial efficiency. This comprehensive evaluation aims to guide dose-finding strategies and contribute to a better understanding of alternative designs that may enhance trial efficiency and accelerate access to safe and effective combination therapies.

BIOS_PPT_022: Advancing Clinical Trial Analysis with Trimmed Means: Estimands, Intercurrent Events, and Regulatory Alignment

Prasanth Palanisamy
Sensan Biosciences

This presentation provides a detail analysis of trimmed means, particularly addressing ICHs (Intercurrent Events) within the framework established by the ICH E9(R1) addendum. Based on Permutt and Li approach is highly related to handling the ICEs such as treatment discontinuation due to lack of efficacy or the occurrence of severe adverse events. This particular use of trimmed means basically redefines the "missing data problem." As an alternative of looking, it solely as a statistical imputation challenge, it treats dropout as a clinical outcome. This approach efficiently treats these events as "bad outcomes", that are then accounted for by trimming them from the analysis with one-sided trimming method is particularly suited for Missing Not At Random (MNAR) scenarios. FDA recognize Estimands framework and robust statistical techniques, trimmed means are currently more commonly used for sensitivity analyses than as primary efficacy endpoints. This presentation outlines calculation, assumptions, and application through SAS.

BIOS_PPT_023: Navigating Treatment Landscapes: Multi-Level Network Meta-Regression (MLNMR)

ABDUL MASUD TARAFDER
Eli Lilly Services India Pvt. Ltd

Multi-Level Network Meta-Regression (ML-NMR) offers a pivotal advancement in evidence synthesis, directly confronting the pervasive challenge of real-world heterogeneity in Network Meta-Analysis (NMA). This Bayesian framework rigorously adjusts for varying treatment effect modifiers (TEMs) across studies using both individual patient data (IPD) and aggregate data (AgD), enabling more accurate indirect comparisons. Leveraging the multinma R package, ML-NMR models complex treatment-covariate interactions to estimate effects tailored for specific target populations. We demonstrate its application with a case study on PASI75 response in plaque psoriasis, using mixed IPD (three RESOLVE trials) and AgD (FIXTURE study) across six treatment arms. The model yielded robust, population-adjusted treatment effects and SUCRA rankings, identifying JXT_Q2W and JXT_Q4W as highly effective. Demonstrating superior robustness over standard NMA and MAIC, ML-NMR significantly minimizes bias, providing highly reliable comparative effectiveness estimates crucial for nuanced Health Technology Assessment (HTA) and clinical guideline development.

BIOS_PPT_025: A 2-in-1 Adaptive Phase 2/3 Design for Oncology Drug Development - An Innovative Adaptive Design Option

Suresh Chenji
Eli Lilly Services India Pvt. Ltd

While adaptive phase 2/3 design intends to reduce the risk of a false Go-decision of a straight phase 3 design, it inevitably increases the risk of a false No-Go decision. An alternate innovative design, the 2-in-1 adaptive design allows to expand an ongoing phase 2 trial into a phase 3 trial to expedite a drug development program with fewer patients. An intermediate endpoint can be used for the adaptive decision. Under a mild assumption that is expected to generally hold in practice, both the phase 2 trial(in case of no expansion) and phase 3(in case of expansion) can be tested at the full alpha level without inflating the overall type 1 error of the study. Extensions to the 2-in-1 design include allowing for multiple adaptive decisions, incorporating group sequential methods for data monitoring, and expanding the design to handle multiple endpoints. This presentation will include an example study and a hypothetical case study in advanced solid tumors to demonstrate the benefits of the 2-in-1 design.

BIOS_PPT_028: Meta-Analysis of multi-omics data for reproducibility scoring in Rheumatoid Arthritis: A statistical framework for biomarker robustness

Sabari Sanghami P
Pfizer

Reproducibility challenges in omics-based biomarker discovery hinder progress in translational research across diseases. This study proposes to identify the reproducibility of potential biomarkers by performing a meta-analysis of omics data from multiple publicly available datasets in Rheumatoid Arthritis (RA). Using statistical methods, we aim to calculate reproducibility scores for genes across studies. An interactive dashboard will be developed to visualize reproducibility scores, study contributions, and biomarker rankings, offering a dynamic tool for prioritizing reliable candidates. The goal is to support data-driven patient stratification and endpoint refinement in RA clinical trials. By combining robust statistical methods with user-focused visualization, this work aims to enhance transparency, foster reproducible science, and align with the growing emphasis on evidence robustness in precision medicine.

Keywords: Rheumatoid Arthritis, meta-analysis, reproducibility scores, biomarker, statistical, dashboard, patient stratification.

BIOS_PPT_030: Stopping Early: Insights into Group Sequential Design

Swati Gopal
Novo Nordisk

A group sequential design is an emerging methodology used in planning a pivotal trial as this approach has the ability to stop the study either for efficacy or futility at an interim look at the data before reaching the planned sample size. In scenarios where early outcomes are available, this method offers significant potential to lower sample size substantially by stopping at interim.

In this presentation I will be using simulated data from an example to illustrate the graphical approach to control the family-wise type I error rate for the confirmatory tests with multiple endpoints. Along with alpha spending on all endpoints in the testing strategy and the look back approach. Finally, provide inference on the statistical procedure for the confirmatory endpoints while adjusting for estimates, p-values and confidence intervals using stage-wise ordering.

BIOS_PPT_034: Advancing Survival Analysis: Time-Dependent RMST Modelling for Oncology

Nagadharshini M
Pfizer

Restricted Mean Survival Time (RMST) has emerged as a robust and interpretable alternative, to traditional hazard-based survival models, particularly when the proportional hazards assumption does not hold. RMST modelling can be enhanced, by incorporating time-dependent covariates through an inverse probability of censoring weighting (IPCW) approach. In this presentation, we will be using Monte Carlo simulations, to assess the estimation accuracy and predictive performance of the time-dependent RMST (T-RMST) model, in comparison to the time-dependent Cox (T-Cox) and fixed-covariate RMST (F-

RMST) models. We will be practically analyzing with Oncology data, to research each model, and to identify the best fitting model that would enable more accurate survival analysis and predictions.

BIOS_PPT_036: Missing data in clinical trials: from clinical assumptions to statistical analysis using patter mixture models

Naseema Parveen Shaik
PAREXEL Pvt. Ltd

The need to use rigorous, transparent, clearly interpretable, and scientifically justified methodology for preventing and dealing with missing data in clinical trials has been a focus of much attention from regulators, practitioners, and academicians over the past years.

New guidelines and recommendations emphasize the importance of :

- Minimizing the amount of missing data and carefully selecting primary analysis methods on the basis of assumptions regarding the missingness mechanism suitable for the study at hand.
- The need to stress-test the results of the primary analysis under different sets of assumptions through a range of sensitivity analyses.

This presentation illustrates several strategies for missing data based on pattern mixture models that embody clear and realistic clinical assumptions.

Pattern mixture models provide a statistically reasonable yet transparent framework for translating clinical assumptions into statistical analyses.

BIOS_PPT_037: Statistical Challenges and Modeling Strategies in Multi-Centre Clinical Trials: Ensuring Robustness Across Diverse Sites

Shreya Singh
Inference

Multi-centre clinical trials are essential for improving generalizability and efficiency in clinical research. However, they also introduce challenges due to heterogeneity in site characteristics, patient populations, and study conduct. This presentation explores these challenges through the lens of a simulated case where two trials using the same drug and placebo produce divergent outcomes-one statistically significant, the other not. We investigate the role of site-level variability, randomization imbalances, and interaction effects in influencing results. Various modeling strategies, including mixed-effects models and Bayesian hierarchical approaches, are discussed to ensure robustness of conclusions. Differences in how the treatment works and how patients respond at different study sites can make the results confusing or hard to trust.

An example using data on dichotomous outcomes, such as responder/not responder demonstrates how difference in treatment effects and variation between study sites can lead to confusing or unreliable results.

BIOS_PPT_040: Reference Based Multiple Imputation

Sandipan Ghosh
Inference

Reference-based multiple imputation methods have become popular for handling missing data in randomized clinical trials. In this presentation some popular reference-based technics, such as jump-to-reference (J2R), copy increments in reference (CIR), copy reference (CR), and last mean carried forward (LMCF) and their specific assumptions and ideal scenario to implement for each imputation methods are briefly discussed. Imputation of real-world data (RWD) using these different technics with SAS and the analysis of the variance of those data is done to understand which imputation methods fit better for which situation. Reference-based multiple imputation is more popular for a continuous outcome, but in this presentation, we dive deep into developing and evaluate the reference-based multiple imputation approach for longitudinal binary data via joint modeling with the multivariate normal distribution and joint modeling with a latent normal model.

BIOS_PPT_042: Bayesian Model Averaging in Dose-Response Models

Manusree Banerjee
Bristol Myers Squibb

Selecting an optimal and safe dose remains a pivotal challenge in drug development, with uncertainties significantly influencing regulatory success. Traditional dose-finding methods relying on pairwise comparisons of treatment arms with multiplicity adjustment are increasingly being replaced by dose-response modeling, which leverages all available data for more efficient and informed decision-making. This presentation discusses about a novel class of longitudinal dose-response models within a Bayesian model averaging framework, designed to incorporate multiple time points and accommodate complex response profiles, including non-monotonic patterns. The methodology enhances dose estimation accuracy, supports early interim decisions, and effectively handles missing data. Through extensive simulations, the proposed approach demonstrates superior performance in dose estimation and trial efficiency, offering a flexible and robust tool for modern dose-finding studies.

BIOS_PPT_043: Response-Adaptive Randomization (RAR)

Somnath Karan
Inference

Respond-Adaptive Randomization (RAR) is a dynamic clinical trial design that optimizes therapeutic benefits and trial efficacy by adjusting treatment allocation probabilities based on interim patient outcomes. By using frequentist or Bayesian frameworks, RAR reduces patient exposure to less effective therapies by favoring treatments with higher efficacy or safety. In doing so, statistical rigor is maintained while ethical standards are raised. Because of the wide range of therapeutic responses, RAR is frequently employed in cancer.

This abstract examines the fundamentals, advantages, drawbacks, and revolutionary potential of RAR in aligning scientific and ethical goals for improved patient outcomes. Using the BATTLE case study as an example, Bayesian Response-Adaptive Randomization (RAR) was employed to allocate 255 lung cancer patients to targeted therapies based on biomarker-driven responses. This adaptive design allowed the trial to preferentially assign patients to more effective treatments as evidence accumulated, ultimately achieving a 46% disease control rate. Despite its advantages, RAR presents several challenges, including potential inflation of type I error, the need for complex real-time data analysis, and the risk of biased treatment effect estimates. Nevertheless, the flexibility and patient-centric nature of RAR make it particularly well-suited for personalized treatment strategies, especially in complex and heterogeneous diseases such as cancer.

BIOS_PPT_048: Application of BOIN (Bayesian Optimal Interval) design and its operating characteristics in dose-finding trial

Vishal Dhamal
Novartis Healthcare Pvt. Ltd.

BOIN designs are a class of model-assisted dose-finding designs that can be used in oncology trials to determine the maximum tolerated dose (MTD) of a study drug based on safety. BOIN designs are operationally simple to implement and have good statistical operating characteristics compared to other dose-finding designs. One feature of BOIN design that may not be optimal is the behavior of the decision rule at the lowest dose. The extra-safe option allows for a stricter stopping boundary at the lowest dose level to address this.

This approach is applied to a dose-finding trial to address unmet need for safe and effective therapy in patients with relapsed/refractory Acute Lymphoblastic Leukemia (ALL). The trial setting assumes 25% target toxicity, maximum of 15 participants, with up to 9 treated at same dose level and 4 dose levels. The operating characteristics (OC) based on simulations shows that the design selects robust dose.

BIOS_PPT_055: A feasibility study to assess the adequacy of historical clinical trial data using standard of care to inform the development of a new treatment for a rare disease

Aditya Joshi
Novartis Healthcare Pvt. Ltd.

We are developing a compound for a rare disease. The program is in phase II, to be followed by phase III study, to demonstrate non-inferiority of new treatment to control. Due to fewer new cases, challenges in patient recruitment are expected. So, there is a potential opportunity for using already-collected data (external control) to be augmented to the control group. We intend to assess the feasibility of leveraging external database to the ongoing Phase II clinical trial in an integrated analysis with more controls. This feasibility assessment will be conducted in two parts. Part I will include identifying the source data, applying target trial emulation framework and assessing the population heterogeneity. Part II will consider a MAP* prior approach based on propensity scores, utilizing historical data from earlier conducted trials. The team will engage with regulatory authorities to confirm the adequacy of the approach and acceptance for phase III study.

BIOS_PPT_056: Enhancing Continuous Glucose Monitoring Analysis: The Advantage of Beta Regression Models for Proportion Data

Jevitha Lobo
Novo Nordisk Service Centre India Private Ltd.

Continuous Glucose Monitoring (CGM) is a technology that allows individuals with diabetes to monitor their blood glucose levels in real time, both day and night. In clinical trials, a treatment is considered effective if it helps patients achieve their CGM targets, which indicate the percentage of time their glucose levels stay within a designated range. Traditionally, these targets are estimated and compared between treatments using the ANCOVA model, based on the assumption of normality. However, given that percentage data ranges between 0 and 100, this presentation suggests for the use of a Beta regression model as a more suitable analytical approach. This model is specifically designed for response variables that are continuous and restricted to the interval (0, 1), making it particularly well-suited for percentage data. From the analytical results this presentation concludes Beta regression model performs better compared to the ANCOVA model and an accurate representation of the relationships and variability present in the data, making it an ideal approach for studies involving percentage/proportions, thereby enhancing the insights derived from the analysis.

BIOS_PPT_061: Forced Randomization in Clinical Trials

Anand Kumar
ICON Clinical Research India Pvt. Ltd.

In randomized controlled trials (RCTs), clinical sites may encounter situations where an eligible participant is randomized to a treatment arm for which the investigational drug is temporarily unavailable. This challenge can lead to either refusal to enroll the participant or the use of “forced randomization” (FR), where the participant is allocated to an available treatment arm without disclosing the drug stockout to the site, facilitated by Interactive Response Technology (IRT).

While FR is increasingly accepted in confirmatory trials when used sparingly, its statistical properties and impact on trial integrity require further investigation. Results show that FR helps maintain enrollment and treatment balance, mitigating disruptions from drug stockouts. However, excessive FR use may introduce bias, highlighting the need for risk mitigation. FR offers a practical approach to managing drug supply challenges while preserving trial integrity when applied judiciously.

BIOS_PPT_063: Integrating Estimand and Target Trial Emulation Approaches with Propensity Score-Based IPTW and Cox Model based IPCW for External Control Analyses using Observational Data

Soumyajit Nandy
PAREXEL Pvt. Ltd.

Formulating the correct scientific question of interest, in causal inference is a crucial step. Here causal inference principle is applied to external control analysis using observational data in metastatic non-small cell lung cancer and illustrate the process to define the estimand attributes. The "Estimand Framework" (EF) can be used to define attributes of the ideal trial, while the "Target Trial Framework" (TTF) can address specific issues in defining the estimand attributes using observational data. The comparison of long-term survival outcomes between patients from three pooled randomized phase 3 trials receiving front-line chemotherapy and similar patients treated with front-line chemotherapy in routine clinical care (observational data) was performed (Polito et al., 2024). The combined approach (EF and TTF) allowed for clear definition of the estimand and aligned estimator while accounting for key baseline confounders, index date, and subsequent therapies. Propensity score-based IPTW implemented to mitigate confounding in the electronic health record-derived dataset for key baseline confounders and Cox model based IPCW implemented to address potential bias from informative censoring. Results showed a hazard ratio estimate close to 1, indicating similar survival between the pooled control arm and the external control (Polito et al., 2024). The proposed combined framework provides clarity on the causal contrast of interest and the estimator to adopt, facilitating design and interpretation of analyses using observational data for external control.

BIOS_PPT_064: Assess Directionality of Multiple Endpoints

Sarah Anderson
Emmes

In clinical trials primary and secondary endpoints are usually tested separately. A single endpoint may not capture the impact of treatment across multiple key endpoints, for example for rare diseases or small sample size. If the endpoints are not met but all favor the intervention arm, this can provide some indication for the next steps. A global rank test is used to combine multiple endpoints to increase power without needing to adjust for type I error. It focuses on ranks rather than original data to assess directionality. Using the permutation test with the global rank test provides evidence that the directionality in the data did not happen by chance and shows the benefit of the invention arm.

BIOS_PPT_065: Understanding the Closed Test Procedure in Oncology Trials for Effective Outcome

Pradeep Acharya
Ephicacy Lifescience Analytics

The Closed Test Procedure is one of the most powerful and efficient test procedures to control multiple type I error rate when multiple hypotheses have to be tested. This potential inflation of the type I error also refers to Multiplicity which occurs due to various reasons including comparisons across multiple treatment arms which significantly reduces the power of the statistical test.This presentation talks about the potential reasons for the occurrence of multiplicity and its impact on study results. This also explains how the use of Closed Test

Procedure controls multiple alpha-levels and mitigates the multiplicity to get the reliable and generalizable results. A Simulated data from Oncology therapeutic area shall be used to test multiple hypotheses and impact of multiplicity. The application of Closed Test Procedure is explained along with defining the user interface, creation of closed set of hypotheses automatically and test them and then present the results

BIOS_PPT_066: Tipping Point Analysis in R and SAS: A Comparative Approach to Handling Missing Data in Clinical Research

Jeet Agrawal, Sushil
PAREXEL Pvt. Ltd.

This presentation explores the implementation of Tipping Point Analysis (TPA) in R and SAS, comparing their approaches to assessing the robustness of clinical trial results in the presence of missing data. TPA is a vital sensitivity analysis method in clinical research, helping to determine how much missing data can influence study conclusions. We will discuss the theoretical foundations of TPA and demonstrate its practical application using both R and SAS, utilizing simulated data representing a typical randomized controlled trial with a continuous primary outcome and various patterns of missing data. The presentation will cover the syntax, functionality, and output interpretation for TPA in each platform, highlighting their respective strengths and limitations. We'll examine how R's flexibility compare with SAS's structured procedures and established presence in the pharmaceutical industry. Additionally, we'll also briefly discuss the regulatory perspective on accepting R-based results for this purpose. By the end of this session, we will have a comprehensive understanding of how to perform TPA using both software packages, enabling them to choose the most suitable tool for their specific clinical trial analysis needs.

BIOS_PPT_068: Beyond Toxicity : Accelerating Dose-Finding with Pharmacokinetic-Driven Bayesian Design PKBOIN-12

Anetta Mary Xavier
eClinical Solutions

In early-phase clinical trials, especially those involving immunotherapies and targeted agents, the focus has shifted from identifying the Maximum Tolerated Dose (MTD) to determining the Optimal Biological Dose (OBD), which aims to achieve the best balance between therapeutic effect and safety. Although pharmacokinetic (PK) data—used to assess how a drug behaves in the body—are routinely collected, they are often overlooked in existing dose-finding strategies. To leverage this valuable information, we introduce **PKBOIN-12**, a model-assisted Bayesian design that incorporates PK data alongside toxicity and efficacy outcomes to improve OBD determination. This design can be expanded to handle late-onset treatment responses using time-to-event modeling. Simulation will be performed, to demonstrate that PKBOIN-12 enhances OBD selection accuracy, increases patient allocation to beneficial dose levels, and reduces the risk of underexposure. This design represents a significant step toward more precise and data-informed dosing strategies in early-phase clinical trials.

BIOS_PPT_069: Illustrating treatment switching using Rank preserving structural failure time model in Oncology study

Swarupa Changan
eClinical Solutions

In many randomized clinical trials, where survival is a key endpoint, subjects are usually intended to be on their assigned treatment until loss of clinical improvement or death. In oncology trials, the crossover from the control to the experimental treatment is reasonable when disease progression is observed or any other clinical criteria is noticed —a process known as *treatment switching*. Traditional Intent-to-treat analyses that ignore such crossover which can underestimate the true effect of the experimental treatment on overall survival. To address this, we have the novel strategic model the **Rank-preserving structural failure time model (RPSFTM),** which effectively corrects the impact of crossover, and this method is used to estimate counterfactual survival—what would have occurred had patients not switched. RPSFTM adjusts the effects of treatment switching by modeling, what the survival times of patients who switched treatments would have been if they had remained on the control treatment.

BIOS_PPT_072: Overall Survival as a Safety Endpoint in Indolent Cancer Clinical Trials

Elbin Siby
Bristol Myers Squibb

Indolent cancers, such as follicular lymphoma and chronic lymphocytic leukemia, have long overall survival (OS) times, often exceeding 5–10 years, complicating clinical trial design. Pivotal trials typically use intermediate outcomes like progression-free survival (PFS) for regulatory approval. However, post-approval studies have shown concerning OS trends despite favorable PFS results, highlighting the need for structured OS monitoring as a safety endpoint. This abstract proposes a quantitative framework for OS monitoring in indolent cancer trials, based on regulatory experience and statistical modeling. The approach includes pre-specified OS monitoring guidelines defined by three parameters: a hazard ratio threshold for unacceptable detriment (HRdet), a plausible benefit threshold (HRben), and the expected number of OS events (Lx). These guidelines aim to inform benefit-risk assessments rather than trigger early trial termination. Inspired by cardiovascular safety monitoring in Type 2 Diabetes Mellitus, the framework supports regulatory decision-making by ensuring transparency, stakeholder alignment, and the continuation of trials post-approval to gather meaningful OS data.

Ind_SP_PPT_004: An Overview of IWG Criteria and Efficacy Endpoints for MDS & CMML

Priyanka Bisoyi
Novartis Healthcare Pvt. Ltd.

In recent years, the increasing number of therapies for Myelodysplastic Syndrome (MDS) has led to the development of standardized response criteria to compare different treatments. The IWG criteria were developed to monitor disease response through clinically relevant measurements. This evaluates four areas of disease response: alteration of disease history, cytogenetic responses, hematological improvement (HI) categorized into HI-E, HI-N, and HI-P responses, and quality of life assessment. Additionally, CRh, defined by specific bone marrow and blood count criteria, will be assessed. Other major efficacy endpoints include ORR, BOR, DOR, PFS, and TTP, reduction in spleen volume and MPN SAF total symptom score calculations for MDS and CMML.This abstract emphasizes the derivation of HI Categories (HI-E, HI-P, and HI-N) according to the IWG 2006 criteria, with a particular focus on HI in patients who were RBC/Platelet transfusion-dependent at baseline and achieved transfusion independence for at least 8 consecutive week.

Ind_SP_PPT_005: Optimizing Clinical Data Quality Through CTCAE v5 Laboratory Toxicity Implementation

Reshmanath R
Catalyst

Laboratory data are essential for evaluating patient safety in clinical trials. While traditional interpretation relies on reference ranges, toxicity grading using NCI-CTCAE v5 standardized framework for grading lab-based toxicities. Integrating CTCAE-based grading into the SDTM LB domain presents challenges due to varied grading logic—some tests require comparison to ULN (e.g., ALT), others to LLN (e.g., hemoglobin), baseline comparisons (e.g., creatinine), or fixed thresholds (e.g., urine output). To address this, we developed a dynamic SAS macro that automates toxicity grading by reading configurable rules from an Excel file, applying specific scenarios per test and combining these results to SDTM LB domain. The macro generates grade outputs and merges them into the LB domain. This approach improves flexibility, accuracy, and consistency, while reducing programming effort and simplifying maintenance. The paper outlines the macro's design, implementation strategy, and benefits across studies.

Ind_SP_PPT_008: Cost Effective Automation of SDTM aCRF Bookmarking and ToC

Uma Nithyanandhan
Emmes

The SDTM annotated Case Report Form (aCRF) is a PDF document that maps data collection fields to corresponding SDTM variables and domains. When properly prepared with bookmarks and a Table of Contents (ToC), it becomes a valuable tool for efficient review. The SDTM Metadata Submission Guidelines recommend a specific bookmarking structure, but creating these bookmarks is often a manual, time-consuming task, typically requiring relatively expensive software along with plugins. Manual creation becomes

increasingly complex as the number of forms and visit timepoints grow. This paper presents a standardized, scalable, and cost-effective approach using the free tool jpdfbookmarks and the reasonably priced Foxit Phantom PDF. Bookmark, ToC text and hyperlinks are easily controlled via an input text file template, allowing for automation and consistency. This method supports any number of CRFs and visits schedules and can be quickly implemented across projects by programmers at any experience level.

Ind_SP_PPT_009: Innovative and Efficient Techniques for Developing Efficacy ADaM Datasets in Diabetes Mellitus

Sarath Chandar S
Ephicacy Lifescience Analytics

Diabetes mellitus is a chronic, multifactorial metabolic disorder characterized by persistent hyperglycemia. The two major forms—Type 1 and Type 2 diabetes—are associated with severe complications affecting the cardiovascular, ocular, renal, and nervous systems. Diabetes can lead to life-threatening outcomes and is closely linked to other major health conditions, including obesity and certain cancers. World Health Organization estimates that nearly 800 million people are affected by diabetes globally. The International Diabetes Federation (IDF) reports an estimated global incidence of 589 million in 2024, with projections rising to 853 million by 2050.

The application of CDISC ADaM and TAUG-Diabetes standards to derive efficacy analysis datasets in diabetes studies. ADaM facilitates traceability, consistency, and regulatory compliance by standardizing the structure and content of analysis datasets. The TAUG-Diabetes offers specific implementation guidance for endpoints relevant to diabetes trials. In antidiabetic clinical trials, the primary efficacy endpoint is typically the reduction in glycosylated hemoglobin (HbA1c) from baseline, as it reflects long-term glucose control. Secondary endpoints include the assessment of hypoglycemic events and glucose monitoring profiles, providing additional insights into the safety and effectiveness of investigational treatments.

This paper presents the rationale, structure, and implementation of CDISC ADaM and TAUG-Diabetes standards for deriving efficacy analysis datasets in diabetes clinical studies. It outlines methodologies for constructing the glycosylated hemoglobin dataset (ADHB1Ac) and the glucose profile dataset (ADSMBG) from SDTM Laboratory (LB) data and further details the derivation of hypoglycemic event analysis datasets (ADHYPO and ADHYPOSUM) using information from the SDTM Clinical Events (CE) domain.

Ind_SP_PPT_012: FHIR Meets CDISC: Bridging Clinical Standards

Sarinraj Premraj
Inference

Integrating real-world data into clinical research requires interoperability between healthcare and research standards. This paper outlines an implementation strategy for mapping Health Level Seven Fast Healthcare Interoperability Resources (HL7 FHIR)-commonly used in clinical systems to collect and exchange healthcare data-to Clinical Data Interchange Standards Consortium (CDISC) standards such as the Study Data Tabulation Model (SDTM) and Clinical Data Acquisition Standards Harmonization (CDASH).

The process includes defining use cases, mapping data elements, aligning terminologies (e.g., LOINC, SNOMED), and building Extract, Transform, Load (ETL) pipelines. Using tools like FHIR APIs and Pinnacle 21, the framework enables accurate, automated transformation of data from Electronic Health Records (EHRs) into structured clinical trial datasets. This mapping enhances data quality, reduces manual entry, and supports regulatory compliance. Ultimately, it enables more efficient use of real-world evidence (RWE) in clinical research and regulatory submissions.

Ind_SP_PPT_013: Engineering ADFACE: Programming Safety Narratives from Complex Reactogenicity Signals

Rajesh Kumar Singh, Robins
ACL Digital

The ADFACE dataset was developed for a Phase Ia/b BCG vaccine trial assessing local and systemic reactogenicity in Vaccine Trials. Deriving ADFACE required consolidating raw data from multiple SDTM sources, primarily FACE, CE, and AE, into a coherent structure aligned with SAP-defined safety objectives. The mission was to provide a clean, submission-ready summary of solicited reactions by anatomical site, intensity, and timing. The complexity stemmed from mapping onset windows post-dose, harmonizing site-specific terms (e.g., pain at injection site, swelling, erythema), managing duplicate records, and ensuring consistency in toxicity grading, Investigational gradings, and resolution flags. The vision was to ensure traceability from source to submission, enabling accurate depiction of vaccine tolerability. By reconciling controlled terminology across domains, aligning events with protocol-defined visit windows, and implementing macro-driven derivations, ADFACE exemplifies how complex reactogenicity data can be structured into meaningful safety narratives tailored to early-phase vaccine studies.

Ind_SP_PPT_015: OMOP vs CDISC: Streamlining Real-World Evidence Programming through Common Data Models

Sreekanth Reddy Yasa
IQVIA

With the rise of real-world evidence (RWE) in regulatory and clinical decision-making, statistical programmers are increasingly working with diverse and unstructured real-world data (RWD). In this evolving landscape, the OMOP Common Data Model (CDM) offers a standardized, scalable approach to organizing and analyzing observational health data—unlocking new efficiencies for programming teams. This session provides a practical overview of how OMOP CDM supports rapid, reusable analytics across heterogeneous data sources, and contrasts it with CDISC standards (SDTM and ADaM), which remain the gold standard for clinical trial submissions. Key differences in structure, terminology, tools, and output readiness will be discussed from a programmer's perspective. Attendees will also explore how OMOP enables modular programming, metadata-driven automation, and accelerated cohort identification—core benefits for real-world study delivery. Ideal for statistical programmers and data managers, this session offers actionable insights on integrating OMOP into RWE workflows and clarifies its role alongside CDISC in today's hybrid data environment.

Ind_SP_PPT_017: Beyond the Curve: Survival Stories Through Kaplan-Meier with SAS & R

Chaithanya Vlupam
Fortrea

Survival analysis is a cornerstone of clinical research yet often remains intimidating to newcomers. This presentation aims to demystify the Kaplan-Meier (KM) estimator by introducing survival analysis concepts through simple, real-world examples. Attendees will gain a clear understanding of censored data, survival probabilities, and the step-function nature of KM curves. Using a hands-on approach, we compare the development of KM plots in both SAS and R, highlighting key differences in usability, customization, and statistical outputs such as log-rank tests and hazard ratios. Special emphasis will be placed on visual storytelling—how survival curves can reveal treatment effects, patient outcomes, and time-to-event dynamics. Tailored for aspiring programmers, analysts, and anyone curious about survival data, this session offers a practical path to mastering KM plots using both SAS and R. By the end, attendees will be equipped to build, interpret, and enhance KM plots with confidence.

Ind_SP_PPT_018: Optimizing SDRG and ADRG Validation: Enhancing Submission Quality through Traceability and Conformance Checks

Manisha Chaudhuri
Ephicacy Lifescience Analytics

With growing regulatory scrutiny and evolving CDISC standards, a well-validated Study Data Reviewer's Guide (SDRG) and Analysis Data Reviewer's Guide (ADRG) are essential to ensure a seamless regulatory submission. This presentation focuses on implementing a standardized and traceable validation approach using PHUSE templates, Pinnacle 21 outputs, and cross-checks with metadata and controlled terminology. By emphasizing conformance summaries, document consistency and traceability across datasets, this approach supports transparency, reduces the risk of reviewer misinterpretation and minimizes queries. This Presentation will include case studies based on lesson learnt and best practices to demonstrate practical examples on how issues such as inconsistent MedDRA versions or missing annotated CRF variables can be proactively addressed. We will also discuss how aligning SDRG/ADRG with regulatory expectations through automation and structured QC enhances submission quality and review efficiency. The presentation aims to equip sponsors and programmers with robust framework for optimizing reviewer guides and ensuring regulatory compliance.

Ind_SP_PPT_019: Addressing Ceiling Effects with Modeling Techniques for PK/PD Submission

Swaroop Kumar Koduri
Ephicacy Lifescience Analytics

Pharmacokinetic (PK) and pharmacodynamic (PD) analyses are essential for evaluating drug efficacy and safety, yet ceiling effects—where responses plateau despite dose increases—can compromise model accuracy. This abstract presents practical modeling strategies to address ceiling effects in PK/PD analyses using non-linear mixed-effects models (NLMEM) and Bayesian methods. We demonstrate how SAS procedures can be applied to model non-linear responses, account for inter-individual variability, and quantify parameter uncertainty. These techniques enable improved model fit, enhanced prediction accuracy, and better understanding of dose-response relationships. In parallel, we propose a structured workflow for integrating PK/PD data into electronic Common Technical Document (eCTD) packages. This includes traceability, metadata alignment, and validation steps to support regulatory compliance. By combining statistical modeling with robust data preparation, this approach equips clinical SAS programmers and statisticians with tools to overcome ceiling effects, support FDA submissions, and streamline the drug development process—ultimately accelerating access to effective therapies.

Ind_SP_PPT_020: Smart Model Selection in MMRM: An R-Based Approach to Two-Way vs Three-Way Interactions

Akansh Rajput
Eli Lilly

Mixed Models for Repeated Measures (MMRM) are widely used in the analysis of longitudinal clinical trial data due to their flexibility in handling missing data and varying time points. Statisticians frequently struggle with choosing between two-way and three-way interaction models, as well as identifying the most suitable variance-covariance structure. These decisions, when explored manually, can be subjective, time-consuming, and prone to inconsistencies, ultimately impacting the reliability of statistical inference.

To address these challenges, an R function has been developed that automates the selection between two-way and three-way MMRM models. This function evaluates multiple models by balancing model fit criteria (e.g., AIC, BIC) while simultaneously selecting an optimal variance-covariance structure from a range of common options (e.g., unstructured, compound symmetry, autoregressive). It provides comprehensive outputs including diagnostics, convergence checks, and summary statistics. Particularly useful in clinical trials and longitudinal studies where robust modeling of time and treatment interactions is critical.

During this presentation, this R function will be explained with a case study along with its benefits of consistency, model transparency, reduction of analyst burden, and reproducibility of workflows in clinical data analysis.

Ind_SP_PPT_024: Evolution from a SAS Programmer to an R Programmer: A Functional Approach

Jagadish Khatam
Princeps Technologies Inc.

This presentation explores the evolution of programming practices during the transition from a SAS programmer to an R programmer, using a common task—generating descriptive statistics by group—as a case study. The journey begins with a naive approach, where statistics are calculated separately for each numeric column and combined manually. It then progresses to writing reusable functions that reduce redundancy by looping over columns. Further refinement incorporates the rlang package, enabling the use of unquoted column names through tidy evaluation, thus improving flexibility and aligning with modern R programming standards. In the final stage, the function is enhanced to also support quoted column names, making it adaptable to various data manipulation workflows. This progression highlights the shift from procedural to functional and tidy programming paradigms, demonstrating how R programmers develop more efficient, reusable, and scalable code. It offers practical insights into writing cleaner, more expressive R code for clinical and statistical programming tasks.

Ind_SP_PPT_025: CDISC adaptations and standards in Decentralized Clinical Trials (DCTs)

Monisha Thiyagaraj
Sensan Biosciences

Decentralized Clinical Trials (DCTs) are transforming clinical research by enabling remote participation and real-time data capture. However, integrating diverse data sources such as eConsent, electronic Patient-Reported Outcomes (ePRO), wearables, and telehealth poses challenges for standardization and regulatory compliance. The Clinical Data Interchange Standards Consortium (CDISC) plays a critical role in ensuring data consistency across DCTs. The CDISC adaptations support decentralized models, including guidance for mapping novel data streams into CDASH, SDTM, and ADaM formats. It highlights practical approaches to maintaining traceability, managing data quality, and ensuring regulatory readiness in hybrid and fully remote trials. Emerging collaborations across industry and standards organizations further support harmonization efforts. By aligning DCT innovations with established CDISC standards, sponsors and researchers can ensure efficient, compliant, and patient-centric clinical trials.

Ind_SP_PPT_031: The Power of Centralized Safety Database

Archana Anbuchezhian
AstraZeneca

The **Centralized Safety Database** stands as a transformative force in delivering consistent and easily accessible FAIR (Findable, Accessible, Interoperable and Reusable) safety data, significantly advancing timely safety surveillance at the project level. In this presentation, we will explore transformative strategies for accelerating the seamless integration of data from individual studies into a cohesive, unified platform.

We will examine decision-making processes at the project level, evaluating the adequacy of individual study datasets in addressing comprehensive project-wide inquiries while recognizing the evolving nature of standardization practices. Our discussion will further highlight strategies to dynamically align all studies within a project, offering a comprehensive and holistic view of project-level information that enables faster, more incisive exploratory data analysis.

Furthermore, we will highlight the remarkable advancements made, detail the challenges successfully navigated during implementation, and underscore the practical value achieved.

Ind_SP_PPT_032: Automating Statistical Analysis in Clinical Research Using R and Tidyverse

Surenkosimin Murugan
Sensan Biosciences

Manual statistical programming in clinical research can be time-consuming, error-prone, and difficult to scale across studies. This presentation highlights a practical framework for automating core statistical analyses using the R programming language and the Tidyverse ecosystem. Leveraging tool such as dplyr, ggplot2, broom and purr, the session will demonstrate how standardized pipelines can be created to produce reproducible tables, figures, and listings (TFLs), streamline data transformations, and enable dynamic reporting.

Through real-world case studies, we will showcase how automation improves consistency,

reduces programming hours, and enhances traceability and audit readiness. Special focus will be given to how reusable R functions, parameterized scripts, and markdown-based outputs can be tailored for cross-functional needs ranging from biostatistics to data management.

Ind_SP_PPT_034: Accelerating iteration in R with Parallel Processing

Sarita Singh
Veramed

As the clinical industry transitions toward wider adoption of R for statistical programming, optimizing run time of iterative operations becomes increasingly critical. **Multiple imputation, bootstrapping, simulations** serve as prime examples of the computational challenges posed by intensive, repetitive workflows in R. Despite R's strengths in data analysis, it often struggles with runtime inefficiencies and memory limitations in large-scale, iterative tasks. In this work, we explore and benchmark various parallel processing strategies, including implementations using the **parallel**, **future**, and **foreach** frameworks, to significantly improve execution time. Using multiple imputation as a representative case study, we evaluate the trade-offs, scalability, and performance gains of each method. The findings provide practical guidance for selecting the most efficient approach based on dataset size and computing resources. These optimization strategies extend beyond imputation, offering broader applicability to other iterative functions in clinical data workflows and paving the way for more efficient R-based pipelines.

Ind_SP_PPT_035: Automating Define.xml Generation Using R: A Modern Approach to CDISC Compliance

Monisha Shree Venkateshan
Symbiance Pvt. Ltd.

Define.xml plays a pivotal role in clinical trial data submissions by capturing metadata for SDTM and ADaM datasets in compliance with CDISC standards. Manual processes and reliance on proprietary solutions can limit transparency, reproducibility, and scalability in this crucial step of regulatory documentation.

This abstract introduces an efficient, open-source pipeline for generating Define.xml version 2.0 using the R programming language. Using the **{defineR}** package, metadata structured in Excel is converted into XML output that aligns with regulatory specifications. Supplementary packages like **readxl**, **openxlsx**, and **writexl** streamline file manipulation, while **haven**, **dplyr**, and **tidyr** support seamless data integration and restructuring. Specialized handling of date-time values and terminology is accomplished with **hms** and **lubridate**.

This presentation will explore how XPT-derived metadata and modular Excel templates can be harnessed within R to automate Define.xml creation, enhancing submission quality through greater consistency, auditability, and regulatory alignment.

Ind_SP_PPT_036: Introduction to Dosimetry

Bijoy Kumar Dey
Novartis Healthcare pvt.ltd

Radioligand Therapy (RLT) is a modern cancer treatment that uses radioactive molecules to target cancer cells directly, reducing side effects compared to traditional radiation therapy. Dosimetry, the process of measuring and calculating radiation doses, is essential in RLT to assess how the radiation affects the body and ensure the treatment is safe and effective. The dosimetry process involves collecting and processing clinical and imaging data, as well as transferring it to specialized software for analysis. Standard steps include imaging to track radiation distribution, analyzing activity over time, and calculating radiation absorbed by organs and tumors using tools like OLINDA. To further illustrate its significance, VISION sub-study will be highlighted as a prime example demonstrating the critical role of dosimetry in RLT. Future advancements in dosimetry include personalized (Patient-specific) treatment planning, finding the best radiation dose for RLT, Integration with AI and Machine Learning, developing advanced models like PBPK, Biological dosimetry.

Ind_SP_PPT_038: Claiming Prospect in Retrospect: Statistical Programming Journey in RWE Submissions

Amitabh Manral
VAICS consulting

Real-World Data (RWD) from **Managed Access Programs (MAPs)** offers a unique opportunity to assess treatment outcomes in patients outside of traditional clinical trials, especially in rare or life-threatening conditions. Effective collaboration between **data management and statistical programming** is essential to ensure data quality, structure, and traceability. One key challenge in RWE is the absence of protocol-defined visits. **Visit mapping and windowing**, along with **date imputations**, are crucial to create coherent timelines for analysis. Transforming RWE into **CDISC-compliant datasets** introduces complexities, particularly in the **LB (Laboratory)** domain due to inconsistent units, non-standard test names, and missing reference ranges. Standardization, controlled terminology, and unit harmonization are vital. For regulatory submission, source data must be traceable and well-documented, including metadata, define.xml, and derivation logic. Adhering to structured frameworks enables RWD from MAPs to evolve into credible, regulatory-acceptable **Real-World Evidence (RWE)**.

Ind_SP_PPT_039: Modernizing Clinical Reporting in R: A Two-Part Approach to Innovation and Scalable Automation

Himani Narang, Ritika
Janssen (Johnson & Johnson)

As R gains traction in clinical trials, the Pharmaverse ecosystem has become a survival guide

for statistical programmers. Rooted in CDISC standards and regulatory needs, it offers open-source R packages that promote automation, reproducibility, and efficiency. This presentation introduces a two-part approach to clinical reporting using Pharmaverse tools.

Part 1 highlights the use of rtables and tidytlg packages to create high-quality, submission-ready TLFs. In parallel, it explores how teal package enables interactive dashboards for dynamic, CDISC-compliant data review and cross-functional collaboration. Part 2 looks to the future—to use these reporting workflows to reduce redundant programming through automation and interactivity.Many outputs are created primarily for internal review and never appear in the CSR. By using teal dashboards to handle such exploratory analyses,

teams can streamline reporting processes and focus efforts on high-value, submission-critical deliverables. This supports a leaner SAP and a more efficient, scalable approach to clinical trial reporting.

Ind_SP_PPT_042: Comparison of SAS Unit Validation and Unit Validation in R

Drisya CA, Rahul
Others

Reliable code is crucial in clinical trial programming, where regulatory compliance and reproducibility are paramount. Validating code is essential as it ensures quality, minimizes errors, and improves efficiency. Programmers typically use SAS and R in clinical programming, with SAS primarily serving regulatory reporting purposes and R being used for exploratory analysis and data visualization.

SAS Unit® is a unit testing framework for SAS® macros that manages the execution of test scenarios and generates documentation in HTML format. In contrast, R provides structured testing for functions through packages like testthat, enabling automation, test coverage, and integration with modern development workflows. Early unit validation improves traceability, maintain code integrity, and prepare for regulatory inspections.

This paper discusses the validation processes of R functions and SAS macros. It highlights the key differences in their purposes, limitations, and opportunities for improvement while offering practical recommendations to enhance validation strategies.

Ind_SP_PPT_044: Translational Biomarkers as Key Drivers of Clinical Outcomes in Oncology: Insights from Next-Generation Sequencing and Multi-Omic Profiling

Vikram Venugopal
Pfizer

Exploratory biomarker analysis integrates data from early and late phase oncology trials to deliver translational insights for scientific presentations (AACR, ASCO, ESMO) and future publications. Using Next-Generation Sequencing and Multi-Omic profiling (Genomic, Transcriptomic, Proteomic), we analyse tumour tissue and liquid biopsy samples longitudinally to evaluate treatment-related biological changes.

In early-phase trials, we observe biomarkers indicating target engagement and pathway modulation, supporting pharmacodynamic activity and potential predictive value. Late-phase analyses reveal dynamic biomarker shifts associated with clinical outcomes, resistance mechanisms, and tumour evolution. Integrated data highlights signatures related to immune activation, DNA damage response, and adaptive resistance.
Although not included in the Clinical Study Report, these findings provide critical translational context. This work demonstrates the value of embedding biomarker research across development stages to inform trial design and guide precision oncology strategies. Key results will be presented through visual data displays, case examples, and comparative analyses to emphasize clinical relevance.

Ind_SP_PPT_045: Implementation of 3.4 updates with automated specifications- A Scalable Approach

Govinda Swamy Bhavana Sruthi
Pfizer

During the SDTM IG 3.4 upgrade, we automated the development and maintenance of SDTM specifications using a standardized Diff file aligned with IG 3.4 updates. This approach enabled the systematic identification of changes, including variable additions, label and core updates, dataset renaming, variable migrations from SUPP to parent domains, and the deprecation of MO and PR domains. To streamline this complex and error-prone process, we developed an automated solution that retrieves and updates specifications accordingly. Post-automation, all specifications were validated using SAS programs. Minor issues such as naming inconsistencies and file corruption were resolved manually. This automation significantly enhanced efficiency, reducing the specification development timeline from two months to one week while improving accuracy and consistency.

Ind_SP_PPT_049: The Power and Pitfalls of Real-World Evidence

Sibi Chakkaravarthi M
Precision For Medicine

Real-World Evidence (RWE) plays a crucial role in statistical programming, providing insights from real-world data (RWD) sources such as electronic health records, claims data, and patient registries. It continues to enhance regulatory decision-making, and clinical practice improvements. Despite its potential, handling RWE poses challenges, including data heterogeneity, missing values, unstructured formats, and lack of standardization. These issues complicate data integration, analysis, and interpretation. Additionally, concerns around bias, confounding, and patient privacy require careful methodological and ethical considerations. The dynamic nature of real-world data also demands continuous updates and version control, adding complexity to reproducibility. Furthermore, aligning RWE studies with regulatory expectations across global agencies remains a significant hurdle. This paper discusses the benefits, applications, and practical difficulties of working with RWE, emphasizing the need for robust statistical techniques and transparent practices to ensure reliable outcomes. In this session, we will also be discussing practical solutions to navigate these challenges and effectively leverage RWE to optimize development strategies.

Ind_SP_PPT_050: Unlocking Data Provenance: The Key to Uniqueness and Stability for Data Review

Tanushree Shrivastava
Janssen (Johnson & Johnson)

The incremental clinical data entry presents significant challenges in accurately identifying changes necessary for medical reviews. Additionally, post-production modifications disrupt the distinction between reviewed and unreviewed data, complicating the review process and potentially impacting data integrity and patient safety. Addressing these challenges is essential for maintaining the efficacy of CDM and review workflows. Therefore, data

provenance, referring to the origin and lineage of records, is vital in CDM. To accelerate the availability of clinical data and enable near real-time visualizations, an in-house data model was developed based on SDTMIG and established business rules. After an initial setup, this model automates outputs and feeds data daily into the visualization tool. This model incorporates a mechanism to create a stable and unique record ID that remains constant throughout the study, aiding medical reviews with review flag assignments. Unlike the natural keys in SDTM, the process uses more stable keys for unique ID creation, managed by a dedicated set of jobs. A final evaluation report highlights data changes, facilitating data acceptance or rejection based on predefined thresholds and stabilizing the review process as data volume grows.

Ind_SP_PPT_054: Explore More with dataviewR: The Next Step Beyond View()

Madhan Kumar N
ICON Clinical Research India Pvt. Ltd.

Efficient data exploration is essential in clinical research to derive insights and ensure reliable downstream analyses. While R provides the traditional View () function for inspecting data frames, this approach remains limited to static exploration without interactivity or reproducibility features.To address these limitations, I have developed **"dataviewR"**, an R package provides an interactive and feature-rich environment for exploring and analyzing data.frame and tibble objects. It allows users to **filter**, **select**, and **summarize data** visually without coding. One of its key features is the automatic generation of **reproducible dplyr code**, ensuring that all data manipulations can be seamlessly integrated into analysis pipelines.Additionally, the package displays detailed **attribute metadata**, such as variable types and labels, supporting comprehensive data understanding. Designed for both novice and experienced R users, **dataviewR** streamlines the exploratory data analysis process, enhances reproducibility, and promotes efficient handling of clinical data.The package has been published in CRAN. For more details, please visit CRAN: Package dataviewR.

Ind_SP_PPT_057:My Journey into Open Source with AdmiralMetabolic : From Learner to Contributor

Siddhesh Pujari
Novo Nordisk Service Centre India Private Ltd.

In the evolving landscape of R-based clinical reporting, Admiral provides a modular framework for ADaM dataset creation. However, its functionality remains sparse in specific therapeutic areas like metabolic studies, where tailored derivations and documentation are still needed. To address this gap, AdmiralMetabolic was developed through a cross-company collaboration between Novo Nordisk, Roche, Boehringer Ingelheim, and Novartis. As a first-time open-source contributor, I joined this initiative to help build reusable functions and robust templates for creating ADaM datasets such as ADVS, ADCOEQ, and ADLB. In this talk, I'll share my learning journey—from onboarding with Admiral to contributing validated code and documentation. A key focus will be how AdmiralMetabolic simplifies and standardizes workflows previously handled by manual code. I'll also share what's coming in version 0.2.

Ind_SP_PPT_059: Visualizing Patient Journeys: The Role of Sankey Diagrams in Clinical Research

Shwetha Padmanaban, Archana V
Zifo RnD Solutions

Longitudinal graphs have long been used in clinical research to track patient progress over time, capturing changes at defined intervals. However, they often fall short in representing the complexity of treatment pathways, especially in large datasets with varied therapeutic journeys. These limitations can lead to cluttered visuals and hinder clear interpretation.

Sankey diagrams present a compelling alternative, offering a more intuitive and visually engaging way to depict patient flow across multiple treatment stages/timepoints. They excel at illustrating transitions and outcomes, making them valuable for understanding complex clinical trajectories. Nonetheless, Sankey diagrams are not without challenges - particularly when handling missing data, a frequent issue in clinical studies.

This presentation showcases examples using test data to demonstrate the Sankey diagrams, explores strategy for managing incomplete data, and highlights their limitations. Practical guidance is provided to help researchers interpret Sankey diagrams effectively, ensuring they are used to their full potential in clinical contexts.

Ind_SP_PPT_060: Closing the Gap: Boosting SDRG Drafting with R Shiny

Rathika Muthu, Sneha R
Zifo RnD Solutions

While Pinnacle Enterprise offers efficient SDRG template generation, its subscription model can be a barrier to accessibility. To address this, we developed a user-friendly tool, Gen-DRG using the R Shiny framework that enables the creation of SDRG drafts without licensing constraints. This tool accepts a wide range of inputs, including SDTM specifications in Pinnacle format, define.xml files, annotated CRFs (aCRF), XPT datasets, and the latest Pinnacle documents with reviewer comments. One of its key innovations is the automatic generation of an acronym list, which users can customize or edit manually. It also supports regulatory authority selection (FDA or PMDA), tailoring the output accordingly.

The tool intelligently identifies and inserts missing QNAMs, ensuring a more complete and accurate draft. Compared to the PHUSE template, this solution significantly reduces drafting time and enhances quality, offering a cost-effective alternative that rivals - and in many ways surpasses - Pinnacle Enterprise's capabilities.

Ind_SP_PPT_062: D-Pack: A Scalable Solution for Efficient m5 Delivery Systems

Gayathri Ayyaru, Priyatharsini M
Zifo RnD Solutions

M5 is a core component of eCTD submissions to both the FDA and PMDA, encompassing the clinical documentation and datasets required for regulatory review of safety and efficacy. Manual structuring of the M5 package increases the risk of errors and inconsistencies. Our D-Pack tool, addresses this risk by automating the creation and organization of the M5 package, ensuring consistency and compliance.This robust SAS

macro tool compiles essential components such as raw datasets, vendor files, annotated CRFs, SDTM and ADaM specifications, SAS datasets, XPT files, TLF outputs together with CRT packages, comprising define XML files and regulatory review guides for ADaM and SDTM datasets (ADRG and SDRG). With built-in automated validation features, the SAS macro facilitates precise, timely, and comprehensive submissions. This not only enhances operational productivity but also significantly reduces the chances of common manual errors, such as missing files or incorrect document placement.

Ind_SP_PPT_068: One Command to do it All: Why to do It Manually When Bash Can't

Kunal Yadav
Novartis Healthcare Pvt. Ltd.

What if you could replace the hours of repetitive program checks with a single command? In the world of clinical programming, running code, checking logs for errors or warnings, and verifying output listings can feel like a grind — until automation steps in.

This presentation showcases a simple yet powerful Bash-based solution that handles it all in one go: it runs your programs, scans logs for issues, detects mismatches in outputs, and summarizes the findings — all automatically. Even better, the script can be scheduled to run at specific times, ensuring routine checks happen hands-free and on time.

The result? Faster delivery, fewer manual errors, and more time to focus on what truly matters: insights and innovation. We'll explore practical examples, benefits for both small and large studies, and how this approach can transform your workflow.

Prepare to rethink what's possible with just one scheduled line of code.

Ind_SP_PPT_070: Teal: An R Shiny Framework to Unlock the Power of Interactive Data Exploration

Sabharish Sekhar
Inference

Teal is an open-source R/Shiny framework tailored for interactive clinical data exploration and reporting. It enables users to **dynamically filter** CDISC-formatted trial data, generate visual outputs (e.g., Kaplan-Meier, forest plots, adverse-event tables), and view the underlying R code for full reproducibility. **Modular components** from **teal.modules.clinical** support standard clinical analytics like summary tables, regression models, patient profiles via intuitive UI panels. Users can compile and export results into regulatory-ready reports (HTML, PDF, R scripts) with the help of built-in **teal.reporter.** This framework empowers clinical statisticians and programmers to **interactively explore data**, ensure **traceable analysis**, and streamline formal trial reporting workflows.

Ind_SP_PPT_072: Statistical Programming Outputs Comparison and Reporting Using Python

Vinayak Mane
Inference

This paper presents the **"Outputs Comparison Tool"** an automated Python-based executable (.exe) that streamlines the comparison of clinical trial outputs across different versions. Designed for statistical programmers, the tool addresses the challenge of manually reviewing iterative updates to tables, and listings (TLs), which often result from evolving analysis needs.It automates output tracking and comparison, offering key features such as: **File count comparison, Table of Contents alignment, Highlighted content differences, Consolidated comparison report with page references**. User-friendly and efficient, the tool supports standard file formats (.rtf, .txt, .doc, .docx, .pdf) and requires no programming expertise. It enhances productivity, accuracy, and traceability in output review workflows.

Ind_SP_PPT_073: Speaking R with a SAS Accent: Translating DM Table Creation from SAS to R

Devprakash Manoharan
Symbiance Pvt. Ltd.

As the pharmaceutical and clinical research industry embraces open-source technologies, the R programming language is emerging as a powerful complement to traditional SAS workflows. However, many clinical programmers with a strong SAS background face a steep learning curve when transitioning to R, often due to differences in syntax, package ecosystems, and data handling paradigms. This session aims to demystify R for SAS programmers by providing a practical, side-by-side comparison of common clinical programming tasks in both languages. Programmers will gain hands-on knowledge of popular R packages such as **tidyverse**, **tidyr**, **dplyr**, **pharmaversesdtm**, **pharmaverseadam** and learn key steps such as deriving variables (mutate(), case_when()), merging datasets (left_join()), and summary statistics (summarize()) are demonstrated with side-by-side code comparisons. Helping SAS programmers speak fluent R - so you can juggle both languages without dropping the data, and ace clinical analyses without any compromise.

Ind_SP_PPT_074: Automating TLF Generation Using Metadata-Driven Workflows in R

Mustafa Bookseller
Orixyon

This presentation introduces a powerful, metadata-driven approach to automating the generation of Tables, Listings, and Figures (TLFs) in R. By decoupling business logic from statistical programming, this method enables scalable, reproducible, and efficient TLF creation across multiple studies. A centralized metadata sheet drives the entire process—

governing dynamic summary and frequency table generation, conditional filtering, and customized formatting. With built-in flexibility for titles, footnotes, and numeric precision, the framework ensures consistency while significantly reducing manual effort. Real-world case studies and reusable macros illustrate how metadata can streamline and standardize clinical trial reporting workflows through intelligent automation.

Ind_SP_PPT_077: Next-Gen Transport Standards: Dataset-JSON from a Programmer's Lens

Naveen Reddy P, Venkata Koteswar Rao
Merck KGaA

The rapid advancement of AI and digital transformation is prompting a reassessment of traditional data submission methods in clinical research. In response, regulatory bodies such as the FDA and CDISC have introduced Dataset JSON-a modern transport format designed to enhance efficiency, improve metadata transparency, and enable seamless integration with evolving technologies. A key benefit of Dataset JSON is support for direct dataset access via Define.xml and enhanced human readability. This paper offers a programmer-centric analysis of Dataset JSON, comparing it with legacy formats like SAS XPT. It examines common implementation challenges, including schema, legacy system compatibility, and limited tool support, while proposing practical mitigation strategies. Tools such as Pinnacle 21 are discussed for validation and compliance. Additionally, the paper outlines practical approaches to generating Dataset JSON using SAS or R, equipping programmers with hands-on guidance. The objective is to facilitate a smooth transition toward more efficient, future-ready clinical data submissions.

Ind_SP_PPT_080: Mapping of Analysis Output Tables and Figures in ADRG by using Python

Sonali Kumbhar
Inference

The Analysis Data Reviewer's Guide (ADRG) plays a critical role in regulatory submissions by providing clear documentation and traceability of analysis datasets and outputs. This presentation focuses on the automated mapping of analysis output tables and figures into the ADRG using Python into a structured ADRG format in section 7.2 Analysis Output Program. The objective is to ensure transparency and traceability between the statistical outputs and their documentation in the ADRG. Using Python, output tables and figures are systematically identified, imported, and cross-referenced with the corresponding sections of the ADRG. This approach enhances consistency, reduces manual effort, and minimizes the risk of documentation errors. The resulting mapping improves the traceability between analysis outputs and their corresponding SAS programs, supporting regulatory transparency and review efficiency.

Ind_SP_PPT_081: Challenges in Rare Disease Trials Using Cell & Gene Therapy: A Statistical Perspective

Mrityunjay Kumar
Ephicacy Lifescience Analytics

Rare disease trials using cell and gene therapy (CGT) face complex challenges due to limited patient populations, ethical considerations, and high treatment costs. Traditional trial designs are often unsuitable, requiring innovative statistical approaches to generate reliable evidence. This presentation talks about key issues such as small sample sizes, lack of control groups, long-term efficacy tracking, and regulatory uncertainties. Emphasis will be given on advanced methodologies including Bayesian inference, use of real-world data (RWD) as external controls, adaptive trial designs, and surrogate endpoints. Strategies for addressing data sparsity and ensuring robust statistical inference under regulatory constraints will be discussed. It will be shown that how these approaches have supported successful CGT approvals in rare indications. Attendees will gain a deeper understanding of how programmers and statisticians can play a critical role in optimizing CGT development strategies for rare diseases.

Ind_SP_PPT_082: Regulatory Bodies & Approval Process in Clinical Research (India)

Nagarjuna Reddy Badduri

Fortrea

In this presentation, I would like to present the details of Regulatory Bodies & Approval Process in Clinical Research (India), Why Are Regulatory Approvals Important in Clinical Trials, I will give brief about Who Oversees Clinical Trials and Key Regulatory Bodies in India like CDSCO, DCGI, ICMR, ECs and PvPI. In this we will discuss about Clinical Trial Approval Process in India Timeline for Regulatory Approvals, Challenges in the Indian Regulatory System and Recent Developments & Future Trends.

Ind_SP_PPT_085: Interactive PK Graphs Simplified: Building Dynamic Visualization Tools Using R Shiny

Pradeep Subramaniam
Inference

Pharmacokinetic (PK) data plays a key role in understanding drug behavior over time. Traditionally, generating PK concentration-time plots from clinical trial data involves manual coding and static outputs. This presentation introduces an interactive and flexible approach using R Shiny to create dynamic PK visualizations directly from ADaM datasets, such as ADPC. We will demonstrate how users can upload SAS datasets, apply filters (e.g., treatment arms, analysis flags), and select variables to instantly generate informative PK graphs—both individual subject profiles and summary plots (mean ± SD). This eliminates the need to rewrite code for each plot and empowers users to explore data in real time. The session will include a step-by-step walkthrough, practical tips, and a reusable template to help clinical programmers and statisticians build their own visualization tools. Whether for quality checks, exploratory analysis, or reporting, this approach simplifies PK graphing and enhances data review efficiency.

Ind_SP_PPT_087: Using Python, R, and SAS Together in Clinical Programming

Rakesh Eppa
Inference

In today's dynamic clinical research environment, relying on a single programming language can limit efficiency and flexibility. This presentation highlights the advantages and practical strategies of using **Python, R, and SAS together** to build integrated, streamlined

workflows in clinical programming. Each tool offers unique strengths—**SAS** for regulatory submissions and CDISC standards, **R** for advanced visualizations and statistical modelling, and **Python** for automation, data ingestion, and process optimization. Using these languages together makes it easier for teams to work quickly, get useful insights, and handle the new challenges in clinical trials.The session will showcase real-world examples of hybrid workflows, discuss key integration methods, and outline best practices for maintaining **traceability, validation, and regulatory compliance**. Attendees will gain practical insights into creating efficient, multi-language pipelines that align with industry standards and foster innovation in clinical programming.

Ind_SP_PPT_088: Pharmaverse Packages: Simplifying R Programming for Clinical Programmers

Nikita Raut
Inference

In the clinical research industry, many professionals use R and its packages to develop automation tools and manage clinical trial data efficiently. One widely recognized ecosystem is **Pharmaverse**, which offers valuable resources particularly beneficial to SAS users transitioning to or integrating R into their workflows.Pharmaverse provides a suite of R packages designed to support standardized clinical trial data workflows, including SDTM, ADaM and summary report generation. These tools simplify data handling, promote consistency, and support regulatory readiness. By automating repetitive and complex tasks, Pharmaverse helps reduce manual errors and improves overall efficiency.

The ecosystem also includes specialized packages such as **sdtmchecks** and **sdtm.oak**, which assist in identifying issues in raw datasets and validating data structures early in the process.This presentation showcases the effectiveness of Pharmaverse in conducting robust raw data checks, facilitating the creation of standardized datasets, and streamlining the submission process.

Ind_SP_PPT_090: What is Analysis Results Dataset (ARD)? How to Create ARD using R cards package?

Shweta Rai
Inference

The Analysis Results Standard (ARS) is a CDISC standard for organizing and describing analysis results data, with the goal of improving automation, reproducibility, and reusability. An Analysis Results Dataset (ARD) is the structured dataset that conforms to the ARS, encoding statistical analysis outcomes in a machine-readable format. Essentially, the ARS provides the framework, while the ARD is the specific implementation of that framework.

The **cards** [CDISC Analysis Results Standard] R package supports the creation of ARD. In this presentation, we will demonstrate how the ARD approach is more efficient than the conventional method of presenting analysis results by using one of the approach supported by R '**cards**' package.

Ind_SP_PPT_091: Innovative Automation for Robust QC Processes of Clinical Data Packages

Raju Vankamaddi
Inference

Delivering high-quality clinical outputs demands reliable and systematic quality control. This paper presents an automated SAS macro that streamlines oversight for Study Leads by offering real-time visibility into deliverable status and quality metrics. The macro generates an Excel summary with multiple sheets covering: one track highlights task completion, program dependencies, and output status for SDTM, ADaM, and TLFs, including zero-observation datasets. It automates validation by comparing production and QC outputs, ensures correct execution order, flags pending validations, and detects discrepancies across domains, records, values, attributes, and dates along with an interactive dashboard. The second track provides detailed log reviews with chronological comparisons, warnings, issues, SAS log status, and a dashboard summary. This dual-level tracking facilitates efficient monitoring and rapid issue resolution, giving Leads a consolidated view for informed decisions. Customizable and cost-effective, the macro integrates smoothly into existing workflows unlike commercial tools bridging operational gaps without additional investment. It cultivates continuous learning and encourages programmer engagement in tool enhancement. More than just a QC tool, this macro is an enabler of efficiency, confidence, and excellence in clinical programming.

Ind_SP_PPT_093: Automating the Repetitive: How AI Tools Are Reshaping Day-to-Day Activities in Statistical Programming Workflows

Muhammad Alshad K, Kanika
ICON Clinical Research India Pvt. Ltd.

This presentation proposes a visionary custom GPT based AI assistant designed to support TLF review in clinical trial workflows. Grounded in CDISC standards (ADaM, SDTM), the assistant conceptually performs intelligent comparisons between mock shells and outputs - conducting reviews that integrate traceability, logic validation, controlled terminology alignment, and compliance with regulatory frameworks. It accepts input formats such as RTF, DOCX, PDF, and TXT, and identifies issues related to formatting, footnotes, sorting, and standard adherence. This concept remains exploratory, developed with sensitivity to regulatory expectations, data privacy, and ethical review principles. Though not deployed in

production environments, the assistant models how AI could transform QC in statistical programming by automating repetitive review tasks while preserving traceability. The session will present its conceptual design, future application potential, and invite discussion on AI's evolving role in compliant, secure programming settings.

Ind_SP_PPT_095: Beyond Traditional Controls: Exploring the Potential of Synthetic Control Arms

O P Nishana
ICON Clinical Research India Pvt. Ltd.

Traditional randomized controlled trials (RCTs) remain the gold standard in clinical research but present ethical and practical limitations, particularly in trials involving placebo control or standard-of-care comparators. Placebo-controlled designs may delay access to existing approved treatments, prolong suffering, and increase the emotional and physical burden on participants. These challenges are compounded by recruitment difficulties, retention issues, extended timelines, and high costs. Synthetic Control Arms (SCAs), which use historical trial data or real-world data to create comparator groups, offer a promising alternative. SCAs can reduce the need for placebo groups, thereby improving trial efficiency, minimizing patient risk, and ensuring ethical standards are maintained. In pediatric, rare disease, and oncology trials—where timely access to treatment is critical—SCAs can significantly enhance trial design by balancing scientific rigor with participant welfare. This approach has the potential to transform the conduct of clinical research in vulnerable populations by promoting ethical integrity.

Ind_SP_PPT_110: Bridging SAS and R: A Practical Framework for Clinical Trial Automation

Mohan Palanisamy
ICON Clinical Research India Pvt. Ltd.

As cinical trial data becomes increasingly complex, the need for automation, reproducibility and efficiency continues to grow. This session presents a practical framework for integrating R programming into SAS-based workflows, enhancing flexibility and productivity in clinical trial programming.While SAS remains the industry standard for generating regulatory deliverables such as SDTM, ADaM, and TLFs, R offers advanced capabilities for data visualization, dynamic reporting, and statistical analysis. Tools like tidyverse, ggplot2, rmarkdown, haven, and SASPy support seamless interaction between both environments.Real-world examples highlight how this integration reduces manual tasks, accelerates reporting, and promotes traceability. The approach enables automated document generation, reusable scripts, and version-controlled workflows that align with regulatory standards.By combining the strengths of SAS and R, this framework empowers

programmers and biostatisticians to build efficient, future-ready workflows for modern clinical development.

Ind_SP_PPT_112: Exploring Biomarker: ADaM - Challenges and Solutions in Hematology Trials

Sugunesh Shivalingan
ACL Digital

Biomarker data play a critical role in hematology clinical trials by providing insights into disease progression, treatment response, and patient stratification. The development of Analysis Data Model (ADaM) datasets for biomarker endpoints requires a deep understanding of both clinical and molecular data. This abstract presents the key considerations and methodologies involved in ADaM programming for biomarker data in hematology trials. It outlines the transformation of raw biomarker data into ADaM-compliant structures, such as BDS (Basic Data Structure), while ensuring traceability, consistency, and regulatory compliance. Special emphasis is placed on integrating longitudinal biomarker measurements, aligning visit-level data, handling limits of quantification (LOQ), and mapping biomarker types (e.g., genetic mutations, cytokine levels, flow cytometry markers) to meaningful analysis-ready variables. Programming strategies for linking biomarker data with clinical outcomes such as overall survival, progression-free survival, and response rates are also discussed. This work highlights how robust ADaM programming can support exploratory and inferential analyses, thereby enhance data integrity and accelerate biomarker-driven decisions in hematology research.

DS_PPT_003: Architecting Intelligence for Clinical Analytics

Likith GK
AstraZeneca

Background: Cardiovascular clinical trials face significant challenges in selecting appropriate statistical methods from over 75 available SAS procedures for cardiovascular risk management (CVRM) analysis. This complexity often leads to suboptimal analyses and delayed insights for critical endpoints including MACE, stroke, and myocardial infarction.

Innovation: We developed an intelligent web-based platform built with modern technologies (React, Node.js, PostgreSQL) that transforms statistical method selection through an intuitive wizard interface. The system guides researchers through therapeutic area selection, endpoint specification, and automated statistical method recommendations based on study design and data characteristics. By integrating comprehensive procedure documentation, ready-to-use SAS code templates, and educational resources, our platform bridges the gap between statistical complexity and practical application.

Real-Time Impact: Our platform demonstrates how intelligent decision support systems can revolutionize clinical trial statistical programming. The system significantly reduces analysis time while improving statistical rigor, enabling both novice programmers and experienced statisticians to deliver high-quality cardiovascular analyses with confidence. This innovation enhances efficiency and accuracy in complex cardiovascular research, accelerating time-to-insight for critical drug safety decisions.

DS_PPT_004: Revolutionizing Clinical Query Management with AI-Driven Automation

Priyam Sarkar, Soma Bhadra
Eli Lilly Services India Pvt. Ltd.

The integration of advanced artificial intelligence (AI) and automation technologies into clinical data management systems presents a transformative approach to enhancing operational efficiency and accuracy. Traditional query management workflows are characterized by labor-intensive processes, including manual logging of actions, data verification, drafting of query text, and iterative follow-up until resolution. These processes are not only time-consuming but also susceptible to human error.

Our technological solution employs AI-driven tech to automate the identification of data discrepancies and the generation of precise query text. Additionally, integration technologies facilitate seamless connectivity with electronic data capture (EDC) systems, enabling automated query posting and real-time data verification through secure login mechanisms. The implementation of these technologies is projected to significantly reduce manual effort, improve data accuracy, and optimize the efficiency of clinical trials, thereby expediting the query resolution process and minimizing errors. Additionally, powerful analytics and

visualization tools are employed to monitor and analyze query performance and data integrity in real-time.

Key components of the solution include a Clinical Data Query Processor, which refines query text using standardized language and metadata, and an CDMS based integration API that ensures seamless interoperability between disparate systems. This strategic implementation aims to revolutionize clinical data management practices, fostering a more efficient and error-free environment.

DS_PPT_008: Enhancing R Shiny Application Performance Using data.table::fcase()

Dineshkumar Subbu
Ephicacy Lifesciences Analytics

R Shiny is widely used for developing interactive web applications in the clinical domain however performance optimization remains a key challenge, especially when managing large datasets or complex conditional logic. This presentation explores the use of fcase(), a function introduced in the data.table package, as an efficient alternative to traditional ifelse() or nested case_when() statements within Shiny applications.
By leveraging the vectorized and optimized nature of fcase(), measurable improvements in data processing speed can be achieved. Through real-world examples and benchmarks, we highlight how fcase() simplifies conditional logic while reducing execution time, particularly in reactive environments where performance bottlenecks can degrade user experience. With the help of this presentation, the user aims to provide practical insights and code-level comparisons, offering best practices for integrating fcase() into Shiny workflows. Developers and data scientists working with large-scale R Shiny applications will benefit from these techniques to build faster and more scalable solutions.

DS_PPT_009: Empowering Clinical Research Teams: Efficient Pre-Screening Using Epic Slicer Dicer and EDC Technology

Shruthi Anthay
Ephicacy Lifescience Analytics

In the realm of clinical trials, patient recruitment and pre-screening are crucial steps that determine the success or failure of a study. The efficiency with which clinical trials identify eligible participants can significantly impact the trial timeline, cost and overall success. Traditional methods of pre-screening are time-consuming, prone to error, and require substantial manual work. However, using advanced technologies like Epic Slicer Dicer and Electronic Data Capture (EDC) systems has transformed the way healthcare institutions conduct patient pre-screening for clinical trials. By automating and streamlining the process, these tools not only make patient selection faster and more accurate but also ensure that the right candidates are enrolled in studies, ultimately improving the quality of clinical trial outcomes.

Slicer Dicer is a data exploration tool in the EPIC EHR that allows one to customize searches on large patient populations. It is a self-service analytics tool for cohort discovery and feasibility analysis

DS_PPT_011: Integrating DevSecOps Principles into R for Clinical Data Programming: Enhancing Compliance, Reproducibility, and Efficiency

Indraneel Chakraborty
Ephicacy Lifescience Analytics

In this evolving landscape of working with clinical data, the integration of DevSecOps—combining development, security, and operations into R programming has become increasingly critical. As developers in the pharmaceutical domain handle sensitive clinical trial data, ensuring reproducibility, confidentiality, and compliance with regulatory standards such as 21 CFR Part 11 and GxPs is critical. This talk explores the technical relevance and practical value of adopting DevSecOps principles in R-based developments. Real world references will be utilized to explain the integrations.

Key practices to be discussed include version control with Git, automated checks, CI/CD workflows using platforms like GitHub Actions, secure credential management in cloud workspaces, and environment reproducibility with {renv}. Additionally, containerization tools like Docker enable secure, isolated execution environments that enhance data integrity and auditability.

Attendees will gain understanding on embedding DevSecOps into R-based clinical workflows; ultimately boosting efficiency, scalability, and compliance readiness in statistical programming.

DS_PPT_012: Synthetic Clinical Trial Data Generation in R Using Bayesian Modelling: Preserving Uncertainty and Clinical Fidelity

Liza Monica
Ephicacy Lifescience Analytics

Access to clinical trial data remains limited due to patient privacy and regulatory concerns, hindering secondary research and innovation. Synthetic data generation offers a solution by simulating realistic datasets that retain statistical fidelity without exposing individual-level information. This work introduces a Bayesian approach using R packages such as rstanarm and bayesplot to generate synthetic clinical trial data. Unlike frequentist or rule-based methods, the Bayesian framework explicitly incorporates prior knowledge and quantifies uncertainty, resulting in transparent, reproducible simulations of treatment groups, endpoints, and adverse event patterns. A case study on an oncology trial demonstrates how time-to-event data and treatment effects can be accurately modelled and validated. The probabilistic nature of Bayesian inference ensures clinically meaningful synthetic data that align closely with real-world distributions. This approach not only advances methodological rigor but also supports regulatory readiness, open science, and safe data sharing—addressing key challenges in data access, reproducibility, and transparency.

DS_PPT_014: Empowering Clinical Trials: Real-Time Hepatic Safety and Endpoint Detection for Enhanced Patient Care

Pradnya Kadam
Eli Lilly Services India Pvt. Ltd.

Case Study 1: Safety Monitoring:

- Ensuring hepatic safety is crucial for the success of clinical trials. Manual reviews are often protracted and fallible leading to delays and inaccuracies in hepatic safety monitoring and reporting.
- An interactive dashboard can track lab values and abnormalities in real-time, integrating data from labs, patient history, medications and events. This comprehensive view helps in identifying early signals and assessing liver-related risks effectively.

Case Study 2: Endpoints Identification:

- In Outcome trials, worsening symptoms or signs related to the targeted disease are significant to participants and necessitate intensified treatment. These critical endpoints can be overlooked if only urgent/hospital visits associated with events, including death, are considered.
- The dashboard offers a comprehensive review of multiple datasets like medical history, adverse events, medications, and queries to provide early detection and reporting of endpoints. It is key to the success of the outcome trial and ensures patient's safety.

Conclusion:

The field of data science is evolving, and so are dashboards. Next-generation dashboards should offer more than just a visual summary; they must be real-time, interactive, and actionable.

DS_PPT_016: Biomedical Event Classification Using Bio BERT & Python: A Deep Learning Perspective

Gayathiri Premkumar
Sensan Biosciences

Accurate classification of adverse events (AEs) from patient reviews is vital for healthcare monitoring and clinical decision-making. This paper employs **BioBERT**, a biomedical language model, within a **Machine Learning Model (MLM)** framework to classify patient-reported text into **Drug related event** and **non-event** categories. Using **Python**, unstructured textual data is transformed into numerical representations through tokenization, embedding extraction, and contextual feature learning. Fine-tuning **BioBERT** on medical datasets enhances the model's ability to capture adverse event indicators. Performance metrics, including **accuracy, precision, and recall,** validate the classification effectiveness, improving automated AE detection for clinical research and patient safety assessment. This approach is particularly useful for pharmacovigilance, AE analysis, and automating structured data extraction from clinical narratives.

DS_PPT_017: Integrating GenAI in RShiny for PubMed API Querying via rentrez()

Namrata Pant

The R package rentrez provides a powerful and flexible interface to the National Center for Biotechnology Information (NCBI) databases, enabling efficient retrieval of biomedical literature and related data. When paired with Generative AI (GenAI) tools for search query formulation and data extraction, rentrez streamlines critical steps in evidence synthesis. GenAI assists in crafting optimized search queries and extracting relevant information, allowing researchers to swiftly gather abstracts, full texts, and metadata from PubMed and other NCBI repositories with reduced manual effort.

This presentation introduces rentrez's key functionalities (search, fetch, summary, and link) and demonstrates how GenAI-augmented keyword generation improves PubMed search processes for implementing transparent and efficient evidence-gathering pipelines.

DS_PPT_018: Genomics Findings

Ruba Sivasamy
Sensan Biosciences

A findings domain that contains data related to the structure, function, evolution, mapping, and editing of subject and non-host organism genomic material of interest. This domain includes but is not limited to assessments and results for genetic variation and transcription, and summary measures derived from these assessments. The GF domain is used for findings from characteristics assessed from nucleic acids and may include subsequent inferences and/or predictions about related proteins/amino acids. Genomics Findings (GF) domain is introduced in (SDTMIG) v3.4 which replaces PF domain from the previous version. If we are using SDTMIG v3.2 or v3.3, GF domain can be included as a custom domain, which would be a promising way to handle genetic testing and assessment results. This paper covered a small intro about Genome and list of some important variables, some GFTESTCD values from GF domain

DS_PPT_019: Enhanced Data Review with AI

Santosh Aravind Vinnakota, Viswa Prakash
Bristol Myers Squibb

Enhanced Data Review with AI project is aimed at transforming traditional data review processes using advanced AI technologies.

Purpose:

- Establish a streamlined and robust framework for data oversight.
- Ensure high accuracy and efficiency across critical and non-critical data sets.

Key Objectives:

- Integrate AI into the trial protocol for continuous monitoring and verification of essential data points.
- Optimize management of non-critical data points to reduce manual effort and enhance operational efficiency.
- Identify and address gaps within the existing data review plan using tools like the Protocol Data Review Plan (PDRP) and Edit Checks.

Goals:

- Achieve operational excellence.
- Deliver actionable outcomes, including improved monitoring efficiencies.
- Ensure compliance with standard review guidelines.

Focus: Forward-looking approach emphasizing precision, reliability, and efficiency through AI-enabled advancements.

DS_PPT_021: Julia-fy Your Data

Surabhi Chaudhari
Pfizer

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_025: Predicting Patient Dropout in Clinical Trials Using Interpretable Machine Learning on Simulated Data

Vishal V
Pfizer

Patient dropout is a persistent challenge in clinical trials, impacting timelines, cost, and data integrity. Existing predictive models often depend on real-world clinical data and yield limited interpretability, making them difficult to generalize across studies or use in operational settings. This study presents a proof-of-concept framework leveraging simulated clinical trial data to model patient dropout risk using a tree-based machine learning classifier. Key features such as visit adherence, engagement score, and comorbidity burden are used to train the model. Interpretability is achieved through SHAP (SHapley Additive exPlanations), which quantified the influence of each variable on individual predictions. The model offers early, transparent insight into dropout risk, enabling targeted interventions to improve retention. This approach demonstrates potential for scalable integration into trial planning and monitoring systems, bridging clinical operations and data science through interpretable AI.

DS_PPT_028: Operationalizing AI Credibility and Real-Time Decision Support in Clinical Trials through Data Science

Anshu Aman

As the clinical development paradigm shifts toward decentralized, data-rich ecosystems, the infusion of AI/ML into trial operations necessitates statistically robust, explainable, and regulatory-compliant data science strategies. This study presents how clinical data science (CDS) orchestrates real-time decision support via feature-engineered data pipelines, predictive signal detection, and adaptive trial intelligence. Emphasis is placed on operationalizing AI model credibility by aligning with pre-defined Contexts of Use (COU), applying algorithmic fairness metrics, performing bias mitigation, and enforcing model governance protocols including versioning and concept drift detection. The deployment of interactive analytics dashboards and auditable ML workflows supports cross-functional collaboration and regulatory transparency. By embedding these capabilities within GxP-aligned infrastructure, CDS enables scalable, trustworthy, and patient-centric innovation in modern clinical research.

DS_PPT_030: Leveraging GENai for Proactive Safety and Efficacy Monitoring Through Laboratory Parameters

ABHINAV SRIVASTAVA
Pfizer

Accurate and timely clinical diagnosis is essential for effective disease management. Many pathological conditions result from molecular and cellular changes, which are often reflected in routine laboratory parameters. These biomarkers may contain hidden patterns that are not apparent through conventional clinical assessment. However, the complexity and volume of lab data often remain underused in predictive modelling.
We present a Genai driven approach that helps uncover important patterns in routine lab test results, like blood and urine values. Our method uses Eclat Algorithm to find lab tests that often show abnormal results together. These combinations are then grouped into meaningful patient categories. Next, we use a machine learning model – XGBoost to analyze these groups and predict who might develop a disease or face complications.
Our framework aims to predict disease onset and complications earlier in the clinical workflow, enhancing personalized decision-making.

DS_PPT_032: Evaluating the Effectiveness of Hierarchical Bayesian and Reinforcement Learning Models in Optimizing Treatment Regimens for Lung Cancer Patients Using Real-World Evidence

Yash Chaudhari
Fortrea

Lung cancer treatment remains challenging due to diverse patient responses and the disease's complex progression. This study evaluates the use of Hierarchical Bayesian and Reinforcement Learning model to optimize treatment regimens using large-scale real-world data, including electronic health records, claims, and biomarker profiles. Hierarchical Bayesian models effectively capture variability across patient populations and treatment centers, offering refined estimates of survival and treatment impact. Reinforcement Learning models continuously learn from individual treatment histories, recommending therapy

sequences that adapt over time as new data emerges. Together, these models enhance prediction of progression-free survival and quality of life, supporting more personalized, data-driven clinical decisions. Our findings highlight the potential for integrating these advanced approaches into routine care, improving outcomes and tailoring treatments to individual needs. This work demonstrates a promising path forward in leveraging real-world evidence to support adaptive, precision oncology in lung cancer.

DS_PPT_033: Optimizing Clinical Data Integrity: A Dual-Path Strategy with Intelligent Automation for Pre-SDTM Validation

Sarika Selvaraj
Precision For Medicine

Maintaining data integrity in clinical trials is increasingly complex due to the rise of large-scale, decentralized studies and diverse data sources. Traditional, reactive approaches often result in inconsistencies, delays, and increased resource burden during SDTM programming. This paper introduces a dual-path solution combining data management and programming strategies to proactively address data quality before SDTM begins. From the data management perspective, deploying standardized, logic-driven electronic Case Report Forms (eCRFs) across sites ensures consistent entry and captures justifications for out-of-range values at the point of entry.

In parallel, the programming approach leverages automated, reusable scripts to detect anomalies in raw data early, minimizing rework and improving workflow efficiency. As an innovative enhancement, we propose integrating AI-driven pattern recognition to predict data quality risks and recommend resolution logic based on historical trends. This intelligent, adaptive framework streamlines clinical data pipelines, accelerates submissions, and positions teams for future-ready, real-time data validation.

DS_PPT_035: AI driven coding assistance - Safeguarding data and compliance through Human Intervention

Surendar Sekar
ICON Clinical Research India Pvt. Ltd.

As artificial intelligence (AI) becomes increasingly integrated into clinical programming workflows, the emergence of code assistance tools such as Generative AI (GenAI) ChatGPT, Gemini, GitHub Copilot provides intriguing opportunities for automation in clinical trial programming. However, their use in GCP-regulated systems presents significant questions about validation, traceability, and regulatory compliance. While these AI powered tools can accelerate operations like ADaM derivation, TLF automation, and documentation drafting, they remain non-validated systems and require additional human intervention.

This abstract explores how to align AI generated code aligns with core clinical regulations, including FDA, GCP guidelines. Furthermore, this abstract outline the practical use of AI in statistical programming, highlighting real-world examples, benefits, and risks whilst being compliant with the regulatory compliance with some case study analysis.

Finally emphasizing “AI can assist, not replace” a clear grip of when and how to employ AI responsibly, as well as when human oversight remains indispensable.

DS_PPT_036: Foundations of Machine Learning in SAS: PROC FOREST and HPFOREST Methods

Deepak Marella
ACL Digital

The integration of machine learning techniques into the SAS ecosystem has enabled robust, scalable solutions for data-driven decision making across industries. This work explores the foundational aspects of machine learning in SAS, with a focus on the PROC FOREST and PROC HPFOREST procedures. These tools implement the Random Forest algorithm, offering strong predictive performance, resistance to overfitting, and automated variable importance measures. While PROC FOREST is suitable for traditional analysis, PROC HPFOREST is optimized for large-scale, parallel processing environments, making it ideal for high-volume domains like healthcare, finance, and manufacturing. The discussion also extends to modern enhancements in SAS workflows, including integration with open-source tools and model interpretability frameworks. By combining the analytical power of machine learning with the data governance and processing strengths of SAS, organizations can build interpretable, production-ready models that comply with regulatory standards and support real-time operational use.

DS_PPT_037: Clinical Data Security Augmentation Through Next-Gen Steganography

Vinothkumar Subramanai
Fortrea

As clinical research data becomes increasingly digital, protecting sensitive patient information is more critical than ever. Previously, we explored SAS® encryption and image-based steganography to secure and conceal clinical data. This presentation introduces **video steganography** as a powerful next step. Unlike static image hiding, video steganography embeds data across multiple frames, leveraging motion and audio layers to enhance concealment, capacity, and resistance to detection. It enables not only secure storage but also covert transmission of clinical information through video files. When combined with SAS-based encryption, this dual-layered method offers a robust approach to protecting data in compliance-driven environments. This session outlines technical methods, use cases, and a prototype framework for embedding clinical data into video streams—opening new possibilities for secure, invisible, and scalable data exchange in the clinical research.

Ind_MGT_PPT_001: Revolutionizing Productivity and Innovation with Microsoft Copilot: Transformative Applications in Business, Healthcare, and Education

Linga Reddy
Syneos Health

In the rapidly evolving landscape of artificial intelligence, Microsoft Copilot stands out as a transformative tool designed to augment human capabilities. This poster explores the multifaceted applications of Copilot, an AI-powered assistant built on the GPT-4 architecture, in enhancing productivity and creativity across various domains.

Copilot integrates seamlessly with Microsoft 365 applications, providing users with intelligent suggestions, automating repetitive tasks, and offering real-time assistance. By leveraging natural language processing, Copilot can understand and respond to complex queries, draft documents, generate code, and even create visual content. In healthcare, it supports clinical documentation, patient management, and research activities, improving patient care and operational efficiency. In education, Copilot aids in curriculum development, grading, and personalized learning experiences. Overall, the integration of Copilot into different fields promises to transform workflows, enhance productivity, and foster innovation.

This Presentation will highlight the following key features of Copilot.

- **Intelligent Suggestions**: Provides context-aware recommendations to enhance productivity.
- **Task Automation**: Automates repetitive tasks, allowing users to focus on more strategic activities.
- **Natural Language Processing**: Understands and responds to complex queries in a conversational manner.
- **Document Drafting**: Assists in creating and editing documents, emails, and presentations.
- **Code Generation**: Helps in writing and debugging code for various programming languages.
- **Visual Content Creation**: Generates images and visual aids to support presentations and reports.
- **Information Synthesis**: Gathers and summarizes information from the web to support decision-making.
- **Patient Management:** Assist in scheduling appointments, sending reminders, and managing patient records.

The presentation aims to provide a comprehensive overview of how Copilot is revolutionizing the way we work and interact with technology, paving the way for a more efficient and innovative future.

Ind_MGT_PPT_002: Bridging Differences: Effective Conflict Management in Technology Tool Development

Bhushan Kulkarni
Syneos Health

The introduction of new technology-based tools often brings about conflicts among users, which can impede collaboration and the smooth adoption of these tools. When integrating novel technologies, diverse user experiences, expectations, and perspectives frequently lead to disagreements that slow progress. This paper explores effective strategies for managing conflicts during the development and deployment stages of such tools. It emphasizes the importance of promoting transparent communication, defining clear roles, and cultivating environments that encourage constructive feedback. Moreover, the paper highlights conflict resolution frameworks and decision-making models that contribute to maintaining a cooperative atmosphere. A balanced approach, incorporating both proactive and reactive methods such as mediation, user training, and inclusive stakeholder engagement, is recommended for resolving conflicts and enhancing team dynamics. By addressing these challenges, organizations can facilitate a more seamless adoption process, ultimately improving both productivity and user satisfaction.

Ind_MGT_PPT_004: Mentorship - Catalyst for growth?

Monika Rajesh
Ephicacy Lifescience Analytics

In today's evolving statistical programming landscape, technical skills alone aren't enough for long-term success. Mentorship plays a critical role in shaping competent, confident, and collaborative professionals. Though the value of mentorship is undeniable with the increasing complexity of clinical trials and rising expectations around data quality and compliance, structured mentoring is often undervalued or overlooked in fast-paced project environments.

In statistical programming, mentorship extends beyond guiding junior staff, it fosters technical excellence, ensures compliance, and preserves institutional knowledge. With growing regulatory demands, evolving tools (like R, Python, and automation platforms), and increasing complexity of clinical trial designs, transitioning from academic or entry-level coding to industry-grade statistical programming can be overwhelming without support.

Whether you are a new programmer seeking guidance or a seasoned professional looking to give back, this session will illustrate why mentorship isn't just an added benefit – it's a strategic investment in both people and quality.

Ind_MGT_PPT_005: Beyond Generations: Cultivating a Workplace of Belonging

Geetha DB, Akhila
ICON Clinical Research India Pvt. Ltd.

In the contemporary workplace, generational diversity has emerged as a defining characteristic, with Baby Boomers, Generation X, Generation Y (Millennials), Generation Z coexisting and contributing to organizational growth in distinct ways. However, this diversity brings with it complex challenges, as each generation exhibits differing attitudes, values, work expectations, communication styles, and learning preferences. The rapid retirement of Baby Boomers and the growing presence of Generation Z have intensified concerns surrounding knowledge retention and skill transfer across generations.

Organizations are increasingly facing difficulties in recruiting and retaining employees with the right mix of skills and experience, particularly in the context of rapid technological advancements and evolving work practices.

This abstract focuses on the implications of generational diversity in the workplace, with a specific focus on the skill and knowledge disparities among generations. It examines the root causes of these gaps and their impact on organizational performance and talent management. We will also deep dive into practical strategies and recommendations for professionals to foster effective intergenerational collaboration, facilitate knowledge transfer, and bridge skill gaps, ensuring a more inclusive and future-ready workforce.

Ind_MGT_PPT_006: Harnessing the Power of Modified Agile Methodologies

Stephen Prawin
Precision For Medicine

AGILE, introduced by The Agile Alliance in 2001, offers a flexible, iterative project management approach compared to the rigid, sequential waterfall model. While AGILE has transformed software development and broader project management, clinical statistical programming often remains tied to linear workflows.

This presentation introduces **MAGIC** (**M**odified **AGI**LE for **C**linical Stat Programming) to challenge this norm and promote a more adaptive way of working. We compare AGILE methodologies with current processes, explore transition challenges, and highlight practical AGILE practices that can be easily adopted to enhance efficiency and collaboration. These include:

- **Scrum**: Organizes work into short sprints with daily stand-ups to track progress.
- **Sprint Planning**: Defines sprint goals and selects tasks collaboratively.
- **Kanban**: Visualizes workflow to manage tasks and identify bottlenecks.
- **Iterative Processing**: Encourages continuous refinement of deliverables.
- **Cyclical Timelines**: Enable regular feedback and incremental delivery.

When applied effectively, these practices improve transparency, adaptability, and team performance.

Ind_MGT_PPT_008: Bridging the Industry-Academia Gap: Introducing Statistical Programming as Electives to Build an Industry-Ready Workforce

Sathishkumar Manoharan
ICON Clinical Research India Pvt. Ltd.

The clinical research industry continues to see high demand for skilled statistical programmers. Currently, most organizations hire fresh graduates and spend 3–6 months training them in SAS, clinical research concepts, and regulatory standards. While effective, this onboarding model is time-intensive and delays project contribution.

This abstract proposes a strategic academic partnership model, where selected colleges introduce **industry-relevant elective papers** aligned with statistical programming roles. The curriculum would cover **SAS and R programming**, **clinical trial fundamentals**, **basic**

statistical methods, and an **overview of CDISC standards (SDTM and ADaM)**. Course materials would be curated with input from industry experts to ensure real-world applicability.

Students trained in these subjects during college would be evaluated and hired based on academic performance, requiring only 1 month of focused corporate training. This model reduces onboarding time, builds a job-ready talent pipeline, and creates a sustainable collaboration between academia and the clinical research industry.

Ind_MGT_PPT_009: The Half-Life of Skills: Leading When Today's Expertise is Tomorrow's Obsolete

Pratibha Jalui
Cytel Inc

What if your most valuable skill today is already on its way to becoming irrelevant? In a world where the half-life of skills is shrinking rapidly, sometimes to just two years, leaders face a hidden crisis: how to stay relevant when expertise comes with an expiry date. This session will challenge the way you think about learning, leading, and letting go. We will explore why some teams fade while others thrive, and how learning agility, rather than static knowledge, is becoming the ultimate leadership currency. You will also discover practical steps to prepare yourself and your teams for the future. Through real-world case studies, surprising industry data, and modern leadership strategies like reverse mentoring and quick learning sprints, you will walk away with a blueprint to build teams that can adapt quickly and stay ahead. Indeed, in this era, your biggest risk is not falling behind. It is standing still.

Ind_MGT_PPT_011: Transforming Mindsets, Transforming Operations: The Lean Principles Advantage

Anagha Gumaste
ACL Digital

The pharmaceutical industry is undergoing a significant transformation with the rise of Decentralized Clinical Trials (DCTs). While DCTs promise enhanced efficiency and patient centricity, their successful adoption hinges on robust change management strategies. This abstract focuses on effective leadership during the transition to DCTs, emphasizing the critical need to address organizational resistance and ensure comprehensive technology adoption among clinical teams. Implementing DCTs necessitates profound cultural and operational shifts, demanding active stakeholder buy-in, targeted training programs for remote monitoring, and meticulous measurement of adoption success. By proactively managing these human and technological aspects, organizations can navigate the complexities of this paradigm shift, unlocking the full potential of decentralized trial methodologies.

Ind_MGT_PPT_014: Rescue, Recover, Repeat: Effective Management of Outsourcing Failures in Biostatistics Projects

Tony Mathew
ACL Digital

Outsourcing has become integral to many pharmaceutical companies and contract research organizations, enabling them to maintain lean operations, scale effectively and focus on core

competencies. However, this model sometimes fails when vendors miss critical deliverables in terms of both timelines and quality—leading to wasted time and resources, damaged morale, strained vendor–sponsor relationships and even delays in regulatory approvals. This abstract presents a tested, pragmatic, step-by-step framework from the vendor's perspective to identify signs of failure in outsourced biostatistics projects (e.g., SDTM, ADaM, TLFs, define.xml, reviewer guides) and outlines mechanisms to rescue or salvage such engagements. It further offers preventive strategies and process improvements to enhance the likelihood of success in future outsourcing of critical studies, promoting predictability and reliability in outcomes.

Poster Presentations

Ind_BIOS_PSTR_001: Biomarker Analysis using Composite Measure(CM)

Pooja Puniya
Parexel Pvt. Ltd.

When evaluating drugs that have nonclinical signals of kidney injury in early clinical studies, it is critical to ensure safety of volunteers, particularly if the volunteers are healthy volunteers, since healthy volunteers have no prospect of benefit from participation in study. To mitigate risk to subjects, researchers often attempt to maintain a sufficient safety margin to the dose/exposure at which renal toxicity was seen in animals; however, this may prevent development programs from evaluating doses/concentrations that are needed to achieve efficacy.

Biomarkers that may be more sensitive indicators of renal injury than current standard measures will aid in monitoring acute and sub-acute drug-induced tubular injury in clinical trials so that renal injury may be detected at an early and potentially reversible stage.

This presentation includes general considerations and detailed explanation of method on how to use CM to study presence of kidney injury demonstrated using simulated data example.

Ind_BIOS_PSTR_002: Selection and Impact of Covariance Structures on MMRM

Angela Rose Thomas
Sensan Biosciences

Mixed Model Repeated Measures (MMRM) is a powerful statistical tool in clinical informatics, enabling robust analysis of longitudinal data. A critical aspect of MMRM is the choice of covariance structure, which directly affects model stability, interpretability, and computational efficiency.

This presentation highlights key covariance structures—unstructured, compound symmetry, autoregressive, and Toeplitz—through the lens of programming implementation. It explores challenges such as convergence issues, runtime trade-offs, and software-specific optimizations in SAS, and R. By bridging statistical theory with practical coding considerations, this discussion empowers informaticians to make data-driven decisions that balance precision and efficiency in real-world applications.

Ind_BIOS_PSTR_003: Robust Strategies for Clinical Trials with Non-proportional Hazards

Srinivas S
Precision For Medicine

Traditional survival analysis methods such as the log-rank test and Cox proportional hazards model rely on the assumption that treatment effects remain constant over time. However, in

oncology trials, especially with immunotherapies, this assumption is often violated leading to reduced power and potential misinterpretation of results. To address this challenge, the FDA initiated discussions with industry statisticians to determine the most appropriate analysis strategies tailored to different NPH patterns, which led to a proposal of a robust test called Max-Combo test and a sequential approach (three-step analysis plan). This presentation briefs in detail about the three-step analysis plan for handling non-proportional hazards, emphasizing innovative statistical techniques to improve the accuracy of treatment effect estimates followed by a case study.

Ind_BIOS_PSTR_005: Determining sample size for Rheumatoid Arthritis based on the ACR 20 response using Single Arm Proportional Meta-analysis

Santosh Kumar
Siro Clin Pharm

Introduction: There are some clinical trials where sample size estimation is not robust or accurate due to the absence of previous evidence; therefore, the pooled effect such as meta-analysis of the intervention as well as placebo provides the best support with the robust data in determining the sample size. ***Methods***: a meta-analysis of randomized studies of Abatacept's effect on Rheumatoid Arthritis based on the ACR 20 response. ***Results***: the proportion of ACR 20 response for 3 months of SC Abatacept was in Common effect model is 0.6538, 95% CI [0.6261; 0.6804], and Random effects model is 0.6571, 95% CI [0.6061; 0.7047], and similarly pooled analysis of 4 studies reported that the proportion of ACR 20 response for 3 months of IV placebo was Common effect model is 0.2687, 95% CI [0.2294; 0.3121] and Random effects model is 0.2660, 95% CI [0.1971; 0.3485]. ***Conclusion***: Since there has been no head-to-head comparison between SC Abatacept and placebo, we conducted a single arm meta-analysis. This single arm analysis has paved a new path in sample size determination

Ind_BIOS_PSTR_006: Implementing Mixed Models with Repeated Measures (MMRM) in R

Nisha Borse
Sanofi Healthcare

Linear mixed models are widely used in clinical trials for analyzing repeated measures, incorporating random effects and handling missing data under the missing at random assumption.

MMRM (mixed model repeated measures) is used in clinical trials to model repeated measures over time, assuming multivariate normal residuals with an unstructured covariance matrix for flexible correlation.

For fixed effects, the model includes time, treatment group, and their interaction, creating a saturated mean model. Baseline covariates can be adjusted either as main effects or with time interactions to allow variations in outcome association.

Ind_BIOS_PSTR_007: Sample Size Determination for Vaccine Efficacy Trials with Lower Bound

Monisha Selvakumar
Emmes services Pvt. Ltd.

Vaccine efficacy trials often require large sample sizes due to low incidence rates. Interim analyses play a key role in guiding trial decisions while minimizing bias, enabling early termination, design modifications, or declaration of trail success. While many studies address vaccine efficacy and interim analysis, few focus on sample size calculations for interim analysis across varying efficacy thresholds. Integrating exact conditional tests under the Poisson assumption with O'Brien-Fleming group sequential design helps address this gap, especially when lower efficacy bounds needs to be incorporated. We employed a two-stage statistical framework combining the exact conditional test modified by Chan and Bohidar with O'Brien-Fleming group sequential boundaries to establish both efficacy and futility thresholds for interim analysis with expected number of cases. The integration of methods addresses challenges in determining optimal sample size while maintaining Type I and Type II error rates across interim looks, ultimately improving trial efficiency and reducing resource expenditure.

Ind_BIOS_PSTR_008: Adjusting for Truncated study duration in Recurrent Event Analysis: A Weighting Approach for Clinical Trials

John Michael Raj
Others

Background: In recurrent event analysis with fixed follow-up intervals, truncated follow-up due to early dropout or study termination introduces bias and reduces precision in risk

estimates, particularly in clinical trials where shorter observation periods may underestimate event risks.

Methods: We propose a time-based weighting approach using the ratio of observed-to-expected follow-up duration in the Prentice-Williams-Peterson Gap Time (PWP-GT) model. The method was evaluated in simulations and applied to a double-blinded trial (N=4000) comparing 500 mg vs. 1500 mg daily calcium supplementation for preeclampsia prevention. For demonstration of problem and application of weighting method drug non-adherence at follow-up visits taken as the recurrent event.

Results: Simulations showed the weighted PWP-GT model had lower bias (1.0% vs. 1.3%) and improved precision compared to the unweighted model, with coverage probabilities >94%. In the trial data, weighting yielded smaller standard errors and a more conservative hazard ratio for hypertension family history (weighted HR=1.14, SE=0.054 vs. unweighted HR=1.23, SE=0.065).

Conclusion: Truncated follow-up in recurrent event studies will impact the risk estimation if unaccounted for. Our findings demonstrate that total time-based weighting effectively addresses this bias and enhances precision in simulated and real datasets.

Ind_BIOS_PSTR_009: Percentile Matching Estimation of Uncertainty distribution

Anjana Viswanathan
ACL Digital

This work considers the application of method of percentile matching available in statistical theory of estimation for estimating the parameters involved in uncertainty distributions

An empirical study has been carried out to compare the performance of the proposed method with the method of moments and the method of least squares considered by (Wang and Peng, 2014) and (Liu B, 2010), respectively

The empirical study clearly establishes the superiority of the proposed method over the other two methods in estimating the parameters involved in linear uncertainty distribution when appropriate orders of percentiles are used in the estimation process.

IND_BIOS_PSTR_049: Timing Matters: A Survival Analysis of Breastfeeding Initiation Among Children Under Two in India Using NFHS-5 Data

Tejal Lakhan

Novotech

Timely initiation of breastfeeding—within one hour of birth—is critical for neonatal survival, yet many newborns in India face delays, even with rising institutional deliveries and maternal care access. Most research employs binary outcomes, obscuring the timing

and determinants of delay. This study applies a Cox proportional hazards model to National Family Health Survey (NFHS-5, 2019–21) data from 46,763 children under two, modeling time (in hours) to breastfeeding initiation. Cesarean delivery was strongly associated with delays (HR: 0.66), while institutional births slightly increased the likelihood of early initiation (HR: 1.13). Greater maternal age and antenatal care were linked to earlier initiation, but higher maternal education and wealth modestly delayed breastfeeding, indicating gaps in postnatal counseling across socioeconomic groups. Urban–rural differences were negligible after adjustment. These findings highlight the need for targeted breastfeeding support—especially post-cesarean—and integration of early initiation counseling into maternal health programs across all facility types and population groups. The use of time-to-event analysis offers a more precise lens on breastfeeding patterns, informing context-specific policy interventions.

Ind_BIOS_PSTR_054: Strengthening Observational Study Accuracy: The Role of Propensity Score Matching in Bias Mitigation

Sarani Selvakumar, Vijaya Sankar
Zifo RnD Solutions

Observational studies often encounter challenges such as confounding and selection bias, which can compromise the validity of causal inferences. Confounding arises when external variables are associated with both the treatment and the outcome, potentially distorting the estimated treatment effect. Selection bias occurs when the study sample is not representative of the target population or when comparison groups differ systematically. Addressing these biases is crucial for accurate and reliable conclusions.

Propensity Score Matching (PSM) is a statistical method designed to mitigate these issues by enhancing the precision of treatment effect estimation. This paper examines how PSM effectively reduces confounding and selection biases in observational research. It demonstrates the practical application of PSM using SAS code and compares its performance with alternative methods, outlining their respective advantages and limitations. Additionally, the paper discusses the interpretation of results derived from PSM to ensure robust and credible inferences in observational study contexts.

Ind_BIOS_PSTR_057: Unlocking the Untapped potential of External Control Arms: The transformative power of MAIC

Rahul Misra
Bristol Myers Squibb

For decades, randomized controlled trials have been the gold standard for evaluating treatment efficacy. However, as medicine evolves and patient-centric designs gain prominence, traditional methods often struggle with logistical, ethical, and financial

challenges especially in cases of limited patient recruitment, rare diseases, or the absence of direct treatment comparisons.

External control arms (ECAs) offer a viable and transformative solution by leveraging curated trial datasets and real-world evidence to generate meaningful comparisons without necessitating a fully randomized control group. Among the methodologies employed, the Matching-Adjusted Indirect Comparison (MAIC) is recognized for its robustness in comparative effectiveness research.

Matching Adjusted Indirect Comparison (MAIC) methodology leverages individual patient-level data (IPD) from one trial for adjustment and aggregate data (AD) from the comparator trial. Through weighting estimated via the method of moments, MAIC re-weights the IPD to align baseline characteristics with those reported in the comparator trial. This adjustment mitigates cross-trial differences, enabling robust indirect treatment comparisons, supporting outcome evaluation and regulatory decision-making in clinical research.

Ind_BIOS_PSTR_058: The Emerging of RWD and RWE - Is it really an alternative to RCT

Venkatesan Mudaliar
Navitas Life Sciences

Innovation and Technology have revolutionized the industry and led to the invention. When it comes to the pharmaceutical industry, Innovation and Inventions presents an ocean of untapped opportunities for business transformation and ease of conducting clinical trials. According to Tractica, the global artificial intelligence software market is forecast to grow from $10.1 billion in 2018 to $126 billion by 2025.

In the similar front, Real World Data studies are in the limelight and unquestionably the trials of the future which could emerge more hybrid study designs combining RCT and RWD. As this dynamic field evolves, innovative based approaches enable the efficiency, productivity, and sustainability. Lawmakers and regulatory agencies are actively engaging in discussion on the value of real-world data. In fact, the FDA has given framework for Real-World Evidence Program to accelerate medical product development and bring new innovations and advances in drug development which helps to reach faster and more efficiently to the patients. While RCTs remain the gold standard for evidence generation, Real World Evidence serves as a disrupter, bringing greater efficiency in improving regulatory decision by making ultimate means of maximizing revenue and minimizing the duration of drug to the market overall trial cost.

With an increasing technology in clinical trials usage of real time data capture and Electronic Health Records, biosensors, and wearables, etc., would help to record, process, review and statistically analyses them. However, the real challenge is to harmonize RWD and seamlessly allow them to use within planned hybrid RCT study design to produce clinically and statistically meaningful decision in drug development. This article emphasis

more on designing hybrid study designs i.e., bridging RWD and RCT to derive valuable evidence which is statistically valid for decision making.

IND_PSTR_SP_002: Optimizing ADaM Workflow with {AiR} Internal R Package – BDS

Dhivya Kanagaraj
Pfizer

Thsis package{AiR} can be used to create ADaM datasets(ADSL & BDS) targeting for FDA Submission, in which we will be focusing on BDS(Basic Data Structure) Datasets. Even though there are so many R packages for clinical programming in CRAN still there is need for such internal package's creation because of the gaps we see between packages in CRAN & company specific Standards. This package has functions/templates/metadata that covers generic BDS Standards with which we can create BDS specific ADaM datasets in R. Each template in the package consist of a main function which creates BDS specific variables along with domain specific functions used for derivations/imputations. These templates can be modified for study needs by the users.We will be talking about the workflow of the functions/templates, package structure, Validation and challenges while creating the package along with a demo of creating ADLB dataset.

IND_PSTR_SP_003: Smart Storage for Smarter Programming: Minimizing R memory with External Data Binding

Samyuktha Jagadeeswaran
ACL Digital

Handling large-scale clinical datasets in R often results in memory limitations and slow performance. This approach presents a lightweight framework that enables importing raw data from multiple formats (SAS, CSV, Excel, TXT) and storing them externally as Parquet files—keeping R's memory usage minimal. Only the R code runs in-memory, while both raw and processed datasets remain stored outside. Users can seamlessly access datasets via active bindings, apply transformations, and auto-save results externally. Additional utilities like variable search across all datasets and raw file cleanup improves usability. The system also displays real-time metadata, full dataset viewing is supported, surpassing RStudio's 1,000-row display limit. The current version is ready for internal use, supporting efficient programming without memory overload. Future plans include extending it into a comprehensive R package with GitHub-integrated validation pipelines, Shiny dashboards for interactive exploration, and RMarkdown for automated reporting.

IND_PSTR_SP_004: Optimizing Clinical Data Flow: A Standardized End-to-End Process from CRF Annotation to CRT Package Submission

Nagalakshmi Kudipudi, Ramya
ACL Digital

In the increasingly regulated and data-intensive world of clinical trials, standardization and traceability across data processing stages are critical. This abstract outlines a structured, end-to-end process encompassing CRF Annotation, SDTM, ADaM, TLF generation, and CRT Package preparation, ensuring regulatory compliance and data integrity. The process begins with **CRF Annotation**, where the case report form variables are meticulously mapped to SDTM domains, providing a foundation for accurate data transformation. This feeds into **SDTM development**, involving the creation, programming, and validation of Study Data Tabulation Model datasets, followed by rigorous quality checks and *P21 compliance*. The SDTM datasets serve as input to the **ADaM (Analysis Data Model)** stage, where analysis-ready datasets are developed, validated, and assessed for consistency and adherence to CDISC standards. Next, **TLF (Tables, Listings, and Figures)** generation is performed using validated ADaM datasets. These outputs undergo multiple layers of review to ensure accuracy and traceability to the statistical analysis plan. Finally, the **CRT (Case Report Tabulation) Package** is compiled, including define.xml, annotated CRFs, reviewer's guides, and validation results, ready for regulatory submission. The standardized flow improves cross-functional collaboration, traceability, and submission readiness while reducing redundancy and errors. This model provides a scalable framework adaptable to varying study complexities and is aligned with regulatory expectations, positioning it as a best practice in clinical data lifecycle management.

IND_PSTR_SP_007:Blueprint to Submission: Building SDTM with Python and R

Gayathiri Premkumar
Sensan Biosciences

R and **Python** are transforming clinical research through powerful, open-source tools for data analysis and regulatory reporting. **R** is widely recognized for its statistical modeling capabilities and rich visualization libraries like **ggplot2**, making it ideal for exploratory data analysis, adverse event trend visualization, and SDTM variable derivation. Packages such as **tidyverse**, **haven**, and **sdtm.oak** streamline data wrangling and compliance with CDISC standards, enabling efficient transformation of raw clinical data into submission-ready formats. **Python**, on the other hand, excels in automation, machine learning, and integration with modern data platforms. Libraries like **pandas**, **numpy**, and **scikit-learn** support data manipulation and predictive modelling, while tools like **pyreadstat** and **xptwriter** facilitate SDTM-compliant XPT exports. Together, they support reproducible research, accelerate regulatory submissions, and foster transparency in clinical development. This presentation outlines strategies for leveraging R & python in clinical standardization efforts.

IND_PSTR_SP_009: Submission Package Checklist: A Validator Perspective

Navesh Saravanam
Sensan Biosciences

Regulatory submissions play a crucial role in clinical trials. Regulatory Agencies requires all new drug applications (NDAs) to be submitted electronically. A clinical study is considered submission-ready when the SDTM, ADaM CRT packages meet high-quality standards. The Statistical Programming team is responsible for preparing the following components: SDTM datasets (XPTs), annotated CRF (acrf.pdf), define.xml, P21 report, and CSDRG (csdrg.pdf).ADaM datasets (XPTs), define.xml, P21 report and ADRG (adrg.pdf). Pharmaceutical companies conduct acceptance, and verification checks before submitting these packages to regulatory agencies. Based on our experience with submission activities, we have developed a checklist to ensure submission packages are of good quality. This presentation will mainly focus on the checklists with challenges faced with packages received from a CRO and provides solutions to get submission-ready CRT packages.

IND_PSTR_SP_013: Advancing SDTM Standards: Integrating NSVs and Phasing Out SUPPQUAL in IG v4.0

Abhinaya Rajachidambaram
Pfizer

The transition from Supplemental Qualifiers (SUPPQUAL) to Non-Standard Variables (NSVs) represents a major evolution in upcoming SDTM IG v4.0. SUPPQUAL datasets, used to capture sponsor-defined non-standard variables, follow a vertical data structure and reside in separate domains often resulting in fragmented data, complex linkage, and increased effort during data review and regulatory submission. NSV's transition integrates non-standard variables directly within the main SDTM domains using a horizontal format. Leveraging controlled terminology and the NSV Registry promotes consistent implementation of new or study-specific variables across datasets. This alignment supports better validation in Pinnacle 21 and enhances metadata transparency through Define-XML. This paper presents a practical approach for transitioning from SUPPQUAL to NSVs by identifying non-standard variables using NSV Registry fragments, mapping them to target domains, and generating metadata for NSVs. This shift improves data standardization, streamlines analysis, and enhances regulatory readiness through a future-oriented model for managing clinical trial data.

IND_PSTR_SP_019: My AI-Driven Voice-Assisted Coding Tool for R Programming

Mahendran Vibin
ACL Digital

This abstract presents an AI tool I'm currently developing for R programming which is a lightweight, offline-compatible R programming assistant that leverages local LLMs for automatic R code generation based on user commands — both voice and text. Designed for data scientists, statisticians, and life science professionals. The tool facilitates rapid code generation, visualization, and statistical analysis, eliminating the need for internet access or API keys. The system works by passing user commands to a locally hosted LLM via an HTTP interface, receiving functional R code, and executing it seamlessly in RStudio. This innovation can significantly reduce coding time and lower the entry barrier for non-programmers in pharma and research settings. Future versions will incorporate a Shiny GUI, real-time dataset awareness, and multi-turn conversations. The tool is poised to support adaptive analytics in clinical and biomedical data workflows.

IND_PSTR_SP_021: Customizing and Debugging SDTM Development in R with sdtm.oak

Indira Govindhan
ACL Digital

sdtm.oak is an open-source R package that supports structured and metadata-driven SDTM development. It simplifies the creation of SDTM domains from specifications and clinical trial data, ensuring consistency and traceability. This presentation outlines practical ways to extend and customize sdtm.oak for real-world study needs. It focuses on adapting internal logic, applying custom rules, and improving flexibility within the framework. These strategies help programmers align automated workflows with sponsor guidelines and study protocols, while maintaining transparency and regulatory compliance. By combining automation with customization, this approach enables efficient, audit-ready SDTM pipelines suited for dynamic clinical trials

IND_PSTR_SP_022: Traceability at Risk : The Hidden Pitfalls in Clinical Events Annotation

Naga Harshitha
ACL Digital

Clinical Event (CE) Case Report Form (CRF) annotation is critical in ensuring accurate data capture and regulatory compliance in clinical trials. However, the complexity of cardiovascular studies, multiplicity of data fields, and interdependencies between CRF modules make annotation a challenging task.This presentation explores the core challenges faced in CE - CRF annotation using annotated cardiovascular CRF samples and proposes solutions to streamline the process.Key issues include code list ambiguities, misclassification of findings, inconsistencies across domains and reliance on manual interpretation of source data.Recommendations include the adoption of AI-assisted annotation, standardization via CDASH/CDISC terminologies, and cross-functional training. Proper annotation significantly improves data quality, reduces regulatory queries, and supports robust endpoint adjudication.

IND_PSTR_SP_026: Effective Data Visualization for Regulatory Submissions in Psychotic Disorder Treatments

Mirnalini Amirthalingam, Indira
ACL Digital

A wide range of Schizoaffective disorder is included in psychotic illness. A non-invasive neuromodulation method that has gained popularity, Transcranial Magnetic Stimulation (TMS) is just one aspect of the rapidly changing field of neurotherapeutic interventions. By rigorously examining a variety of treatment modalities used by top medical device firms, such as Deep Brain Stimulation (DBS), Vagus Nerve Stimulation (VNS), Electro Convulsive Therapy (ECT), this work expands the scope beyond TMS. In order to depict treatment outcomes and patient trajectories across psychotic diseases, this study looks at a variety of graphical and data visualization techniques. In order to determine how well they can depict treatment pathways, clinical response patterns and neurobiological targets, techniques including Sankey diagrams, Kaplan-Meier survival plots, heatmaps and network graphs are investigated towards regulatory acceptance of data.

IND_PSTR_SP_024: Automating SAS, Python, R Code Using AI/ML & NLP Techniques

Prakash Palaniswamy
ACL Digital

As clinical trials evolve, the need for faster, more reliable and scalable programming solutions has become increasingly critical. Manual programming using specifications is time-consuming, error-prone and resource-heavy-especially when changes are frequent or cross-platform support is needed (e.g., SAS, Python and R). This abstract presents an AI-driven framework that automates the generation of statistical programs directly from structured specification files (e.g., SDTM, ADaM specs in Excel or Define-XML). Leveraging natural language processing(NLP) and machine learning(ML), the proposed system reads variable metadata and derivation rules, then intelligently produces validated base code in SAS, Python or R. When specifications are updated, the model identifies the changes and updates only the relevant portions of the existing code, reducing rework and ensuring traceability. This approach supports hybrid workflows where AI accelerates coding and human programmers focus on validation, review and domain-specific decision-making. The result is faster delivery timelines, reduced programming effort and a scalable system that empowers teams without replacing them. The talk will also discuss implementation strategies, lessons learned and future scope for integrating such AI workflows into regulatory-compliant pipelines.

IND_PSTR_SP_028: AI/ML Driven Code Migration from SAS to Open Source

Prakash Subramaniam, Naveen Raj
ACL Digital

Refactoring SAS code using AI and embracing open source tools enable organizations to modernize, standardize legacy workflows through intelligent automation and technologies. The Translation or Conversion of Statistical Analysis System (SAS) code to open-source languages like Python or R have become increasingly relevant in modern data analytics and research. This complex process requires meticulous planning and execution to ensure minimal disruption and data integrity. Key steps include assessing the current environment, defining the target architecture, performing a gap analysis, and developing a comprehensive translation strategy. Effective conversion also involves thorough testing, validation, and user training to ensure continuity and performance optimization. Leveraging expertise in SAS programming, data management, and system architecture is crucial for a successful conversion. As an expert in code translation/conversion, we ensure streamlined processes, cost efficiency, and enhanced system capabilities to support organizational goals and data-driven decision-making. This abstract outlines the scope of the topic, highlighting the importance and approach to translating/converting SAS code to Python or R for enhanced data analysis capabilities.

IND_PSTR_SP_029: Revolutionizing medicine: recent developments and prospects in stem-cell therapy

Ram kishan Krishnan Reengaswamy
Syenos Health

Novel therapeutic techniques in clinical trials offer a paradigm shift in the approach to battling prevalent and destructive diseases. Stem-cell therapy is witnessing a surge in clinical trials, reflecting a growing interest in translating laboratory findings into viable treatments. Clinical trials involving various stem-cell types are currently underway and include a wide range of health issues. The goal of ongoing trials is to determine whether stem-cell therapies are effective in alleviating symptoms of neurological diseases such as Alzheimer's, Parkinson's, and spinal cord injuries. Researchers are investigating how stem cells might be able to repair damaged neurons, encourage brain regeneration, and lessen the symptoms of these crippling conditions. Here we discuss recent biotechnological advancements, critical trial evaluations, and emerging technologies, providing a nuanced understanding of the triumphs, difficulties, and future trajectories in stem cell-based regenerative medicine.

IND_PSTR_SP_071: Gen AI, CDISC Compliance, Trustworthy Analytics

Ashwini Reddy
SAS

Many CROs and Pharmaceutical Organizations have established technical processes to be followed for the submission of documents in the eCTD format. The reporting is complex and requires adherence to protocols and compliance with Good Clinical Practice, and precise documentation formatted according to guidelines set by regulatory authorities.

This paper discusses the potential for using LLMs and other Natural Language Processing (NLP) techniques to aid data interpretation, clinical documentation, regulatory submission, and post-trial analysis, while being watchful of data privacy, model bias, and responsible adoption.

The following steps are discussed:

Use SAS to extract contextual information from clinical trial protocol and SAP

1. Employ SAS Natural Language Processing (NLP) techniques and Large Language Models (LLM's)

2. Confidence Scoring in SAS Viya Visual Text Analytics – LITI Assessment of cost, security, and privacy.

The use of LLMs like LLAMA-V2 and RAG-for-LLMs proved beneficial, though resource-intensive. Leveraging SAS Visual Text Analytics LITI rules as pre-filters for confidence scoring reduced costs and gave quicker and more accurate results.

Ind_DS_PSTR_001: Comparative Analysis of Master Protocol Designs

Rashmi Thakur, Soumya N
Eli Lilly

Introduction - Traditional clinical trials often assess one treatment in one disease population, limiting efficiency. Master protocols address this by allowing evaluation of multiple interventions, biomarkers, or populations under a unified structure.

Types of Master Protocols

- **Umbrella Trials**
    - **Definition**: Test multiple therapies in a single disease with different genetic or molecular subtypes.
    - **Example**: *Lung-MAP* (Lung Cancer Master Protocol) – evaluates targeted therapies for non-small cell lung cancer (NSCLC) based on genomic profiles.
    - **Best For**: Heterogeneous diseases with well-characterized subtypes.
- **Basket Trials**
    - **Definition**: Assess one therapy across multiple diseases sharing a common biomarker or mutation.
    - **Example**: *NCI-MATCH* – assigns patients with different cancers to targeted treatments based on genetic alterations.
    - **Best For**: Rare mutations or therapies targeting molecular mechanisms across indications.
- **Platform Trials**
    - **Definition**: Evaluate multiple therapies in a perpetual, adaptive design, allowing arms to be added or dropped based on interim analysis.
    - **Example**: *REMAP-CAP* – tests treatments for community-acquired pneumonia (including COVID-19) in an ongoing, global platform.
    - **Best For**: Dynamic diseases, pandemics, or areas requiring continuous evaluation.

Comparison and Conclusion

- **Platform trials** offer maximum flexibility and efficiency but demand advanced infrastructure.

- **Umbrella and basket trials** are simpler and effective when targeted to specific research goals.

Conclusion: No single design is universally best; the optimal choice depends on disease biology, trial objectives, and resource availability.

Ind_DS_PSTR_002: AI/ML in Clinical Trials: Big Potential, Growing Regulatory Clarity

John Shalen
ACL Digital

Artificial Intelligence (AI) and Machine Learning (ML) are reshaping clinical trials by enabling smarter patient selection, early dropout prediction, and real-time anomaly detection. These tools promise faster, more efficient studies—but their use in regulatory submissions remains limited due to concerns around explainability, traceability, and validation.

In January 2025, the FDA released draft guidance introducing a risk-based framework to assess AI model credibility based on context of use and model risk. It emphasizes transparency, life cycle monitoring, and early engagement with regulators—marking a significant step toward regulatory acceptance.

Despite this progress, most AI/ML models are still used internally and excluded from submission datasets like ADaMs. Broader adoption will require clear validation standards, audit-ready documentation, and training for statisticians and programmers.

Looking ahead, the industry must collaborate to develop standard operating procedures, invest in explainability tools, and align AI development with Good Clinical Practice. With the right frameworks in place, AI/ML can evolve from exploratory tools into trusted, submission-ready assets that improve trial quality and accelerate decision-making.

Ind_DS_PSTR_003: AI-Driven Real-Time Anomaly Detection in Remote Patient Monitoring

Dharani Eswaramoorthy
ACL Digital

Remote Patient Monitoring (RPM) using Continuous Glucose Monitoring (CGM) systems enables real-time tracking of glucose levels, supporting early detection of complications in diabetes. These systems facilitate continuous physiological data collection, aiding clinical decision-making, delivering timely alerts, and empowering patient self-management. However, despite their potential, current CGM alarm systems are limited by high false positive rates and lack the adaptability needed to account for individual variability in glycemic patterns. To address these limitations, this study proposes the development of a machine learning–based anomaly detection framework in R, specifically tailored to CGM data in RPM settings. The model is designed to reduce false alarms, personalize detection thresholds, and enable real-

time identification of clinically significant glycemic events. This approach aims to improve patient outcomes and enhance the reliability and precision of AI-driven alerts in diabetes management.

Ind_DS_PSTR_006: Operationalizing AI/ML for Real Time Data Integration and Interoperability in Life Sciences

Renisha Robinson
Sensan Biosciences

The explosion of diverse healthcare data—from EHRs and lab systems to wearables and real-world evidence—has intensified the need for intelligent, scalable, and regulatory-compliant data integration solutions. Artificial Intelligence (AI) and Machine Learning (ML) are transforming Clinical Data Management by automating processes such as data mapping, normalization, and semantic alignment across heterogeneous sources. Leveraging standards like HL7 FHIR, OMOP CDM, and CDISC SDTM, AI/ML enables real-time data ingestion, natural language processing of unstructured clinical text, and federated learning frameworks that support privacy-preserving collaboration. Advanced technologies including AI-driven APIs, smart ETL tools, and biomedical knowledge graphs are essential to achieving semantic interoperability and actionable data flow. Combined with MLOps for continuous model monitoring and deployment, these innovations address challenges related to data quality, bias, and legacy systems. AI/ML-powered interoperability is emerging as a critical driver in accelerating clinical research, enabling personalized medicine, improving healthcare analytics, and advancing population health management.

Ind_DS_PSTR_007: DuckDB: A solution to big data woes in R

Saheli Das
Merck

R users often struggle with large datasets, such as healthcare claims data, due to memory constraints and slow operations like joins or filtering.

DuckDB, an open-source, in-process SQL OLAP database, transforms R into a powerful big data analytics tool. Seamlessly integrated with R, DuckDB supports complex SQL queries on large datasets without external database servers, handling diverse formats like CSV and Parquet efficiently. Its streaming execution and memory management enable scalable, reproducible analyses within RStudio.

This presentation will provide a brief introduction to this system along with a demonstration to help the audience get a flavor of this powerful tool. Attendees will learn how DuckDB bridges R and SQL, eliminates data import/export bottlenecks, and enhances workflows with packages like ggplot2 and tidymodels. Ideal for data scientists and analysts, DuckDB unlocks big data potential without compromising R's simplicity.

Ind_DS_PSTR_008: AI-Driven Protocol Review: Accelerating and Simplifying Review

Aditya Moogi
Novo Nordisk Service Centre India Private Ltd.

Clinical trials are guided by documents - protocols, plans, manuals - that are rich in knowledge but locked in static formats. As the volume and complexity of these documents grow, traditional ways of accessing and interpreting them no longer scale. This presentation explores how the clinical trial industry must shift from systems that store information to systems that understand it. By adopting a mindset of designing intelligence, we introduce a framework that leverages large language models (LLMs) and retrieval-augmented generation (RAG) to transform passive documents into interactive knowledge interfaces. Rather than focusing on building tools, this session emphasizes rethinking workflows, capabilities, and cognitive bottlenecks in clinical trials. We present an example of this thinking, illustrating how intelligent document interaction can reduce protocol misinterpretation, improve efficiency, and empower trial teams at every level. This is not just about AI adoption – it is about redesigning how we engage with clinical knowledge itself.

Ind_DS_PSTR_009: Fixing the Data Puzzle - Data Repair with SAS|R|AI

Sanjay Srinivasan, Govindhan
ACL Digital

Handling missing data is a critical step in clinical and real-world data analysis. Multiple imputation (MI) has emerged as a robust method to address this issue by generating several plausible datasets and combining the results for unbiased analysis. This study presents a comparative evaluation of multiple imputation techniques using SAS (PROC MI), R (mice package), and AI-based methods (KNN, MICE, and Deep Learning) on a real-world clinical dataset, where missing values were intentionally introduced under MCAR, MAR, and MNAR missingness patterns. We highlight differences in usability, diagnostics, and output interpretation across platforms based on four evaluation criteria: Root Mean Squared Error (RMSE), Unsupervised Classification Error (UCE), Supervised Classification Error (SCE) and execution time.

In healthcare, imputation enables analysis of incomplete clinical datasets, facilitating informed decision-making and personalized treatment strategies.

Ind_DS_PSTR_010: Empowering Clinical Programming through Interactive Data Visualization

Sugunesh Sivalingam, Vijay, Harika
ACL Digital

In the era of data-driven decision making, Clinical Programming teams require tools that enable swift and meaningful interpretation of complex datasets. The Data Visualization Tool is an in-house platform developed to address this need, designed specifically for use in clinical research environments. This tool allows, programmers, statisticians and clinical teams to explore, analyze and present clinical trial data interactively and intuitively. With the input of standard clinical trial datasets such as SDTM and ADaM, this tool enables real-time visualization of patient profiles, adverse event trends, efficacy endpoints, and data quality metrics. Its customizable dashboards support rapid detection of anomalies, patient-level tracking and trend analysis- facilitating earlier insights and more decision-making across functional teams. Data Visualization Tool implementation has improved the efficiency of data review cycles, reduced manual reporting burdens, and enhanced cross-functional communication by transforming static outputs into dynamic visuals. This abstract highlights the role of Data Visualization Tool in bridging the gap between raw clinical data and actionable intelligence, supporting the broader goals of data transparency, regulatory compliance and patient safety.

Ind_DS_PSTR_027: Uncovering Propensity Score Matching with Advanced Data Visualization in Real-World studies

KiranKumar Vyas, Umayal
Veramed

Propensity Score Matching (PSM) is a crucial statistical method used in clinical trials to estimate causal effects by balancing covariates between treated and control groups. While numerical results provide precise assessments, they can be difficult to interpret, especially when dealing with complex propensity score distributions and covariate balance metrics. Relying solely on numbers may obscure important patterns and hinder meaningful conclusions.

Advanced data visualization techniques help translate complex statistical outputs into intuitive formats, making it easier to assess matching effectiveness, detect biases, and understand analytical implications. This proposal highlights the need to integrate qualitative insights through visualization to enhance interpretation and decision-making.

To demonstrate this approach, an interactive dashboard has been developed as a case study, showcasing how qualitative insights complement quantitative results. This methodology improves communication of key findings, making statistical analyses more accessible and supporting data-driven decisions in clinical research.

Ind_MGT_PSTR_001: Commanding Calm: Emotional Intelligence in High-Stakes Conflict and Crisis Leadership: Mastering Influence, Empathy, and Clarity Under Pressure

Jagruti Desai
PAREXEL Pvt. Ltd.

In the high-pressure world of pharma, healthcare, and clinical research organizations (CROs), effective leadership during critical moments is a clinical, commercial, and reputational imperative.

This presentation explores how emotional intelligence (EI) enables leaders to respond to conflict and crisis with clarity, empathy, and ethical influence.

Drawing on industry-specific case studies we introduce a tactical EI framework—Read the Room, Anticipate Resistance, Shift Momentum (RARS) designed for high-stakes moments in science-led environments.

This presentation showcase techniques to de-escalate functional breakdowns, align stakeholders under scrutiny, and maintain composure in project teams, with auditors, regulators, and commercial teams.

This presentation equips leaders with practical tools to bridge communication gaps, bridging Medical–Commercial divides, navigate complex sponsor relationships, and addressing reputational risk, which leads leaders to lead with presence rather than pressure.

Mastering EI becomes a leadership differentiator that drives trust, resilience, and performance—when it matters most.

Ind_MGT_PSTR_002: A Step Towards Leadership from Programming

Nishanth Sungaram
Fortrea

Transitioning from a programming role to leadership is a transformative journey that requires a shift in mindset, skillset, and approach. While technical expertise remains valuable, leadership demands strong communication, delegation, and team-building abilities. This presentation explores the key steps to successfully navigate this transition, including assessing personal strengths, developing essential soft skills, and seeking leadership opportunities.

Attendees will gain insights into balancing technical and managerial responsibilities, staying connected to technology while leading teams, and leveraging mentorship and networking for career growth. By focusing on continuous improvement and embracing leadership challenges, programmers can evolve into effective leaders who drive innovation and inspire teams.

This session provides actionable strategies to help professionals take their first step toward leadership, ensuring a smooth and impactful transition from coding to guiding teams toward success.

Ind_MGT_PSTR_003: Bridging the Generational Leadership Crisis in the era of Data and AI

Damini Sharma

ACL Digital

Contemporary society faces an unprecedented leadership crisis that permeates from foundational family structures to specialized professional domains. This crisis is particularly acute in Organizations, where the intersection of generational disparities (Gen X to Generation Alpha), technological evolution, and regulatory complexity creates a perfect storm of leadership challenges. Generational complexity is compounded by the fact that clinical research often involves multi-decade timelines, requiring leadership strategies that can span multiple generational transitions. The crisis manifests across multiple societal levels: families failing to cultivate character and resilience in children, educational systems prioritizing performance over leadership development, cultural shifts toward superficial influence over authentic leadership, and workplaces confusing management with true leadership. As industries evolve technologically, the human and behavioral side of leadership— adaptability, accountability, resilience—is often underdeveloped in younger cohorts. Within this broader context, statistical programming represent critical domains where leadership failures have immediate implications for public health and therapeutic advancement.

STD_PSTR_001: Comparing Statistical Models for Clustered Survival Data: Insights from Real-World Data

Pravin Sahadevan

National Institute of Mental Health & Neuroscience (NIMHANS)

Clustered data often arises in time-to-event analysis.

Often, these data are analysed using the Cox proportional hazards model, ignoring the clustering effect, which leads to misleading findings.

In this study, we tried to compare alternative survival models like the shared frailty model and the marginal Cox model that can handle the clustering effect using the secondary data from the Center for International Blood and Marrow Transplant registry.

To study the association between neighborhood-level poverty and survival outcome among the pediatric allogeneic hematopoietic cell transplantation (HCT).

We found the results were similar between the models, and the shared frailty model provided a better model fit than the marginal Cox model.

STD_PSTR_002: Application of Censored Quantile Regression to Explore Prognostic Determinants in Major CVD Event

Guru Prasath

National Institute of Mental Health & Neuroscience (NIMHANS)

The Cox proportional hazards model is a widely used model for survival data.

In this study, secondary data from a retrospective 9-year cohort of 977 UAE citizens was used to explore three survival models: the Cox proportional hazards model, the Accelerated Failure Time (AFT) model, and Bottai's censored quantile regression model for identifying the prognostic determinants of major cardiovascular disease (CVD) incidence.

Cox & AFT Models failed to capture the heterogeneity due to the covariates on the major CVD incidence.

Quantile Regression flexibly models the covariate effect on different quantiles of the time to major CVD incidence.

STD_PSTR_004: AI-Guided Statistical Optimization of CRISPR-Loaded Nanobots for Triple-Negative Breast Cancer Therapy

Prabhuraj K

Bharathiar University

Triple-Negative Breast Cancer (TNBC) is an aggressive subtype of breast cancer that lacks targeted treatment options. In this study, we designed CRISPR-Cas12a-loaded lipid nanobots to edit the **BRCA1 gene pathway**, aiming to restore tumor suppressor activity. Using

artificial intelligence (AI)-driven Design of Experiments **(DoE)** and multivariate statistical models, we optimized nanobot formulation and delivery. Preclinical assays in TNBC cell lines and murine models demonstrated high editing efficiency (>80%), significant apoptosis induction, and minimal off-target effects. Statistical analyses using **SAS and R** confirmed robust treatment-response relationships, indicating a strong therapeutic potential. This integrative approach highlights the synergy between nanomedicine, gene editing, and statistical modeling in oncology.
These experimental designs were further enhanced using AI-driven models such as Bayesian optimization and random forest regression, enabling the prediction of optimal nanobot configurations with minimal experimental runs. During the in vitro and in vivo testing phases, classical statistical methods including **t-tests, ANOVA, and post-hoc** comparisons (e.g., Tukey's HSD) were used to assess significant differences in gene-editing efficiency, tumor suppression, and apoptosis between control and treatment groups.

STD_PSTR_011: R_Compare: An R-Based Alternative to SAS PROC COMPARE for Dataset Integrity Checks

Monisha P
Bharathiar University

R_compare is a **programmatic R function** developed to **compare datasets** in-depth, emulating up to 90% of the **fundamental features of SAS PROC COMPARE**. It compares structural similarity by **merging variables** between datasets, *examining differences in* ***variable names, data types, lengths****, and* ***missing patterns***—essential features replicating SAS's structural diagnostics.
The package produces an extensive **HTML report** consolidating **dataset metadata, observation-level discrepancies**, and **cell-level value differences**. With row tracking, styled output, and modular summary tables, R_compare facilitates reproducible, audit-ready clinical, omics, or survey data comparisons. With the use of R packages like **dplyr, kableExtra**, **rlang** and **knitr,** R_compare focuses on **readability and flexibility** while shunning the **licensing issues of SAS**. It is a light, transparent solution, particularly for academic or open-source setups.
R_compare facilitates **reproducible and transparent** workflows.
This makes it a significant tool in the domain of **Data Science in Biology** as it allows **scalable validation** of data for **research and clinical datasets** without the need for licensed statistical software.

STD_PSTR_012: Statistical Modeling and Optimization of Bioplastic from ***Eichhornia crassipes*** (water hyacinth) using **Response Surface Methodology**

Nivedhitha V
Bharathiar University

Plastic is made from non-renewable resources, which poses a significant hazard to both human health and the environment. Bioplastics can minimize harmful chemical release and reduce the risk of microplastic contamination, especially in food packaging applications. The present research work focuses on the preparation of bioplastic from ***Eichhornia crassipes*** **(water hyacinth)**, an invasive aquatic plant species but it is a cellulose-rich plant. It is

biodegradable and acts as a reinforcing agent to enhance the mechanical properties of bioplastics. In this study, cellulose was separated using an alkali treatment method. To optimize the formulation parameters, **Response Surface Methodology (RSM)** was employed using a **Box-Behnken Design (BBD)**. This statistical approach enabled the systematic investigation of the effects and interactions of key factors (such as cellulose concentration, plasticizer content, and drying temperature) on the bioplastic properties. The BBD facilitated the development of a predictive model for identifying the optimal conditions required to achieve desirable strength and biodegradability while minimizing the number of experimental trials.

STD_PSTR_013: Biological control of Allelopathic effects of *Parthenium hysterophorus*

Pavalalosini A
Bharathiar University

Parthenium hysterophorus L., has caused a drastic decline in the yield of crops during the years. It is found to be a congress weed that has spread throughout the globe. The eradication of this weed species has been challenging throughout the years. To control its spread various management strategies like physical, chemical, and biological control have been considered but no single method has proved satisfactory. Integrative approaches have been applied for effective eradication of these weed species. These species tend to affect the crops by releasing an allelopathic chemical known as parthenin that inhibit the germination of surrounding crops, cause toxic effects to environment. Certain microorganisms can be applied to protect the allelopathic effect of these weeds on the surrounding plants. This biological control can be combined with other weed eradication methods to work effectively. These microorganisms capable of degrading or capturing parthenin can be isolated from soil environment, and the optimization of its growth conditions can increase its effectiveness.

STD_PSTR_014: Pan HLA Binding Affinity Prediction Using Transfer Learning Models With Integrated Statistical Analysis

Sathya Priya S
Bharathiar University

Accurate prediction of **peptide–MHC (Major Histocompatibility Complex) binding** is critical for **vaccine design, cancer immunotherapy, and autoimmune research**. HLA **(Human Leukocyte Antigen)** a MHC molecule, is **polymorphic** in nature. The existing tools are limited to well-studied HLA alleles and perform poorly on rare or unseen variants. This project presents a pan-HLA binding prediction model using **transfer learning** with protein language models (e.g., ProtBERT, ESM). Peptide-HLA pairs were retrieved from IEDB and encoded using concatenated sequence embeddings of 9–11mer peptides and HLA pseudo-sequences. A transformer-based regression architecture was fine-tuned to predict log binding affinities. Pretrained transformer models were fine-tuned to predict binding affinities across diverse HLA types. The model was benchmarked to improve generalization to underrepresented alleles on zero-shot allele splits. Feature interpretation using SHAP values can be used to highlight the key amino acid positions in both peptide and MHC sequences.

This approach enables **scalable**, **allele-agnostic immunogenicity prediction and supports equitable vaccine development** across global HLA diversity and **enhances epitope prediction** for genetically diverse populations and supports future integration with TCR-binding models for personalized immuno-oncology.

STD_PSTR_015: Automated and Reproducible Framework for the Comparative Analysis of Clinical Conditions Using R and R Markdown

Velentina K F
Bharathiar University

In modern clinical research, Reproducibility and automation are necessary to ensure consistency, transparency and scalability of analysis. This abstract focuses on creating a comprehensive, automated and reproducible framework for the comparative analysis of clinical conditions using R and Rmarkdown. This comprehensive framework integrates statistical testing, data visualization and most importantly report generation into a cohesive, script-driven pipeline that minimizes manual intervention and promotes reproducibility across diverse datasets and research scenarios.
By utilizing R's robust statistical and reporting capabilities, the framework enables standardized comparison of clinical outcomes and employing Rmarkdown helps to generate dynamical documentation to ensure traceability of analysis results and reproducibility with minimal effort. This approach contributes to higher standards in clinical data analysis and ease of deployment.

STD_PSTR_017: Predictive Modeling In Antiphospholipid Syndrome: A Logistic Regression-based Approach To Pregnancy Outcomes

Hasika SS
Bharathiar University

Antiphospholipid syndrome (APS) in pregnancy is a significant autoimmune condition associated with increased risks of miscarriage, stillbirth, preeclampsia, and intrauterine growth restriction. Current therapeutic strategies primarily involve low-dose aspirin and low molecular weight heparin to improve pregnancy outcomes and reduce thrombotic risks. Despite advances, challenges remain in managing refractory cases and in establishing standardized treatment protocols for women with non-criteria obstetric APS.
Future research focuses on personalized medicine approaches, immunomodulatory therapies, and refining diagnostic criteria to optimize maternal and fetal outcomes. Emerging biomarkers and advancements in laboratory diagnostics are enhancing early detection and risk stratification of APS in pregnancy, enabling timely intervention and improved monitoring throughout gestation.

STD_PSTR_018: Evaluation of drug toxicity on MMP using one-way ANOVA

Jeya Vignesh
Bharathiar University

Evaluating drug toxicity is a crucial aspect of both pre-clinical and clinical drug development particularly within pharmacokinetic investigations. Platinum-based anticancer drugs is commonly administered chemotherapeutic agent where it exerts its cell-killing effects primarily by inducing mitochondrial dysfunction and triggering apoptosis. The mitochondrial membrane potential (MMP) serves as a key marker of mitochondrial health and its loss being among the first signs of apoptotic processes.

In the present study, the alterations in MMP were assessed in HeLa cells exposed to various concentrations of platinum-based anticancer drug, utilizing TMRM staining. The fluorescence data obtained were statistically evaluated through one-way ANOVA. Based on the significant difference among the groups indicate at least one treatment group differed from the others in terms of mitochondrial membrane potential. Platinum-based anticancer drugs treatment at medium and high doses significantly reduced mitochondrial membrane potential in HeLa cells compared to the control suggesting dose-dependent mitochondrial toxicity.

STD_PSTR_019: Multi-Omics Integration: Bridging Genomics, Transcriptomics, Proteomics, and Metabolomics

Nirmal Velu
Bharathiar University

The integration of multi-omics data—comprising genomics, transcriptomics, proteomics, metabolomics, and epigenomics—offers a holistic view of complex biological systems. This approach provides hope for unravelling the intricate details in various aspects of biology and accelerates innovation in healthcare. Effective multi-omics integration relies on advanced biostatistical frameworks and computational tools, including dimension reduction techniques (PCA, t-SNE), clustering algorithms, and machine learning models. Key databases such as GEO,PRIDE database, TCGA, and METLIN provide rich repositories for omics data. Integration tools like iCluster, MOFA, Cytoscape, and mixOmics facilitate combined analysis, while statistical methods such as ANOVA, correlation analysis, and Bayesian networks reveal inter-omics associations and functional patterns. These strategies have enabled breakthroughs in biomarker discovery, personalized medicine, and systems biology. Applications span cancer classification, drug response prediction, and metabolic pathway reconstruction. This highlights the role of biostatistics in transforming high-dimensional omics data into actionable biological insights through integrative, data-driven approaches.

STD_PSTR_021: Statistical Assessment of Molecular Docking Outcomes for Anti-Aging Compounds from *Zingiber officinale* Targeting SIRT1 and Keap1 Proteins

Sakthishree N
Bharathiar University

Growing interest in natural therapies for aging has spurred investigations into plant-based bioactive compounds with therapeutic promise. *Zingiber officinale* (ginger), recognized for its strong antioxidant and anti-inflammatory effects, has been explored for its potential role in promoting longevity. This study investigates the anti-aging potential of selected bioactives from *Zingiber officinale* by targeting two key aging-related proteins, SIRT1 and Keap1, through molecular docking. AutoDock Vina was used to simulate docking and evaluate binding affinities and molecular interactions. Statistical analysis was carried out using descriptive statistics mean, standard deviation), comparative methods (t-tests and ANOVA), and interaction profiling (hydrogen bonding and hydrophobic interaction counts). Notably, 6-gingerol, zingerone, and 6-shogaol exhibited strong binding affinities. These findings, supported by statistical assessments such as mean binding energy comparisons, ligand efficiency calculations, and interaction frequency analysis, suggest these compounds hold promise as lead candidates for further pre-clinical investigation.

STD_PSTR_023: Accelerating Clinical Trials: AI-Powered Real-Time Data Capture and Validation, From Conversation to Database

Berwin N
St.Joseph's Institute Of Technology

Collecting data manually in clinical trials often leads to delays, errors, and high costs. This presentation introduces an innovative system that fully automates this process. Our solution uses AI to convert doctor–patient conversations into text while simultaneously capturing real-time data from connected medical devices. This combined data automatically populates a preliminary eCRF, which is immediately shown in a clear, standardized format on the investigator's tablet or mobile device. Investigators can then review and verify patient information within seconds or minutes. Once approved, the verified data is securely transferred to the sponsor's database. By eliminating the need for manual data entry, this direct "source-to-system" method enhances data accuracy, accelerates trial timelines, and reduces the workload of clinical teams.

STD_PSTR_026: Automated AE & SAE Classification Using NLP and Medical Dictionaries

Paul Ibzhan JB
St.Joseph's Institute Of Technology

AE (Adverse Event) classification is traditionally manual and subjective.

Manual review is time-consuming and inconsistent across reviewers.

Misclassification can impact safety signals, submissions, and compliance.

NLP can standardize this process using MedDRA and clinical context.

Objective: Build a lightweight, explainable system to classify AE vs. SAE.

STD_PSTR_032: Designing Graph Specifications for Clinical Data Visualization Solution

Kiruthikaa Kannan
Madras Christian College

Effective communication of clinical data relies heavily on standardized and statistically sound visualizations. This work focuses on the development of detailed graph specifications for over 60 plot types commonly used in clinical trials, including forest plots, Kaplan-Meier curves, swimmer plots, shift plots, and exposure-adjusted incidence rate plots. Each specification was designed as a blueprint, outlining the purpose of the graph, clinical context, statistical methodology, expected input structure (e.g., ADaM datasets), visual layout, customization parameters, and regulatory relevance.

Key considerations included alignment with CDISC standards, clarity for clinical reviewers, reproducibility, annotation conventions, and support for automation. By decoupling specification from implementation, the framework enables consistent graph generation across studies, reduces validation overhead, and enhances interpretability. These specifications serve as foundational assets for programming teams and ensure that visualizations are both scientifically robust and submission-ready, regardless of the tools or environments used.

STD_PSTR_034: Integrating Clinical, Demographic, and Genomic Data for Robust, Explainable Heart Disease Risk Prediction Using Machine Learning

Harsh Kumar
IMS, BHU

Heart disease remains the leading cause of death globally, with over 17 million fatalities annually. Despite the urgency, early detection remains elusive due to the limitations of traditional diagnostics, which are costly, invasive, and often inaccessible in low-resource settings. While machine learning (ML) has shown promise in predicting cardiovascular risk, most existing studies rely on single, homogenous datasets with limited features, compromising model generalizability and clinical utility.

In this project, we address these critical gaps by integrating **three diverse datasets**: the Kaggle cardiovascular dataset (>70,000 patients), UCI heart disease datasets (e.g., Cleveland), and gene expression profiles. This multi-dataset approach combines demographic, clinical, and molecular features, capturing both traditional risk factors (e.g., age, blood pressure, cholesterol) and genetic predispositions to heart disease. Data preprocessing steps include rigorous outlier removal, feature engineering (e.g., BMI, MAP), and clustering to uncover latent patient subgroups.

We train and compare multiple ML algorithms—Random Forest, XGBoost, Decision Tree, Logistic Regression, and Multilayer Perceptron—using extensive hyperparameter tuning with k-fold cross-validation. Model performance is evaluated using accuracy, AUC, precision, recall, and F1-score, ensuring robust assessment across heterogeneous populations. To enhance clinical interpretability, we apply SHAP and LIME to visualize and explain the contribution of individual features to predictions.

Our findings demonstrate that integrating diverse datasets significantly improves predictive accuracy and generalizability, while interpretability frameworks bridge the gap between high-performance ML models and actionable clinical insights. This work highlights the potential of combining clinical, demographic, and genomic data to enable **scalable, affordable, and explainable early heart disease risk prediction**, supporting proactive healthcare interventions worldwide.

Make IDAR at J&J your top choice

The Integrated Data Analytics & Reporting (IDAR) organization at Johnson & Johnson is looking for fresh perspectives from college graduates with diverse backgrounds and experiences to drive the future growth in IDAR.

What is IDAR?

IDAR is a global organization comprised of several disciplines that **translates data into information to bring medicines to patients in need.** We're a critical part of Johnson & Johnson.

Why Johnson & Johnson? Be a part of something big

Johnson & Johnson is one of the largest and fastest-growing global pharmaceutical companies in the world, with an industry-leading pipeline of new innovations.

Driven by Our Credo, our core focus is improving health at every stage of life. From medicines to cancer treatments, the work we do has the power to impact the lives of billions of people around the world.

People are our passion—from patients, doctors and caregivers to our organization members—we strive to make a difference in the lives of others. Johnson & Johnson is not just a place to work, but a place to belong and be a part of something big.

We're a global company spanning 60 countries. The IDAR organization operates across 5 regions and 30 countries around the world.

Creating a healthier world

As part of the world's largest healthcare company, we recognize our responsibility to our communities, the environment and our planet. We continuously push to think and act sustainably, reducing our environmental footprint whenever possible.

Find your fit in IDAR

No matter what you studied, you can find your fit in IDAR.

Ayesha brings her computer science skills to Clinical & Statistical Programming

Ayesha, a computer science major and math whiz by nature, found her fit in IDAR by joining the Clinical & Statistical Programming organization. Her days consist of planning, designing and developing codes to support clinical data, visualizations, analysis and reporting activities.

Have a love for coding or analysis? Apply your knowledge in IDAR.

Nico's passion for problem solving is a great fit for Data Management

Nico loves a good challenge and has impeccable attention to detail. His background in science and engineering attracted him to IDAR and the Data Management organization where he's an expert in collecting and cleaning clinical trial data.

Are you the most organized person you know? Flex your skills in IDAR.

How we work together

IDAR collaborates across functions and externally. Here's a snapshot of what we do:

IDAR Business Operations (IBO)	Manages strategic and operational business management activities across the IDAR functions.
Clinical and Statistical Programming (C&SP)	Delivers analytical programming solutions that enable early and effective decisions. The role within C&SP plans, designs, collaborates with other functions and develops codes to support clinical data review, visualizations, analysis and reporting activities.
Clinical Data Standards & Transparency (CDST)	Maintains standards of deliverables for all aspects of clinical study from design through reporting. CDST supports clinical data transparency that enhances the patient and investigator experience and advances scientific research.
Data Management (DM)	Provides data delivery expertise in the conduct of our clinical trials. The organization develops a database to capture data, oversees data cleaning activities and delivers data that meets the standards for potential submission to regulatory authorities.
Regulatory Medical Writing (RMW)	Writes high-quality clinical documents to support the development, regulatory approval and maintenance of our products. These documents include investigator's brochures, clinical protocols, clinical study reports, summaries of efficacy and safety from clinical trials and responses to questions from health authorities.
Risk Management Central Monitoring (RM-CM)	Leads the risk-based approach to monitoring clinical trials to help ensure protection of human subjects and the quality of clinical trial data. The RM-CM group is responsible for each detection, analysis and mitigation of potential risks in clinical trials.

Play J&J trivia!
Myth or Fact

1. J&J has been on the Working Mother's list of the 100 Best Companies for over 30 consecutive years.

2. Our Credo stems from a belief that consumers, employees and the community are all equally important.

(Answers on back)

Be you!
We celebrate each other's differences

Johnson & Johnson has a strong commitment to Diversity, Equity and Inclusion (DEI). We believe the best ideas and solutions unfold when employees bring their uniqueness to work with them.

> "On my first day, my organization leader inspired me by saying, Johnson & Johnson will let you shine in your own way."
>
> –Clinical Data Manager Intern

Thank you So much for sponsoring

IASCT ConSPIC 2025

We are grateful for your support