MBSW 2024
BOOK OF ABSTRACTS

RENAISSANCE INDIANAPOLIS NORTH HOTEL, CARMEL, IN

MONDAY MORNING 8:30AM- 12:00PM JUNE 10 – SHORT COURSES

A Statistician's Guide to DMC –Overview, Standards, and Best Practice
David Kerr, Cytel Inc.

PDMCs (Data Monitoring Committees) are an important component of the clinical trial process. They are charged with reviewing interim data and making recommendations to protect the safety of trial participants and ensure the scientific integrity of the study. We will cover beginning, intermediate, and advanced topics regarding the DMC process and how all statisticians – the DMC member, the sponsor, and the SDAC (Statistical Data Analysis Center) supporting the DMC – can better work together.

Specific topics to be discussed will include the history of DMCs and current guidance documents; the organizational flow of the DMC process and the responsibilities of those involved; DMC meeting structure, timing, and purpose; logistics of DMC membership including assessment of conflict of interest; DMC review of study conduct, safety, efficacy (possibly with formal boundaries); closed session interaction between the SDAC and DMC; hallmarks of a good SDAC; DMC recommendations including considerations for recommending premature termination of the study or alternatives; and examples of "tricky situations" sometimes faced by the DMC and the SDAC and how to deal with them.
Finally, we will review the latest draft FDA guidance on DMCs and the implications.

Open-Source Drug Development: ADAM, TLGs & Interactivity with R.
Phil Bowsher and Rich Iannone, RStudio

Posit/RStudio will be presenting an overview of Pharmaverse, GT, Quarto and WebAssembly for the R user community at MBSW. This is a great opportunity to learn and get inspired about new capabilities for working with clinical trials data and generating TLGs (Tables, Listings and Graphs) for inclusion in Clinical Study Reports. The usage of R in pharma, especially in clinical trials, has increased rapidly over the last 3 years. This talk will review various public new drug applications and discuss the usage of open source within various areas of the submissions. This is a great opportunity to learn about current trends of using open-source languages in submissions. This hands-on workshop will provide an introduction to the current landscape of prepping ADAM data and using GT, Quarto and Shinylive TLGs. In this talk, we will review and reproduce a subset of common table outputs used in clinical reporting containing descriptive statistics, counts and or percentages. The talk will provide an introduction to TFL-producing R programs and include an overview of the GT R package with applications in drug development such as safety analysis and Adverse Events. Pharmas have been using Shiny for Exploratory Data Analysis for many years. Now there is much interest in creating interactive reports with Shiny and WebAssembly to help with efficiency gains when delivering results and to save reviewers time and streamline the review process. The Shiny apps would support data review and exploration for clinical outputs. There has also been a public pilot to include Shiny in an FDA clinical trial submission as an opportunity to be more effective at processing the kinds of data and analyses that appear in clinical trial submissions. Moreover, the apps can help the regulatory bodies more quickly and accurately assess the safety and efficacy of new medical products. This talk will discuss the new horizon for new drug submissions with tools like WebR, Quarto and Shinylive. Rich and Phil will also feature ways that you and your pharma can get involved in this exciting new effort that is bringing pharma together!

MONDAY NOON-2PM JUNE 10 – POSTERS

Early and Precise Detection of Adverse Drug Events from Real-world Data

Yi Shi, IUPUI Adverse drug events (ADEs) are significant public health burden. Many ADEs cannot be detected in a timely manner after the corresponding drugs reached to market, let alone precision ADEs in patient subpopulations. The development of large-scale real-world data (RWD) including ample patients facilitate post-marketing drug surveillance. On one hand, we developed a Bayesian method to accelerate ADE detection by integrating epidemiologic design, data on newer drugs with lower sample sizes and older drugs with larger sample sizes, and drug similarity networks. On the other hand, we developed a precision mixture risk model (PMRM) to screen a tremendous amount of patient subpopulations defined by clinical features (e.g., gender, race, and all diagnoses), and to identify signals of subpopulation-specific ADEs with both confounding and false positive control. The proposed methods identified signals of ADEs for newer drugs and in subpopulations from RWD, as well as demonstrated proper false positive control and better performance compared to conventional methods. In conclusion, applications of the proposed methods hold the promise for early and precise detection of ADE.A Statistician's Guide to DMC –Overview, Standards, and Best Practice
David Kerr and Walter Boyle

Deep Learning of Sparse Irregularly Observed Multivariate Longitudinal Data

Yunyi Li, Hao Liu, Sujuan Gao, Indiana University

There is a growing interest in the analysis of multivariate longitudinal data across various scientific fields. These types of data are often observed sparsely and at irregularly spaced follow-up times, which poses unique challenges for data analysis. While statistical models specified by ordinary differential equations (ODEs) in conjunction with neural networks have emerged as a new approach to analyzing the temporal changes in multivariate longitudinal data, existing methods have primarily focused on regularly spaced and densely observed time series data.
In this paper, we present a novel method that utilizes ODEs with neural networks to address the analysis of longitudinal data that is sparsely observed at irregularly spaced follow-up times. Our proposed method combines longitudinal observations and employs stochastic optimization to estimate the initial values of ODEs. This approach is designed to be straightforward and computationally efficient. Multivariate longitudinal data often exhibit complex correlations. To evaluate the performance of our proposed method, we conducted comprehensive simulation studies, including explicit and implicit ODEs with complex variance-covariance structures for the underlying stochastic process. Additionally, we demonstrate the application of our proposed method by analyzing longitudinally observed multivariate PET imaging, CSF biomarkers, and cognitive data from the Alzheimer's Disease Neuroimaging Initiative study. This paper presents a methodology contribution by proposing neural network-based ODE estimators for modeling the dynamic change of multivariate longitudinal data with irregularly spaced follow-up times. Our method is particularly useful when ODEs cannot be formulated as parametric or semi-parametric models, which is often the case in real-world longitudinal studies.

Variable selection in a regularized multi-state model for health status transitions using electronic health records data

Jason Mao, Yang Li, Thankam Thyvalikakath, Wanzhu Tu, Indiana University

In medical and epidemiological research, longitudinal time-to-event data frequently records multiple discrete types of events of interest, as multiple events can play a role in an individual's event history. Multi-state models (MSM) are an extension of traditional time-to-event analysis that models the processes in which individuals can move across a finite number of states. These transitions between different health states are quantified through corresponding transition intensities, which are expressed as functions of patient or care characteristics. Practical analyses of MSM of real-world data (RWD), such as electronic health records (EHR), often involve high-dimensional covariates. Consequently, such analyses may encounter challenges related to model stability, parameter estimation, and inference. In this study, we propose utilizing the "minimum approximated information criterion" (MIC) method to perform variable selection, reducing the dimensionality of the model parameter space. The MIC method performs sparse estimation by incorporating an approximation of an l0 regularization term into and reparametrizing the likelihood function with the hyperbolic tangent function. Through simulation studies, we provide numerical evidence evaluating the performance of this method for model selection accuracy and parameter estimation. We demonstrate the application of the proposed method through a RWD analysis of EHR data on health-state transitions during the COVID-19 pandemic.

Identifying Predictive Combinations of Biomarkers for Early Cancer Detection with Stability Selection in Combination with Ensemble Learning

Apu Chandra Das, Lynette M. Smith, Ran Dai, Anna Lokshin , University of Nebraska Medical Center
Certain rare cancers such as ovarian and pancreatic cancer would benefit if detected early at a stage when they are resectable. Unfortunately, approved biomarkers for these cancers are not adequate for screening the general population, and it is unlikely that a single marker will meet performance criteria for screening. Determining a combination of biomarkers for early detection of rare cancers is a challenge. Often model selection suffers from overfitting in the discovery phase, which leads to poor performance upon validation. Since ovarian and pancreatic cancer have poor prognoses, we aim to identify biomarkers that perform excellently in early cancer detection discovery and validation phases. Stability selection methods have been used to prevent overfitting and to reliably select truly expressed biomarkers. Ensemble learning methods provide robust prediction results in the face of model misspecification. We present a novel framework with a biomarker selection stage with stability selection and prediction stage using an ensemble of machine learning (ML) methods, namely the stability selection ensemble learning (STABEL). The ensemble consists of random forest (RF), elastic net (EN), linear discriminant analysis (LDA), and support vector machine (SVM). Simulated data, along with ovarian cancer data from Prostate, Lung, Colorectal, Ovarian (PLCO) Cancer Screening Trial and pancreatic cancer data are used to test the methodology. Our results demonstrate the power of integrating ensemble learning with stability selection to select cancer biomarkers. The STABEL technique outperforms the existing methods, such as LASSO and RF in terms of sensitivity, specificity, prediction accuracy, and area under the curve (AUC). This technique has a broad range of applications in biomarker selection for any cancer.

Identifying Genetic Variants for Alzheimer's Disease and Comorbid Conditions

Minmin Pan, Dongbing Lai, Frederick Unverzagt, Liana Apostolova, Hugh C. Hendrie, Christianna Purnell, Andrew Saykin, Tatiana Foroud, Sujuan Gao, Indiana University

Background: Alzheimer's disease and related dementias (ADRD) frequently co-occur with comorbidities such as diabetes and cardiovascular diseases in elderly populations.

Objective: Utilize a life-course approach to identify genetic variants that are associated with the co-occurrence of ADRD and another comorbid condition.

Methods: Research data from African American participants of the Indianapolis-Ibadan Dementia Project (IIDP) linked with electronic medical record (EMR) data and genome-wide association study (GWAS) data were utilized. Age of onset for ADRD was obtained from longitudinal follow-up of the IIDP study. Age of onset for comorbid conditions were obtained from EMR. The analysis included 1,177 African Americans, among whom 174 were diagnosed with ADRD. A semi-parametric marginal bivariate survival model was used to examine the influence of single nucleotide polymorphisms (SNPs) on dual time-to-event outcomes while adjusting for sex, years of education, and the first principal component of GWAS data.

Results: Targeted analysis of 20 SNPs that were reported to be associated with ADRD revealed that six were significantly associated with dual-disease outcomes, specifically congestive heart failure and cancer. In addition, eight novel SNPs were identified for associations with both ADRD and a comorbid condition.

Conclusions: Using a bivariate survival model approach, we identified genetic variants associated not only with ADRD, but also with comorbid conditions. Our utilization of dual-disease models represents a novel analytic strategy for uncovering shared genetic variants for multiple disease phenotypes.

TUESDAY MORNING 8:30 – 11:30 JUNE 11

Clinical: Modern Design Approach

Bayesian Adaptive Dual Objective Phase I-II designs for Personalized Dose-Finding with Combination Therapies
Brad Carlin, Cencora-PharmaLex
James Willard, McGill University; Lira Pi, Cencora-PharmaLex

We describe a Phase I-II design for testing a new cancer drug, where interest lies in its value as both a monotherapy, and in combination with a second drug. Phase I begins with a monotherapy "run-in" period that can model efficacy alone, or use a clinical utility index to trade off safety and efficacy, and can capture correlation among the two competing endpoints. Our bivariate dosing model employs Bayesian Optimization (BO) over a bivariate Gaussian process approximation, providing smooth and efficient estimation over the two-dimensional dosing grid. Ultimately, the trial identifies a recommended dose region, from which the two or three doses can be selected for Phase II comparison with the optimal monotherapy dose and placebo. We evaluate both stages of our design using simulation, where in Phase I we study the probability of correct dose selection and a related root mean squared error (RMSE) criterion, while in Phase II we return to the traditional benchmarks of Type I error and power. The proposed design appears to satisfy modern Project OPTIMUS-inspired regulatory guidelines for Phase I-II oncology trials, while offering improved efficiency, flexibility, and interpretability. We also offer an illustration of the approach in a non-cancer setting, where a sponsor is interested in a design for the development of an intraocular implant that combines two topical agents. In this setting, response heterogeneity is expected to exist with respect to a key binary covariate (a particular characteristic of the lens of the eye), and we seek a design that accommodates this feature.

Advancing Pediatric Clinical Trials through Case Studies: Extrapolation, Bayesian Methods and Regulatory Perspectives
Fei Chen, Johnson & Johnson Innovative Medicine
Robert (Skip) Nelson, Johnson & Johnson Innovative Medicine
Pediatric populations almost always present unique challenges in clinical trials, including ethical considerations, recruitment difficulties, and physiological differences of children from adults. These challenges frequently result in a scarcity of direct clinical trial data for pediatric treatments, making extrapolation not just beneficial but essential. By highlighting Bayesian methods, the talk will shed light on innovative ways to address these data gaps, ensuring that medical treatments for children are based on robust and scientifically sound evidence.

The primary objective of this talk is to learn from real-world experiences of past pediatric clinical studies, examining both successes and challenges. We aim to provide a talk that will
benefit practitioners involved in pediatric trials and inform innovations in study designs. Case studies will highlight how Bayesian extrapolation can be used effectively and lead to successful outcomes. Limitations of extrapolation in certain contexts and potential pitfalls in statistical methodology will be discussed, focusing on lessons learned and strategies forward. In reviewing past studies, we also hope to share insights into the regulatory review process and scenarios under which different degrees of extrapolation may be appropriate.

Some topics of interest that we will be highlighting during the roundtable discussion:
Variability in treatment effects between pediatric and adult trials: the effect of dosing, duration, and demographics
Enrollment and retention challenges in pediatric trials: ways to mitigate this problem?
Borrowing information from adult trials when pediatric trials fall short: what statistical methods can one adopt and implement?
Bayesian borrowing: key ideas, software, and case studies (examples)
Regulatory guidance and current perspectives

Design considerations for two-stage enrichment clinical trials
Zhantao Lin, Eli Lilly and Company
Rosamarie Frieri; William Fisher Rosenberger; Nancy Flournoy, Eli Lilly and Company
When there is a predictive biomarker, enrichment can focus the clinical trial on a benefiting subpopulation. We describe a two-stage enrichment design, in which the first stage is designed to efficiently estimate a threshold and the second stage is a "phase III-like" trial on the enriched population. The goal of this paper is to explore design issues: sample size in Stages 1 and 2, and re-estimation of the Stage 2 sample size following Stage 1. By treating these as separate trials, we can gain insight into how the predictive nature of the biomarker specifically impacts the sample size. We also show that failure to adequately estimate the threshold can have disastrous consequences in the second stage. While any bivariate model could be used, we assume a continuous outcome and continuous biomarker, described by a bivariate normal model. The correlation coefficient between the outcome and biomarker is the key to understanding the behavior of the design, both for predictive and prognostic biomarkers. Through a series of simulations we illustrate the impact of model misspecification, consequences of poor threshold estimation, and requisite sample sizes that depend on the predictive nature of the biomarker. Such insight should be helpful in understanding and designing enrichment trials.

ENRICH: A Novel Approach to Clinical Trials in ICH
Giorgo Paulon, Berry Consultants, LLC
Past clinical trials of surgical evacuation of intracerebral hemorrhage (ICH) have generally not demonstrated functional benefit. One of the main challenges in ICH clinical trials is the heterogeneity in patient populations, leading to difficulties in detecting treatment effects. Enrichment strategies, aimed at optimizing trial design and participant selection, have emerged as a crucial component in enhancing the efficiency and success of these trials by identifying and selecting patient subgroups more likely to benefit from the investigational arm. Additionally, adaptive trial designs that allow for modification to key trial parameters in response to accumulating data according to pre-specified rules can further optimize the trial conduct by stopping the trial early for futility or expected success.

This talk explores the application of enrichment strategies by presenting a recently completed successful clinical trial of early minimally invasive surgical removal of ICH compared to medical management. Moreover, important considerations for enrichment designs in trial implementation will be discussed.

TUESDAY MORNING JUNE 11

Computation & Visualization

Making an Interactive Shiny App with GenAI…Five Ways
Phil Bowsher, POSIT
Posit/RStudio will be providing a session on using AI to enhance open-source statistical programming and Shiny app creation. This session will discuss opportunities and applications for GenAI to empower statistical programmers. This talk will explore and discuss 5 ways to make Shiny apps with GenAI. Talk will discuss ways to use AI to support programmers in the process of writing code and provide a hands-on opportunity to test it out.

Visualizing Insights, Empowering Discoveries: SAS Viya Unleashed in Biopharmaceutical Analytics
Matt Becker, SAS
The biopharmaceutical industry has recently produced a tremendous amount of data at a quick rate from a variety of sources, including clinical trials, genomics, proteomics, and patient records. For improvements in research, medication discovery, and patient care, it is essential to draw insightful conclusions from this data and make defensible decisions. In this setting, SAS Visual Analytics (SAS VA), which provides sophisticated data visualization, analytics, and exploration capabilities, emerges as a potent tool. This paper explores the application of SAS Visual Analytics in the biopharmaceutical sector, highlighting its benefits, features, and real-world use cases.

Human Perception for Information Visualization Design

Dr. Yingjie Victor Chen, Purdue University
Understanding human perception is vital for creating effective visualizations of complex data. This talk will explore foundational theories of visual perception, including color theory, pattern recognition, and spatial reasoning, and discuss how these theories inform the creation of intuitive and impactful data visualizations. By examining case studies and current research, attendees will gain insights into the best practices for designing visualizations that are not only aesthetically pleasing but also facilitate quick and accurate data comprehension. This talk aims to bridge the gap between cognitive psychology and design practice, providing attendees with the knowledge to craft visualizations that cater to the complex workings of human perception.

Topline Results Dashboard with R Markdown & Flexdashboard
Robert Adams, Bayer AG
Topline Results (TLRs) are key results of a clinical trial comprising pre-selected sets of demographic, safety, and efficacy (primary and secondary endpoints) summaries. Data for TLRs are usually generated within hours after unblinding / locking of database and are therefore a time critical. Due to that specific requirements TLRs are often reported as statistics printed in tables or figure formats that might not provide instinctive insights into the underlying trends in results. The approach is follows: We used R Markdown and the flexdashboard package to generate a visual dashboard from TLRs that allow easier navigation and better visual understanding of TLRs. The method has some advantages, inter alia, the result of the process is a self-contained HTML file that comprises the data, HTML, CSS, JavaScript, figures and tables. Since R Markdown and flexdashboard are used as wrapper packages there is no need to be familiar with any of the web technologies besides basic R knowledge. Although the document can be navigated in an interactive fashion (e.g. by scrolling, tab selection, javascript image carousels), the output is a deliverable is a static HTML document that requires no hosting on servers and has no function for filtering or dynamic loading of data which makes is suitable for a safe communication of TLR. It can therefore be regarded as a modern substitute to slide decks or word / PDF documents.

TUESDAY MORNING JUNE 11

Artificial Intelligence and Machine Learning

A Large Language Model (LLM)-powered Workflow Applied to Pharmaceutical Informed Consent
Jacob Gagnon, Biogen
Hangyu Liu, Andrew Borgman, Jake Gagnon, Dan Boisvert, Haleh Valian, Yuka Moroishi, Biogen
The pharmaceutical industry faces challenges in re-using samples, data, and images for secondary research. Complex guidelines governing data re-use are often buried in Informed Consent Forms (ICFs) that are manually reviewed by subject matter experts (SMEs) on a case-by-case basis. This process is time-consuming and hampers efficiency. We developed a novel Artificial Intelligence (AI)-driven workflow leveraging Large Language Models (LLMs) for computer-assisted analysis of ICFs. We tested a variety of open-source models, including LLaMA, FLAN-T5, Llama-2-13b-chat, Wizard LM 70B, all-mpnet-base-v2, and bge-large-en, as well as some commercial models, such as GPT-3.5-Turbo, GPT-4, text-embedding-ada-002 etc. for text generation and embedding. Although LLMs have many advantages in natural language processing (NLP), it lacks domain knowledge, and sometimes hallucinates. We implemented a Retrieval-Augmented Generation (RAG) framework to address these challenges and simplifies our process by using generative AI to summarize and identify allowable data usage promptly and precisely. This adaptable framework is applicable to any company requiring comprehension and analysis of ICFs or data use agreements. Through a domain-specific validation framework, combining manual SME review and quantitative measurements, we achieved 96% cosine similarity compared to human responses in Biogen study ICFs tested to date. The talk delves into the current workflow design, associated cost estimation, a multi-layered strategy to address potential hallucinations in LLMs, considerations for safeguarding data privacy within a secure computing environment, the design and development of an interactive Streamlit chatbot powered by LLMs, and other pertinent technical details. Our overarching goal is to enhance data reusability by improving the efficiency of ICF document analysis and expediting the dissemination of therapeutic insights to the patient community.

Navigating the Landscape: A Framework for Evaluating Large Language Models (LLMs) in Biostatistics

Sydeaka Watson, Eli Lilly & company
Large Language Models (LLMs) are revolutionizing various fields with their capabilities in information processing, text generation, and data analysis. This presentation explores the potential of LLMs to streamline and enhance biostatistical workflows within the pharmaceutical industry. We present a practical framework for evaluating LLM tools tailored to the specific needs of a biostatistics unit. This framework outlines sample prompts designed to assess LLM performance across various biostatistical tasks, including literature search support, statistical code generation and review, report editing, and data exploration. The talk addresses key considerations for LLM implementation, such as data security, explainability, and integration with existing workflows. We discuss strategies to mitigate potential risks associated with LLM use in biostatistics, ensuring responsible and ethical integration. This presentation concludes by highlighting the importance of ongoing evaluation and human oversight. By effectively evaluating and utilizing LLMs, biostatistics units can leverage these powerful tools to improve efficiency, accuracy, and communication within the drug development process.

Exploring Bias in AI and Machine Learning
Jim Box, SAS
Implementing AI and Machine Learning models has the potential for revolutionizing pharmaceutical research (and society in general). But with this great power comes the responsibility to ensure these systems are produced and implemented in a fair manner. We'll take a look at the ways our very human biases can impact the machine learning process, and explore the responsibilities all of us have as creators and implementors of these systems.

Leverage AI to build end-to-end workflow for research, analysis, and reporting
Michael Man, Eli Lilly

Abstract coming soon.

TUESDAY MORNING JUNE 11

HTA Statistics, Benefit-Risk, HEOR

Presentation Title: Novel yet Practical Bayesian Quantification of Benefit-Risk Balance of Medical Products
Saurabh Mukhopadhyay, PhD, AbbVie
Evaluation of the benefit-risk profile is a key element throughout the life cycle of a medical product. There is an increasing emphasis on quantitative benefit-risk assessments, as it aims to define explicit, transparent, and evidence-based criteria to compare benefit-risk profiles of different treatments that can inform sponsors, regulators, payers, and other stakeholders at different stages of its entire life cycle. One challenge in implementing quantitative benefit risk is the elicitation of utility functions to map observed outcomes to numeric scores. In this presentation, we first discuss a Bayesian framework that provides an explicit and efficient way to estimate preferences using trade-off data from stakeholders. A key gap remains in how to effectively combine preferences elicited with data from clinical trials. Therefore, we next characterize the uncertainty of the performance of a treatment based on clinical trial data. Finally, we discuss how to combine preferences with outcomes of treatment to define an overall benefit-risk utility score of a treatment. The methodology also allows stochastic comparisons between two or more treatments or doses to understand relative preferences and associated uncertainties. This comprehensive approach is also easy to implement in practice and can leverage aggregate-level clinical trial data to estimate treatment performance. We demonstrate the entire process using simulated data and an R-shiny app.

Bayesian Benefit Risk Analysis Using the brisk R Package
Richard Payne. PhD, Eli Lilly
Quantitative methods for benefit-risk analysis help to condense complex decisions into a univariate metric to describe the overall benefit relative to risk. One such approach is the multi-criteria decision analysis (MCDA) framework. Bayesian benefit-risk analysis incorporates uncertainty through posterior distributions which are inputs to the benefit-risk framework. This talk will focus on the brisk R package (available on CRAN) which provides functions to assist with Bayesian benefit-risk analyses, such as MCDA. The package requires users to input posterior samples, utility functions, weights, and the package outputs posteriors of benefit-risk scores. The software and methods will be illustrated with several examples.

Presentation Title: Recent Developments and Innovations in Health Technology Assessment
Hongwei Wang, PhD, AbbVie
Health Technology Assessment (HTA) is a comprehensive evaluation process that systematically examines health technologies, including medications, medical devices, and prevention methods, taking into account medical, economic, social, and ethical factors. Its primary objective is to provide evidence-based information to national authorities to support decisions regarding reimbursement and pricing. To ensure consistency and guidance in HTA submissions, health authorities responsible for reimbursement and pricing have issued draft guidance documents. Some notable recent developments include the mandatory joint clinical assessment of advanced therapies by European Medical Agency and HTA bodies starting in 2025, the US Inflation Reduction Act allowing Medicare to negotiate directly with drug manufacturers on the price that will take effect in 2026. In addition to the changes to processes, the evidence requirements for reimbursement are also evolving with robust real-world evidence, advanced analytics for personalized medicine, and sophisticated statistical analysis methodologies to meet the HTA needs. For example, health authorities not only consider clinical effectiveness and safety but also place significant importance on the long-term effectiveness and comparative effectiveness of a technology compared to the current standard of care. Since clinical trials often have limited durations, innovative statistical methodologies are frequently used in HTA submissions to predict longer-term effects. Comparative effectiveness, both through direct and indirect comparisons, is another statistical application utilized in HTA submissions to assess new technologies against the standard of care. Moreover, in disease areas such as oncology, patients in clinical trials often switch to more effective treatments than the one they were initially randomized to. Determining the actual effect of a technology in the presence of treatment switching is another focus area in HTA submissions. In this session, our speakers will explore the current HTA landscape with the new EU HTA regulation along with presenting innovative statistical methodologies that can address the complex HTA requirements.

CLINICAL: Meta-analyses and indirect treatment comparisons

Tipping point analyses for the between-group correlation in an arm-based evidence synthesis

Wenshan Han, Florida State University

Co-authors: Zheng Wang2; Mengli Xiao3; Zhe He4; Haitao Chu5,6* Lifeng Lin7,*

1Department of Statistics, Florida State University, Tallahassee, FL, USA. 2Department of Biostatistics and Research Decision Sciences, Merck & Co., Inc, Rahway, NJ, USA. 3Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. 4School of Information, Florida State University, Tallahassee, Florida, USA. 5Global Biometrics and Data Management, Pfizer Inc., New York, NY, USA. 6Division of Biostatistics and Health Data Science, School of Public Health, University of Minnesota, Minneapolis, MN, USA. 7Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, USA

Systematic reviews and meta-analyses are essential tools in contemporary evidence-based medicine, synthesizing evidence from various sources to better inform clinical decision-making. However, the conclusions from different meta-analyses on the same topic can be discrepant, which has raised concerns about their reliability. One reason is that the result of a meta-analysis is sensitive to factors such as study inclusion/exclusion criteria and model assumptions. The arm-based meta-analysis model is growing in importance due to its advantage of including single-arm studies and historical controls with estimation efficiency and its flexibility in drawing conclusions with both marginal and conditional effect measures. Despite its benefits, the inference may heavily depend on the heterogeneity parameters that reflect design and model assumptions. This project aims to evaluate the robustness of meta-analyses using the arm-based model within a Bayesian framework. Specifically, we develop a tipping point analysis of the between-group correlation parameter to assess the robustness of meta-analysis results. Additionally, we introduce some visualization tools to intuitively display its impact on meta-analysis results. We demonstrate the application of these tools in real-world meta-analyses with single-arm studies.

Alternatives to Normal-Normal Meta Analysis

Tim Hanson, Medtronic

The common assumption of normality of treatment effects in meta-analysis is often violated, leading to estimation bias, improper interval coverage, and inflated Type I error. In this talk we will discuss these issues and review some alternatives that do not assume normal treatment effects, including richer parametric models that allow for skew, and some nonparametric approaches. Several examples will be provided from the biopharmaceutical and medical device literature as well as R code for carrying out analyses. The talk will culminate with recent work on a nonparametric approach to estimating reference intervals from meta-analytic data.

Network Meta Analysis to Predict the Efficacy of an Approved Treatment in a New Indication

Jennifer Proper, Eli Lilly

Haitao Chu, Purvi Prajapati, Michael D. Sonksen, and Thomas A. Murray, Eli Lilly

Drug repurposing refers to the process of discovering new therapeutic uses for existing medicines. Compared to traditional drug discovery, drug repurposing is attractive for its speed, cost, and reduced risk of failure. However, existing approaches for drug repurposing involve complex, computationally-intensive analytical methods that are not widely used in practice. Instead, repurposing decisions are often based on subjective judgements from limited empirical evidence. In this work, we develop a novel Bayesian network meta-analysis (NMA) framework that can predict the efficacy of an approved treatment in a new indication and thereby identify candidate treatments for repurposing. We obtain predictions using two main steps: first, we use standard NMA modeling to estimate average relative effects from a network comprised of treatments studied in both indications in addition to one treatment studied in only one indication. Then, we model the correlation between relative effects using various strategies that differ in how they model treatments across indications and within the same drug class. We evaluate the predictive performance of each model and find that the model minimizing root mean squared error of the posterior median for the candidate treatment depends on the amount of available data, the level of correlation between indications, and whether treatment effects differ, on average, by drug class. We conclude by discussing

an illustrative example in psoriasis and psoriatic arthritis and find that the candidate treatment has a high probability of success in a future trial.

A comprehensive review and shiny application on the matching-adjusted indirect comparison
Haitao Chu, Pfizer Inc.
Ziren Jiang, Joseph C Cappelleri, Margaret Gamalo, Yong Chen, Neal Thomas , Pfizer Inc.
Population-adjusted indirect comparison (PAIC) is an increasingly used technique for estimating the comparative effectiveness of different treatments for the health technology assessments when head-to-head trials are unavailable. Three commonly used PAIC methods include matching-adjusted indirect comparison (MAIC), simulated treatment comparison (STC), and multilevel network meta-regression (ML-NMR). MAIC enables researchers to achieve balanced covariate distribution across two independent trials when individual participant data are only available in one trial. In this article, we provide a comprehensive review of the MAIC methods, including their theoretical derivation, implicit assumptions, and connection to calibration estimation in survey sampling. We discuss the nuances between anchored and unanchored MAIC, as well as their required assumptions. Furthermore, we implement various MAIC methods in a user-friendly R Shiny application Shiny-MAIC. To our knowledge, it is the first Shiny application that implements various MAIC methods. The Shiny-MAIC application offers choice between anchored or unanchored MAIC, choice among different types of covariates and outcomes, and two variance estimators including bootstrap and robust standard errors. An example with simulated data is provided to demonstrate the utility of the Shiny-MAIC application, enabling a user-friendly approach conducting MAIC for healthcare decision-making. The Shiny-MAIC is freely available through the link: https://ziren.shinyapps.io/Shiny_MAIC/.

TUESDAY AFTERNOON JUNE 11

Real World Evidence

Revolutionizing Infectious Disease Management through Real World Evidence
Gene Olinger, Oyanalitika
Utilization of real world data such as genetic predispositions and community microbial profiles has potential to enhance vaccine development and treatment in the field of infectious diseases. Real world evidence (RWE) has the potential to address challenges like antibiotic resistance and optimizing vaccine responses. For instance, by leveraging real world data obtained from genetic sequencing and immunological profiling, researchers can identify specific vulnerabilities and design targeted interventions for enhanced outcomes for patients. High resolution diagnostics approaches that provide pathogen identification, resistance, and prognostic biomarkers will enable easier stratification of patients during acute phases of infection and clinical sequela. While limited resources in some settings, the implementation of RWE can contribute towards better understanding of unique factors contributing to confounding of outcomes in real world settings which could be quite heterogeneous in for instance different locations, climates and vulnerable populations. RWE applications in infectious diseases continue to evolve by harnessing innovative analytical approaches for tackling infectious diseases, optimizing treatment efficacy, and curbing the spread of antimicrobial resistance.

Real-world evidence for pharmacovigilance and drug repurposing
Pengyue Zhang, Indiana University
Adverse drug events (ADEs) and Alzheimer's disease (AD) are significant burdens to public health. Real-world data mining has the potential to identify precision drug effects in subpopulations. This presentation includes two exemplified health insurance claim data-based analyses to identify: 1) ADEs in patient subpopulations defined by clinical features (e.g., age, gender, comorbidities etc.) and diagnosis codes; and 2) repurposable treatments for AD with a special conscious for African Americans.

Sensitivity analysis for evaluating treatment effects from real world data
Mingyang Shan, Eli Lilly
Abstract Coming soon

Historical data borrowing for predictive biomarker identification
Adam Luo, Abbvie
Identification and evaluation of predictive biomarkers are critical to the development of novel medicine in Oncology as they can guide treatment decisions and help identify disease subpopulation with better treatment efficacy. However, for phase 1 trials with single treatment arms, it is difficult to differentiate predictive from prognostic biomarkers in the absence of control arms. We set out to establish a predictive analysis pipeline for early-phase oncology trials with borrowed controls. Its predictive biomarker signature development by proposed analysis pipeline can shed light on the asset development.
Method: Driving by the research questions during clinical trials, our analysis pipeline consists of three steps: 1) evaluation of data resources, 2) control data matching, 3) downstream analysis. The evaluation step is to evaluate a variety of historical oncology trials as well as real-world data resources that can potentially be used for control data borrowing. The selected datasets will constitute our control data warehouse with a guidance document for key variable evaluation and analysis recommendations. The control matching step first performs treatment landscape analysis to better understand the treatment composition and decide the inclusion/exclusion criteria for control sample pool. We preprocess the clinical and biomarker data by going through outlier detection, missing data imputation, and data filtering. The iterative matching process utilizes propensity score estimation, data matching methods, and balance diagnostics. The downstream analysis follows the conventional univariate and multivariate modules in predictive analysis workflow.
Discussion: Controlled data borrowing in phase 1 clinical trial offers the advantage of optimizing limited resources, cost-effective utilization of historical trial data and real-world data resources, and potentially expediting the downstream biomarker analysis and further evaluation of new intervention. Leveraging control data from historical trials offers the advantage of consistent endpoint criteria and controlled conditions, while borrowing from real-world data resources provides access to diverse patient profiles and reflects actual clinical practices. However, real-world data resources might vary in quality and accuracy, which leads to potential biases, so data quality control is essential.

TUESDAY AFTERNOON JUNE 11

Biomarkers and pre-clinical

Bayesian-frequentist hybrid inference framework for single cell RNA-seq analyses
Dongyan Yan, Eli Lilly and Company
Gang Han, Zhe Sun, Jiyuan Fang, Xinyue chang, Lucas Wilson and Yushi Liu, Eli Lilly and Company
Single cell RNA sequencing technology (scRNA-seq) has been proven useful in understanding cell-specific disease mechanisms. However, identifying genes of interest remains a key challenge. Pseudo-bulk methods that pool scRNA-seq counts in the same biological replicates have been commonly used to identify differentially expressed genes. However, such methods may lack power due to the limited sample size of scRNA-seq datasets, which can be prohibitively expensive. Motivated by this, we propose using the Bayesian-frequentist hybrid (BFH) framework to increase the power of identifying differentially expressed genes. We show in a simulated scenario that the proposed BFH would be an optimal method when compared with other popular single cell differential expression methods if both FDR and power are considered. As an example, the method was applied to an idiopathic pulmonary fibrosis (IPF) case study. In our IPF example, we demonstrated that with a proper informative prior, the BFH approach identified more genes of interest than traditional single cell methods. Furthermore, these genes were reasonable based on the current knowledge of IPF. Thus, the BFH offers a unique and flexible framework for future scRNA-seq analyses.

Molecular Representation Learning for Biological Metrology
Dr. Haixu Tang, Luddy School of Informatics, Computing, and Engineering, Indiana University
In this talk, I will introduce 3D Molecular Network, a deep neural network based on an elemental operation of 3D molecular convolution on the 3D conformations of compounds for molecular representation learning (MRL). I will discuss three applications of this model in biological metrology, for the prediction of tandem mass (MS/MS) spectra, the chromatographic enantioseparability, and the response rate in mass spectrometry, respectively.

Leveraging Generative AI Technology for Accelerated Analysis of Scientific Publications in Pre-clinical Medical Research

Steven Silva, Cencora-PharmaLex
Nikhil Damle, Rafael Mundim, Cencora-PharmaLex

In the realm of pre-clinical medical research, the analysis of vast amounts of scientific publications is a time-consuming task for highly trained subject matter experts (SMEs). Artificial Intelligence (AI) tools based on large language models (LLMs) offer a promising solution to expedite these tasks.
Bibliometric review methods based on traditional relevance and impact criteria are prone to oversight of valuable information due to the sheer volume and diversity of content. AI is able to rapidly extract insights based on queries, which could enable SMEs to uncover relevant subsets of information that might have been missed through conventional filtering approaches done by a human.
Drawing parallels from other industries where the technology is more widely used, we investigated the use of generative AI, both in its generic form and when specifically trained on medical corpora, to streamline the bibliography analysis process.
We set up a proof-of-concept experiment comparing human-led tasks with and without the aid of generative AI. We used openly available models as input to generate initial insights as well as for identifying potentially relevant topics to be reviewed by the SME, according to a set of user prompts. Intrinsic issues of the generative AI models were observed which can be mitigated with the use of relevance and consensus metrics to aid human judgement. Despite these issues, we achieved overall gains in efficiency as compared to the traditional analysis methods. Generative AI can be used in this context with the due caution to expedite the analysis of scientific publications in pre-clinical medical research, offering a pathway to enhanced efficiency, cost savings, and expediting discoveries. We advise for a judicious integration of AI for bibliography review workflows where human checkpoints ensure trustworthiness and decision making based on evidence.

Amyloid Reduction as a Surrogate Biomarker in Alzheimer's Disease Clinical Trials: Exploring the Path Forward Following FDA Guidelines

Guoqiao, WASHU

Abstract coming soon

TUESDAY AFTERNOON JUNE 11

CMC

Frequentist and Bayesian tolerance intervals for setting specification limits for left-censored gamma distributed drug quality attributes

Richard Montes, Alnylam Pharmaceuticals

Tolerance intervals from quality attribute measurements are used to establish specification limits for drug products. Some attribute measurements may be below the reporting limits, that is, left-censored data. When data has a long, right-skew tail, a gamma distribution may be applicable. This paper compares maximum likelihood estimation (MLE) and Bayesian methods to estimate shape and scale parameters of censored gamma distributions and to calculate tolerance intervals under varying sample sizes and extents of censoring. The noninformative reference prior and the maximal data information prior (MDIP) are used to compare the impact of prior choice. Metrics used are bias and root mean square error for the parameter estimation and average length and confidence coefficient for the tolerance interval evaluation. It will be shown that Bayesian method using a reference prior overall performs better than MLE for the scenarios evaluated. When sample size is small, the Bayesian method using MDIP yields conservatively too wide tolerance intervals that are unsuitable basis for specification setting. The metrics for all methods worsened with increasing extent of censoring but improved with increasing sample size, as expected. This study demonstrates that although MLE is relatively simple and available in user-friendly statistical software, it falls short in accurately and precisely producing tolerance limits that maintain the stated confidence depending on the scenario. The Bayesian method using noninformative prior, even though computationally intensive and requires considerable statistical programming, produces tolerance limits which are practically useful

for specification setting. Real-world examples are provided to illustrate the findings from the simulation study. (https://doi.org/10.1002/pst.2344)

OMARS designs, bridging the gap between screening and optimization experimental designs
José Núñez Ares & Peter Goos, EFFEX
In 2011, Jones and Nachtsheim introduced the Definitive Screening Designs as three-level designs for factor screening and optimization. The underlying idea of their proposed experimental designs is that, in the presence of effect sparsity, a single experimental design can be used for both screening and optimization. Ockuly et al. show that effect sparsity holds in a collection of 129 experimental data sets with a total of 183 responses. In recent years, OMARS designs have been introduced. The OMARS family of designs extends the catalog of designs that allow screening and optimization in a single step substantially, and they can be used for quantitative, two-level categorical, and blocking factors. In this talk we will demonstrate a multi-criteria design selection algorithm.
Typically, one-step designs for simultaneous screening and optimization involve a large number of factors, making the analysis of experimental data difficult. The challenge is to select the most influential effects and construct a meaningful model, given the fact that there is a large number of potential effects. Recent work has proposed analysis techniques tailored to these designs. We will show how the analysis of OMARS benefits from the use of an all-subsets model selection technique based on integer programming, which can handle problems with more than a hundred potential effects.
Finally, multi-response optimization problems are the norm in the pharmaceutical industry. Defining the design space in the presence of multiple and often conflicting critical quality attributes is challenging. We will show how interactive graphical interfaces can be used to explore the probability of success of being with specifications and assess the robustness of different candidates for values of the critical process parameters.
The demonstrations will be made using a novel web-based software and an example inspired by a real case in the pharmaceutical industry.

Reliably Assessing Comparability in Autologous Cell Therapy Change Protocols: A Compliance Perspective
Bruno Boulanger & Jean-François Michiels, Cencora-Pharmalex
A Change Protocol (CP) is a prospective plan to evaluate the impact of proposed bioprocess changes post-approval on the identity, strength, quality, purity, and potency of a drug product or biological product. This assessment is crucial for ensuring product quality in compliance with regulatory requirements.
Autologous cell therapy, involving the processing of an individual's cells or tissues outside the body and their subsequent reintroduction, poses unique challenges when changes are introduced to the bioprocess. Given the personalized nature of autologous therapy, the comparison between processes before and after changes, but within donors, becomes imperative. Current regulation published by FDA primarily assume uniformity across batches and do not specifically address the intricacies of within-donor comparison.
Our presentation tackles this gap by proposing a solution that establishes and justifies acceptance limits for a Two One-Sided Test (TOST) while considering within-donor comparison. This involves a comprehensive study design and addresses the fundamental question of predefining acceptance limits. The current FDA regulation lack specific provisions for this scenario, and our approach fills this void.
Additionally, we introduce a method that leverages historical data from the original process. This allows the Change Protocol to be executed with a limited number of donors and batches, facilitating a more efficient evaluation. Furthermore, we emphasize the importance of ensuring the capability of the new process, guaranteeing the maintenance of quality, efficacy, and safety in the future.
Our presentation will demonstrate the application of Bayesian statistics in deriving and justifying acceptance limits, taking into account within-donor evaluations. By integrating historical data and leveraging Bayesian statistics, our approach offers a robust framework for reliably assessing comparability in autologous cell therapy change protocols, ensuring compliance with regulatory standards.

An Overview of Statistical Approaches for Classifying Product Quality Attributes as "Stability-Indicating"
Jeff Gardner, DataPharm SDMS
For mature commercial biopharm products where a manufacturer has many lots that have completed stability testing protocols it can become cost-intensive to maintain the stability testing scope and schedule filed for commercial approval. In such instances the manufacturer may wish to eliminate testing that is non-value-added and instead focus only on maintaining testing for "stability-indicating" quality attributes. This lecture discusses potential statistical methodologies for defining "stability-indicating" quality attributes, including strengths and

weaknesses. Special attention will be given to perspectives from industry and regulatory agencies regarding the potential risks of pursuing stability testing efficiencies via these statistical methodologies.

Low Correlation Designs for Demonstrating Proven Acceptable Ranges and Informing Design Spaces
Tim Kramer, Eli Lilly and Company
Pre-submission synthetic molecule activities include identifying risks and demonstrating proven acceptable ranges for process parameters associated with various drug substance synthesis steps. Frequently, process chemists and engineers will conduct an initial risk assessment of potential parameters and identify three subsets—those that likely to impact the process (key parameters), others that are not expected to affect the process (secondary, low risk parameters), and a third group of no-risk parameters that are not further studied. Additionally, specific parameter combinations are identified that could potentially interact relative to the control of specific impurities. The development teams then conduct experiments that demonstrate acceptable parametric ranges and provide information for developing design spaces. Resource limitations generally require few experiments and designs that have correlations between parameters are frequently used. This presentation will present a class of low correlation designs developed specifically for evaluating many parameters in a small number of runs. These designs utilize the information from the initial risk assessments. Some of the benefits and drawbacks of these designs will be discussed.

WEDNESDAY MORNING JUNE 12

Clinical Advanced Questions

Power Performance for Testing a Composite Outcome with Correlated Endpoints using the Finkelstein-Schoenfeld Test
Yu-Hsiang Shu, Hsin-Yu Hsu, Yu-Chen Su, Inari Medical
In 1999, Finkelstein and Schoenfeld introduced the idea of prioritizing the pairwise comparisons of endpoint components in testing a composite outcome. The priority is based on the clinical importance and realized as a hierarchy structure in comparisons. The FS test is expected to be more powerful when the priority is also true to the treatment effects and each added endpoint can bring additional information to the test. However, a recent FDA report shows that the power may not be consistently increasing with additional endpoints when the endpoints included in the composite outcome are correlated. In our simulations, we generated binomial endpoints using the iterative proportional fitting procedure (IPFP) and continuous endpoint using the rank-based inverse normal transformation (INT) to approximate exact correlation assignments. We explored the impacts of the number of endpoints, priority, effect sizes, and correlations in a composite outcome. When all endpoints are binomial, higher correlations between endpoints may lead to greater reduction in power performance. Adding more endpoints may not necessarily increase power. In fact, the power can decrease if the added endpoints have effect size smaller than the previous ones. When a continuous endpoint is included, one should be aware whether the FS test power may be driven by the single continuous endpoint because all the remaining tie will be explained by it. With higher correlations between the endpoints, more ties may remain after binomial endpoints, and the power performance determined with the additional information brought by the continuous endpoint will be more conditional on the effect sizes of the previous binomial endpoints and itself. For power calculations for the FS test in a RCT proposal, we recommend conducting simulations specific to the study parameters with the possible correlations between endpoints properly included.

Two basic statistical strategies of conducting causal inference in real-world studies
Yixin Fang, PhD, AbbVie
In the literature of causal inference, a variety of statistical methods have been proposed to adjust for confounding bias. However, it is challenging for the users to understand the statistical properties enjoyed by each method and then explicitly specify its underlying model assumptions. In this presentation, we investigate two basic statistical strategies of conducting causal inference in real-world studies, which cover many commonly used methods. These two strategies are the weighting strategy and the standardization strategy. The weighting strategy defines a target

estimand using a propensity-score model (treatment assignment ~ confounders), while the standardization strategy defines an estimand using an outcome-regression model (outcome variable ~ treatment assignment + confounders). Although these two strategies are different at the beginning, at the end they are robust for estimating the treatment effect under the same set of identifiability conditions and therefore same kind of sensitivity analysis is needed for evaluating the impact caused by the violation of these conditions.

Optimal treatment regimes for behavioral interventions using multiple surrogate outcomes
Eric Laber, Duke University
Behavioral interventions delivered by mobile-health (mHealth) often seek to affect long-term outcomes such as body mass index (BMI), prolonged substance-use cessation, resting heart rate, and VO2 max. Often, the effect of any single behavioral intervention on the long-term outcome of interest is imperceptible relative to momentary fluctuations. Thus, adaptive algorithms for mHealth typically work through multiple surrogate outcomes; e.g., an mHealth intervention for BMI reduction in overweight individuals might target surrogates such as step count, calorie intake, or mindfulness. While the relationship between surrogates and the outcome of interest is typically unknown, one often has access to partial directional information; e.g., in the context of BMI reduction, more steps are better than fewer steps, and more app engagement is better than less, but it is not clear if taking five-hundred additional steps is better than spending five additional minutes engaged with the app. We consider optimal sequential decision masking with surrogates that admit a partial ordering. We show that a partial ordering on the surrogates corresponds to a stochastic partial order over the space of treatment regimes from which we derive an estimator of the set of maximal treatment regimes. Empirical experiments show that bandit algorithms that make use of partial ordering by restricting to maximal treatment regimes improve patient outcomes, especially when momentary treatment effects are small. We provide an illustrative example based on the ADAPT mHealth study for BMI reduction in overweight subjects with type 1 diabetes (T1D).

Causal Inference in treatment effect estimation using longitudinal real-world data for precision medicine.
Jing Su, Biostatistics and Health Data Science, Indiana University School of Medicine
Causal Inference in treatment effect estimation using longitudinal real-world data for precision medicine.
In this talk, I will use study cases to demonstrate the importance and challenges of causal inference using longitudinal real-world data. I will also discuss statistical, machine learning, and artificial intelligence models for causal inference. I will further illustrate the shared statistical core across these three types of causal inference models and, from there, explore the application of statistical methodology in causal artificial intelligence models. The study cases include: 1) the benefits and the risk of immune-related adverse events of immune checkpoint inhibitors in cancer patients; and 2) the role of social determinants of health in the observed health disparities of the severe acidosis risk among diabetic chronic kidney disease patients.

WEDNESDAY MORNING JUNE 12

Computation and visualization

R-shiny application for visualizing cough count data collected from a wearable device
Richard Vinisko, Boehringer Ingelheim United States
Saurav Adhikari, Boehringer Ingelheim United States
Pharmaceutical companies are increasingly interested in exploring the benefits and feasibility of incorporating digital health technologies (DHTs), such as wearable devices, into their research initiatives. DHTs have been used to support objective, continuous, and real-time data collection and capture diverse data encompassing physiological, movement, and sleep metrics. Data collected from DHTs presents the chance provide a more complete

understanding of the multifaceted factors influencing a patient's health, which may change in response to a new treatment.

Opportunities exist to explore these data and provide insight into potential associations between different sets of metrics. We have developed an r-shiny application to visualize data collected by a wearable monitor designed to capture coughs and other attributes over a 24 hour recording period. The device concurrently collects physiological data such as heart rate, physical activity (low, medium, high), and body position (upright or not). Visualization of these 24 hour profiles can aid in discovering patterns between cough frequency or other cough attributes and these physiological measures which might lead to hypothesis generation or aid in developing novel endpoints for use in drug development.

Going from SAS v9 Programming to SAS Viya Programming: What's the Same, What's Different
John LaBore, SAS Institute
Josh Horstman, SAS Institute
Current SAS users are likely aware of SAS Viya, the next-generation software that will replace SAS v9 (there will not be a "SAS 10"). As with any change, it can be uncomfortable until it is understood what is the same and what is different. The good news is that v9 users will be able to lift all their code to SAS Viya, with only changes to pointers (i.e., libnames), as SAS is essentially embedding the v9 engine (the "Compute Server") in Viya. The better news is that users will be able to start writing code (Cloud Analytics Services Language, or CASL) that takes advantage of Viya's in-memory and multi-threaded super fast processing capabilities (the "CAS Server"). SAS Viya is a platform that can also easily integrate other languages (R, Python, etc.) with SAS into a seamless program. Even better, SAS Viya can be accessed with the familiar SAS interface, Enterprise Guide, or users can opt to use SAS Studio as the interface. This presentation will provide an overview of SAS Viya, and use a series of statistical programming examples to demonstrate how statisticians and statistical programmers can become rapidly productive in this new environment.

Why Do I Have Missing Data and How Do I Fix it?
Melodie Rush, SAS Institute
What do you do when you have missing values in your data? In SAS we have many ways to manage missing values. In this session we cover what are missing values, why and when missing values occur and how to manage missing values. We discuss functions, procedures and how different products deal with missing values.

Leveraging Visual Analytics in Drug Safety and Benefit-Risk in Multi-Regional Clinical Trials
Melvin Munsaka, AbbVie
Multi-regional clinical trials (MRCTs) are routinely used in drug development. There are many analysis challenges in MRCTs due to both intrinsic factors and extrinsic factors as noted in the International Council for Harmonisation E5 (ICH-E5) guideline. The ICH-E17 guideline provides general principles for planning and design of MRCTs. This guideline outlines high-level general principles for the planning and design of MRCTs with the aim of increasing their acceptability in global regulatory submissions. It also includes some anecdotal mention of drug safety and benefit-risk. In the context of drug safety and benefit-risk, there is a need to better understand, explore, and document differences and their implications as appropriate and to explore biological or practical reasons for apparent differences in drug safety outcomes and benefit-risk when noted in MCRTs. This may be challenging for safety outcomes in single MRCT given that most trials are designed for efficacy. One approach is to leverage pooled safety data across two or more MRCTs as appropriate. In this talk, we will discuss some quantitative and visual analytics approaches in MRCTs for drug safety and benefit-risk assessment with emphasis on the use visual analytics as tool in exploring and assessing potential heterogeneity across regions in safety outcomes and benefit-risk.