# Abstracts

## STATBOLIC

*February 23-24, 2026*

**Hilton Washington DC/Rockville Hotel & Executive Meeting Center**

*1750 Rockville Pike, Rockville, MD 20852*

## Session I: Innovative Trial Designs: Exploring Multiple Indications Within a Single Study

**Time: 10:15 – 11:45 AM, February 23, 2026**

### *P1: A Novel Design to Evaluate a Broad Indication Related to Metabolic Disorders in One Outcome Study*

***Presenter:*** *Yongming Qu, Vice President, Statistics, Eli Lilly*

***Abstract:***

We propose a unified clinical trial design that leverages the cardiovascular-kidney-metabolic (CKM) syndrome framework to simultaneously evaluate treatment efficacy across multiple metabolic disease indications using a composite primary endpoint combining major adverse liver outcomes, renal composite endpoints, heart failure, and major adverse cardiovascular events. The proposal is to evaluate the primary outcome of a broad composite of endpoints in multiple metabolic associated diseases with sufficient evidence showing the trend of improvement in each individual disease endpoint. We discuss different options for the proposed approach with a spectrum of sample sizes to provide different levels of strength of evidence. Implementation challenges are also discussed.

### *P2: Approaches to Move from a Narrow Indication Specific Composite Endpoint to a More Holistic Assessment of Efficacy in Cardio Renal Metabolic Trials*

***Presenter:*** *Stefan Hantel, Expert Statistician, Boehringer Ingelheim*

***Abstract:***

Type 2 diabetes mellitus and obesity are associated with cardiovascular comorbidities like heart failures or chronic kidney diseases. The endpoints major adverse cardiovascular events (MACE) cardiovascular death, myocardial infarction and stroke, heart failure related endpoints like hospitalization for heart failure and kidney related endpoints are usually investigated in separate registration trials for new drugs treating T2DM or obesity. This results in several large outcome trials. Since the enrolled patient populations in these trials is largely overlapping, a broader endpoint combining MACE, hear failure and CKD related components is considered. It will be discussed, how consistency of the components and consistency of results across sub-populations can be addressed.

***P3: TRIUMPH-OUTCOMES***

***Presenter:*** *Cem Kayhan, Associate Vice President, Development, Eli Lilly*

***Abstract:***

The TRIUMPH-Outcomes study is a large, event-driven, phase 3, randomized, double-blind, placebo-controlled clinical trial designed to evaluate whether retatrutide at its maximum tolerated dose reduces the risk of major cardiovascular (CV) and kidney outcomes in adults with overweight/obesity (BMI ≥27 kg/$m^2$) and established atherosclerotic cardiovascular disease (ASCVD) and/or chronic kidney disease (CKD). The trial aims to address two central clinical questions: the magnitude of treatment benefit on CV outcomes and on kidney outcomes compared with standard of care plus placebo. TRIUMPH-Outcomes plans to enroll approximately 10,000 participants, including a targeted 40% with CKD, with an expected mean follow-up of 3.3 years until accrual of at least 1296 CV events and 1085 kidney events. Key inclusion criteria encompass adults aged ≥45 years with or without type 2 diabetes (HbA1c ≤10.0%) and documented ASCVD or CKD based on eGFR and albuminuria thresholds. The study evaluates multiple primary and key secondary endpoints, including MACE-3, heart failure–related outcomes, kidney composite endpoints, all-cause mortality, albuminuria change, metabolic biomarkers, and body-weight outcomes. As obesity, diabetes, cardiovascular disease, and kidney disease continue to rise globally and contribute substantially to morbidity and healthcare burden, outcome trials such as TRIUMPH-Outcomes are essential to determine whether multitarget incretin-based therapies can meaningfully modify long-term risk and reshape standards of care.

***P4: Causal Mediation Analysis in Metabolic Disease***

***Presenter:*** *Jesper Madsen, Director, Statistics, Novo Nordisk*

***Abstract:***

When studies evaluate multiple effects or indications, investigators often want to understand mode-of-action (MoA) relationships — for example, whether effects on different endpoints arise from shared pathways. Regulatory agencies increasingly request exploration of hypothesized MoAs and/or the validity of surrogate markers. Causal mediation analysis can address these questions but depends on strong assumptions and there is a lack of consensus on which methods are most appropriate. We propose initiating a discussion about best practices for causal mediation analysis in metabolic disease and

advocate for approaches that target path-specific effects (see Vansteelandt et al., 2019; Linder et al., 2025).

S. Vansteelandt, M. Linder, S. Vandenberghe, J. Steen, and J. Madsen, “Mediation Analysis of Time-To-Event Endpoints Accounting for Repeatedly Measured Mediators Subject to Time-Varying Confounding”, Statistics in Medicine 38, no. 24 (2019): 4828–4840.

M. Linder, J. Madsen, and S. Vansteelandt, “Mediation Analysis of Path-Specific Effects in Randomised Clinical Trials With Repeatedly Measured Mediators and Outcomes”, Pharm Stat. 2025; 24:e7003038.

***P5: Going Beyond One Trial – Utilization of Data Across Development Programs***

***Presenter:*** *Sille Esbjerg, Director, Statistics, Novo Nordisk Lene Sommer Vestergaard, Associate Director, Statistics, Novo Nordisk*

***Abstract:***

In the FDA draft guidance "Demonstrating Substantial Evidence of Effectiveness With One Adequate and Well-Controlled Clinical Investigation and Confirmatory Evidence," the agency sets out its requirements for what constitutes substantial evidence. The presentation will include two examples illustrating possible forms of confirmatory evidence. Our aim is to stimulate discussion on how to work with the concept of substantial evidence in situations where two well conducted clinical trials are not available.

## Session II: Active Comparator in Obesity Trials

**Time: 1:00 – 2:15 PM, February 23, 2026**

***P1: Applying the Estimand Framework to Non-Inferiority Trials***

***Presenter:*** *Bharani Dharan, Executive Director, Statistics, Novartis*

***Abstract:***

This presentation will explore the challenges of applying the estimand framework to non-inferiority trials, using clinical studies on weight-loss drugs as illustrative examples.

Regulatory guidelines for non-inferiority trials predate ICH E9(R1) and do not discuss the underlying treatment effect of interest. CHMP guidelines discuss the role of two analysis sets, the full analysis set and the per-protocol analysis set, and specify that the two analysis sets are equally important since the full analysis set is seen as generally not conservative for non-inferiority. The first part of the presentation will address how the estimands framework can be applied to non-inferiority trials.

The second part will discuss derivation and choice of non-inferiority margin. The FDA guidance on non-inferiority trials outlines approaches for determining the non-inferiority margin. As part of this approach, the effect of the reference treatment compared with placebo should be estimated based on historical trials (e.g. via a meta-analysis). This effect is named $M_1$ and the final non-inferiority margin, named $M_2$, cannot be larger than $M_1$. We first outline how the $M_1$ effect of the reference treatment versus placebo depends on the targeted estimand, specifically on the strategies for addressing intercurrent events. We conclude that theoretically, any meta-analysis performed to determine $M_1$ needs to be based on the same estimand that is targeted in the new non-inferiority trial. We then illustrate the practical implications, particularly in situations where the historical trials do not clearly specify which estimand was of interest or actually estimated.

References

Lynggaard, H., Keene, O. N., Mütze, T., & Rehal, S. (2024). Applying the Estimand Framework to Non-Inferiority Trials. *Pharmaceutical Statistics*, *23*(6), 1156-1165.

***P2: Active Comparator Studies for Weight Management With a Case Study***

***Presenter:*** *Anna Batorsky, Advisor, Statistics, Eli Lilly*

***Abstract:***

Current FDA regulations require 1,500 placebo participants for registration of weight management (WM) drugs. While placebo-controlled trials demonstrate both efficacy and safety, the primary purpose of this large placebo requirement is to ensure comprehensive safety assessment. However, recruiting and retaining large placebo cohorts is becoming increasingly challenging—and potentially unethical—as safe and effective WM therapies become more widely available. We propose that active comparator studies combined with indirect comparisons can achieve objectives similar to placebo-controlled trials. Although active comparator designs are common in other therapeutic areas, they have not been widely adopted for WM drug development.

We conducted a proof-of-concept analysis to evaluate whether indirect comparison estimates could reduce the need for large placebo cohorts. Specifically, we compared

tirzepatide versus placebo using both direct data from SURMOUNT-1 and indirect estimates derived from SURMOUNT-5 (tirzepatide vs semaglutide) and STEP-1 (semaglutide vs placebo). Efficacy parameters included percent change in body weight (treatment regimen and efficacy estimands), systolic and diastolic blood pressure, LDL cholesterol, and triglycerides. Safety assessments encompassed adverse events by system organ class and preferred term, pancreatic enzyme levels, and pulse.

Direct and indirect estimates using published summary statistics showed good concordance across both efficacy and safety parameters. These findings suggest that augmenting direct comparison data with indirect estimates may support reducing the 1,500 placebo participant requirement for WM drug registration. Prospectively designed active comparator studies using participant-level data and harmonized preferred terms for safety analyses could further improve estimate accuracy. This approach warrants discussion regarding its strengths, limitations, and potential regulatory implementation, with future analyses incorporating additional active comparator studies to validate this methodology.

## Session III: Treat-to-Target in Obesity

**Time: 2:15 – 3:15 PM, February 23, 2026**

***P1: Treat to Target in Obesity Management: The Why and How?***

***Presenter:*** *Volker Schnecke, Senior Real-World Evidence Manager, Novo Nordisk; Abd Tahrani, Vice President, Development, Amgen*

***Abstract:***

Although treatment targets are routinely used for other chronic conditions, none currently exist for obesity management. A treatment target should indicate that disease management has resulted in a reduction in the risk of adverse health outcomes from the disease. Real-world data have suggested BMI 27 and WHtR of 0.53 after weight loss to be associated with acceptably low risk of developing obesity-related comorbidities. Clinical trial data are currently analysed to examine how reaching these targets affects CV risk factors. Initial results suggest that meeting the BMI and/or WHtR targets is a better predictor of "normalized" CV risk factors than percentage weight loss. We will describe how the targets were derived, present initial exploratory results from weight-loss trials, and outline next steps to validate, refine, and establish treatment targets in obesity care.

## Keynote: Statistical Challenges in Developing Products Treating Metabolic Diseases

## Session IV: Covariate Adjustment in Cardiometabolic Health (CMH) Trials

**Time: 10:15 – 11:40 AM, February 24, 2026**

***P1: Covariate-Adjusted Log-Rank Test: Guaranteed Efficiency Gain and Universal Applicability***

***Presenter:*** *Marlena Bannick, Biostatistician, Seattle Children's Research Institute*

***Abstract:***

*The log-rank test is a nonparametric statistical tool for comparing time-to-event outcomes between two groups. In a randomized trial, baseline covariates that are prognostic for the outcome can be leveraged to improve efficiency. The covariate-adjusted log-rank test (CALRT) incorporates these variables and is guaranteed to be more efficient than the standard log-rank test, without making any modeling assumptions. Unlike the standard log-rank test, the CALRT enjoys universal applicability: the same CALRT statistic can be applied under simple randomization and all commonly used covariate-adaptive randomization schemes. This talk will provide intuition and a high-level overview of the CALRT methodology. We will also briefly illustrate its implementation in RobinCar2, a new R package developed by the ASA's Biopharmaceutical Section Covariate Adjustment Scientific Working Group.*

***P2: The effect of covariate adjustment for Cox regression in Cardiovascular Outcome (CVOT) trials***

***Presenter:*** *Ran Bi, Associate Director, Statistics, Novartis*

***Abstract:***

Adjustment for prognostic baseline covariates is commonly done to increase power in randomized controlled trials. Yet time-to-event outcomes are commonly analyzed using the log-rank test or Cox regression with no covariates besides treatment. Covariate adjustment in non-linear models is more complex than in linear models: common effect measures such as hazard ratios are non-collapsible, there is no obvious baseline assessment of the outcome to adjust for, existing risk scores may be insufficiently prognostic, and misspecification of the functional form of covariate effects can inflate type I error.

Nine studies in a pool of cardiovascular outcome trials were re-analyzed with adjustment for existing risk scores and known prognostic factors. We also assessed whether adjustment for these covariates would be outperformed by adjustment for predictions from newly created models (so-called 'super covariates'). We evaluated the strength of prognostic covariate required for adjustment to be worthwhile using simulations.

The effect of adjusting for existing risk scores and known prognostic factors was minimal. Adjusting for newly developed model-based predictions on average improved test statistics, but this was not consistent across studies or cross-validation folds. While the model-based predictions were stronger prognostic factors, the variability of the gain in test statistic also increased. The simulations suggest that stronger prognostic covariates may be required.

For time-to-event cardiovascular outcome studies, we should not automatically assume a meaningful benefit from covariate adjustment. A balanced view that considers each study individually is required.

## Session V: Artificial Intelligence (AI), Digital Health, and Real-World Evidence (RWE) in Cardiometabolic Health

**Time: 1:00 – 2:30 PM, February 24, 2026**

***P1: X-Y-Z framework for Modernizing Drug Development with AI***

***Presenter:*** *Yong Chen, Professor of Biostatistics at Department of Biostatistics, Epidemiology, and Informatics (DBEI), University of Pennsylvania, Perelman School of Medicine*

***Abstract:***

We introduce XYZ, a unified framework that advances clinical evidence generation by jointly modeling the treatment (X), outcome (Y), and population (Z) dimensions. This framework is built to support AI-driven drug repurposing, outcome profiling, and multi-dimensional optimization of eligibility criteria using real-world data. By integrating lossless federated target trial emulation for drug discovery, negative control–based debiasing for robust outcome inference, and AI-guided simulation to evaluate alternative inclusion criteria across real-world populations. XYZ enables principled and scalable evidence generation. Applications in repurposing GLP-1 RA, drug identification for AD/ADRD, and trial design for advanced non-small cell lung cancer (NSCLC) illustrate how XYZ enhances

generalizability, supports regulatory alignment, and ensures reliable insights generated from distributed research networks.

***P2: AI in cardiovascular disease***

***Presenter:*** *Hongtu Zhu, Kenan Distinguished Professor of Biostatistics, Statistics, Radiology, Computer Science and Genetics, University of North Carolina at Chapel Hill*

***Abstract:***

This talk surveys recent AI advances in cardiovascular disease (CVD) analysis, spanning both image-based diagnostics and EHR-based modeling. We organize the imaging literature by key anatomical targets—non-vessel structures such as the ventricles and atria, and vessel structures such as the aorta and coronary arteries—providing a framework to compare methods across modalities including MRI and other cardiac scans. We also review AI applications to EHR data for risk prediction, phenotyping, and outcome forecasting, emphasizing opportunities and pitfalls in real-world clinical data. The talk highlights publicly available datasets and code resources to support reproducibility, and concludes with key challenges and promising directions for future research.

***P3: Beyond Fixed Thresholds: Optimizing Summaries of Wearable Device Data***

***Presenter:*** *Irina Gaynanova, Associate Professor, Biostatistics, School of Public Health, University of Michigan*

***Abstract:***

Wearable devices, such as actigraphy monitors and continuous glucose monitors (CGMs), capture high-frequency data, which are often summarized by the percentages of time spent within fixed thresholds. For example, actigraphy data are categorized into sedentary, light, and moderate-to-vigorous activity, while CGM data are divided into hypoglycemia, normoglycemia, and hyperglycemia based on a standard glucose range of 70–180 mg/dL. Although scientific and clinical guidelines inform the choice of thresholds, it remains unclear whether this choice is optimal and whether the same thresholds should be applied across different populations. In this work, we define threshold optimality with loss functions that quantify discrepancies between the full empirical distributions of wearable device measurements and their discretization based on specific thresholds. We introduce two loss functions: one that aims to accurately reconstruct the original distributions and another that preserves the pairwise sample distances. Using the Wasserstein distance as the base measure, we reformulate the loss minimization as optimal piecewise linearization of quantile functions. We solve this optimization via stepwise algorithms and differential

evolution. We also formulate semi-supervised approaches where some thresholds are predefined based on scientific rationale. Applications to CGM data from two distinct populations—individuals with type 1 diabetes and those with normal glycemic control—demonstrate that data-driven thresholds vary by population and improve discriminative power over fixed thresholds. This is joint work with Junyoung Park and Neo Kok.

***P4: AI to Automate Clinical Event Adjudication in Global Randomized Trials***

***Presenter:*** *Pablo Marti Castellote, Research Fellow in Medicine, Brigham and Women's Hospital*

***Abstract:***

N.A.

***P5: Impact of GLP-1 Discontinuation on HbA1c and Weight in Real-World T2D Patients***

***Presenter:*** *Christophe Tchakoute, Senior Data Scientist, Product Development Data Sciences, Genentech*

***Abstract:***

GLP-1 RAs are cornerstones in Type 2 Diabetes (T2D) management, significantly improving glycemic control (HbA1c) and causing substantial weight loss, as shown in RCTs. However, the real-world impact of treatment discontinuation on these key metrics is not well understood, especially as it relates to wash-out periods for subsequent clinical trials. Understanding the rate and extent of HbA1c and weight rebound is critical for clinical practice and for setting appropriate wash-out periods in future clinical trials.

Study Design and Methods

This retrospective cohort analysis used the IQVIA-AEMR claims database, linking administrative claims with detailed Electronic Health Record (EHR) data (HbA1c, weight) to quantify changes following GLP-1 RA discontinuation in a large US T2D cohort.

To address selection bias inherent in observational studies, rigorous statistical methods (Inverse Probability of Treatment Weighting) were employed. Multiple imputation was used to address missingness in some of the key variables used to derive weights. This

methodological approach minimized confounding and enhances the generalizability of the findings to the broader T2D population.

Conclusion and Clinical Significance

These findings will help provide more insights on the consequences of GLP-1 RA cessation, which is critical for clinical practice and the development of future GLP-1 RA therapies.