THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

MASH Placebo Database Working Group Update

Liver Forum 17 - Paris

4 September 2024

Veronica Miller, *PhD*

Chris Hoffman, *PhD*

Margot Yann, *PhD*

The Forum for Collaborative Research

Berkeley Public Health

Welcome

Veronica Miller

Michael Cooreman

Agenda

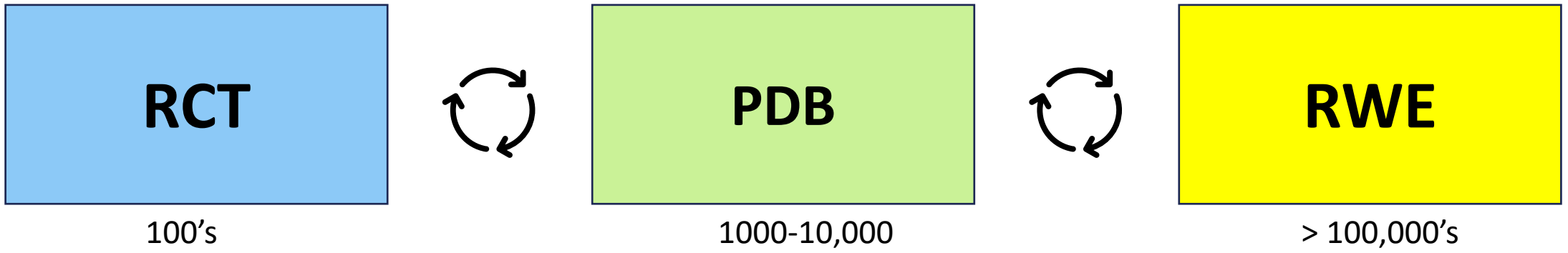
THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Data & Analysis Center (DAC) Team Updates
- Data update
- Research use case
- Data merging and initial analysis
- Emerging vision for the DAC
- Discussion

Placebo Subjects: A unique NH cohort and bridge between RCT and RWE

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Patients selected/experienced RCT - bias reduced
- Patients more diverse than in any individual trial
- Time zero defined
- Patient-level data available to regulatory agencies to support regulatory processes



MASH PDB – Potential Research Questions

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Natural history of MASH in untreated trial patients
- Comparability of RCT patients to "real world" patients
- Predictors of disease improvement, stability, worsening
- Fluctuation in safety parameters in untreated patients
- Analysis and prediction of screen failures
- Application of AI/ML to paired biopsies
- Comparison of causal inference and other analytical methods
- Shared placebo arm for future trials
- Others?

Maximizing Data Potential

Data-Driven Questions & Solutions

MASH Placebo Database Project

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

Community-based project oversight ("governance") based on nearly 25 years of experience building trust with our stakeholders to make collective progress on global health challenges

Co-Chairs

MASH PDB Executive Comm

MASH PDB Steering Comm

Data contributors + experts across stakeholder groups

Statistics & Analysis WG

Open to academic, industry, and regulatory statisticians and folks not afraid of cohorting with statisticians

MASH PDB Working Group

Open to all LF members

The DAC Platform

A safe place for data sharing and analysis

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Data protection by design & by default
  - Approved at UC Berkeley for ePHI and highly sensitive data
- Deep collaborations across UC Berkeley Offices
  - Privacy, Human Subjects, Information Security, and VC Research Offices
- Virtual machines, HPC cluster, and parallel file system storage
- Access restricted to DAC Team
  - Access for regulatory agencies to support regulatory review

Liver Forum 16 – WG summary

A sense of urgency

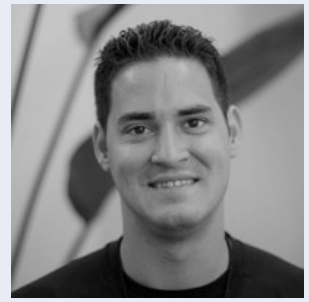
THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Ethical and practical challenges to continuing and new trials
- How will approval of first MASH treatment impact ongoing and future studies?
- Understanding the untreated (placebo) rates of fibrosis/ NAS progression and regression is critical
- Potential value for early phase studies, to improve patient selection (define homogenous populations), is high
- Potential to digitalize biopsy or radiology data, aligning with AI/ML Histology and Radiology WG's, is high
- Potential of connecting to EMR/ registries to obtain longer term data in registries is high (e.g., tokenize to identify patients and data from EMR to obtain longer term outcomes)

DAC Team Updates

The Data & Analysis Center Team

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

Principal Investigator/Director Veronica Miller, PhD		Lead Data Scientist Margot Yann, PhD	
IT & Operations Director Chris Hoffman, PhD		Technical/Infrastructure Zach Rooney, MSCS	
Forum/DAC Advisor Richard Haubrich, MD		**Introducing** Research Data Analyst Mengxi Bai, MSSTAT	

Progress since Liver Forum 16

Building deeper expertise with clinical trial data

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Hiring: Mengxi Bai - Data Analyst with pharma experience
- Team training: CDISC standards and SAS underway
- Building a network of service providers
  - E.g., anonymization, data standardization, tokenization
- Working with MASH PDB data received
- Talking with companies about data contributions
  - Concept note slides available for your use

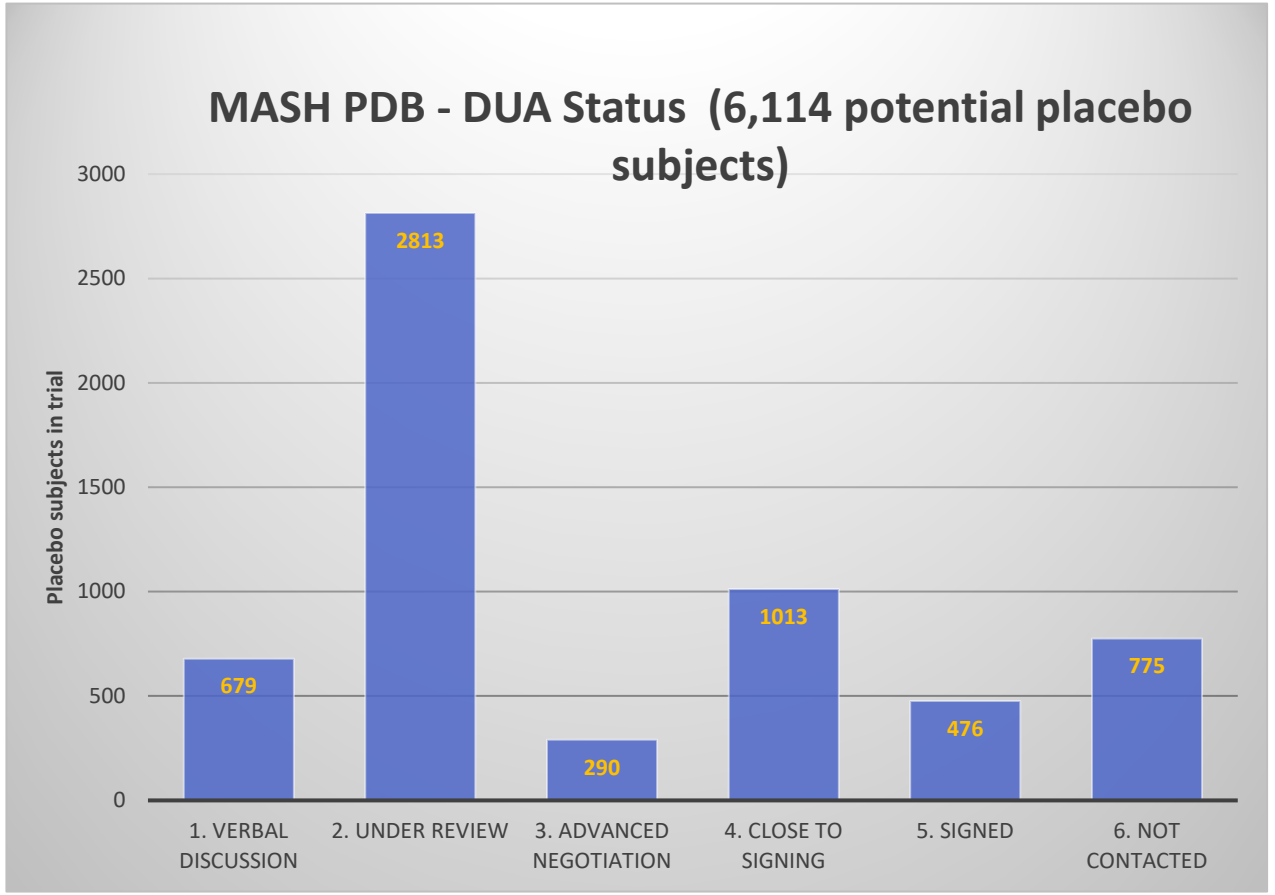
Data Update

Data Sources & Data Availability

As of 30-August-2024

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Invited to participate
  - All completed phase 2 and phase 3 studies
- Potential # of placebo patients: >6K
  - Signed / data received: 476 (2 companies plus five clinical trials from NIH NIDDK)
  - Close to signing – 1,013 (4 companies)
  - Advanced negotiations – 290 (1 company)
  - Under review / strong commitment: 2,813 (10 companies)
  - Verbal discussion – 679 (5 companies)
  - Declined to share data: 1



Our Ideal Request

Always subject to negotiation

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- De-identified versions of the SDTM and ADaM datasets for the placebo subjects who participated in the relevant trials
- Supporting documentation
  - Latest study protocol or plan
  - Dataset specification and/or data dictionary
  - Case report forms
  - De-identification/anonymization methods and other data transformations
  - Primary publication(s) (as attachments or citations)
  - If available: Statistical analysis plan, define.xml
  - Manifest of files received

Data Management Update

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Standardizing and updating documentation, e.g.,
  - Data Ingest Checklist
  - Metadata documentation
  - DAC Security Plan
- Assessing data quality (e.g., Pinnacle 21)
- Handling different versions of data
- Mapping data and developing schema
- Initial merging and exploratory data analysis (below)

MASH Placebo Data Research Use Case - Update

Research Use Case - Background

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Some data providers have requested a research plan
- First draft: PDB Working Group meeting at Liver Forum 16
- Revisions by Margot Yann and Richard Haubrich
- Discussed with Co-Chairs (Michael Cooreman and Manal Abdelmalek)

Primary Objective

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Define a set of parameters that accurately characterizes the population exhibiting stable liver fibrosis in MASH.
- Factors including
  - Demographic details (age, sex, BMI, race, ethnicity)
  - Routine laboratory tests (bilirubin, platelets, AST/ ALT, INR, HbA1c)
  - Non-invasive tests of fibrosis (FIB-4, NFS, ELF, VCTE)
  - Liver histology (fibrosis progression/regression, hepatic collagen content, α-SMA expression)
  - Other potential biomarkers

Primary Endpoint

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Categorical Endpoint:
  - Liver fibrosis by histology at the end of follow up
    - stable (no change in fibrosis category from baseline) or
    - changed (either increased or decreased category from baseline)
- Exploratory Endpoints: Change from baseline to end of follow up in VCTE and ELF
  - Explore changes in continuous biomarkers of fibrosis to provide insights into the more subtle variations and trends in fibrosis markers over time

Secondary Objectives

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Access the short-term fluctuations in biomarkers in MASH patients.
  - Identify the variability in these markers over shorter intervals (i.e. <3-4 months) and ascertain their significance in the context of disease stability and progression.
- Characterize the subset of the MASH patient population identified as not experiencing fibrosis progression and evaluate if there are differences by sex, race and age.

Secondary Endpoints

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Change in the intra-individual variability of biomarkers such as liver enzymes, fibrosis scores over predefined short intervals (e.g., monthly assessments over a six-month period)
  - Evaluate biomaker fluctuation's amplitude and frequency in MASH patients, correlating them with transient clinical events or interventions and assessing their predictive value for long-term outcomes.
- Absence of significant changes in both non-invasive fibrosis markers (such as FIB-4, NFS, ELF, VCTE) and liver histology assessments, from baseline to the conclusion of the study period.
  - Aim to identify unique traits and factors that contribute to fibrosis resistance in a specific MASH demographic.

Study Population & Data Analysis

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Study Population
  - Analysis of data from the placebo arms of clinical trials
- Data Analysis – Statistical Approach
  - Multivariable logistic regression to evaluate the bivariate outcome of change in fibrosis progression (change vs no change) to define factors associated with lack of progression
  - Descriptive analyses of the population with no change in fibrosis
  - Longitudinal data analysis utilizing repeated measures analysis of variance (ANOVA) or linear mixed-effects models to account for intra-individual variability over time to analyze the short-term fluctuations of biomarkers

Data Analysis - continued

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- Data Analysis – Machine Learning Approach
  - Supervised learning
    - Define outcome variables as labels to train algorithms on a labeled dataset
    - Regression models, random forests, and gradient booting etc.
  - Unsupervised Learning
    - Discover patterns or groupings in the data, identifying patient groups with similar profiles
    - Clustering algorithms such as K-Means and Hierarchical Clustering
  - Other forecasting analysis: time series analysis
    - Identify any cyclic or trend patterns
    - Methods like ARIMA or Prophet

Expected Outcomes

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- This analysis seeks immediate benefits by demonstrating PDB's utility and identifying the relationship between the key biomarkers and MASH fibrosis progression/regression.
- The findings will inform strategic patient selection in clinical trial design for MASH, promoting a more targeted and individualized approach for intervention.

We remain open to making amendments and incorporating additional analyses as necessary to ensure the ongoing relevance and efficacy of our study.

MASH Placebo Data – Data merging and initial analysis

Data of Interest

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

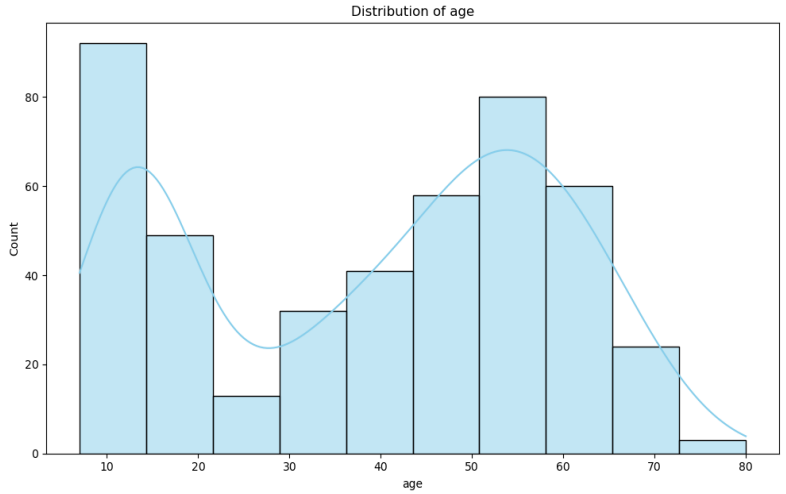
- Define a set of common parameters, including
  - Demographic details (age, sex, BMI, race/ethnicity)
  - Routine laboratory tests (bilirubin, AST/ ALT, HDL)
  - Noninvasive tests (NITs) of fibrosis and steatosis (CAP, LSM, MRI-PDFF)
  - Liver histology (fibrosis, steatosis, lobular inflammation, ballooning)
- Baseline & endpoint follow-up
  - Other regular follow-ups to be added

Patient Data Distribution

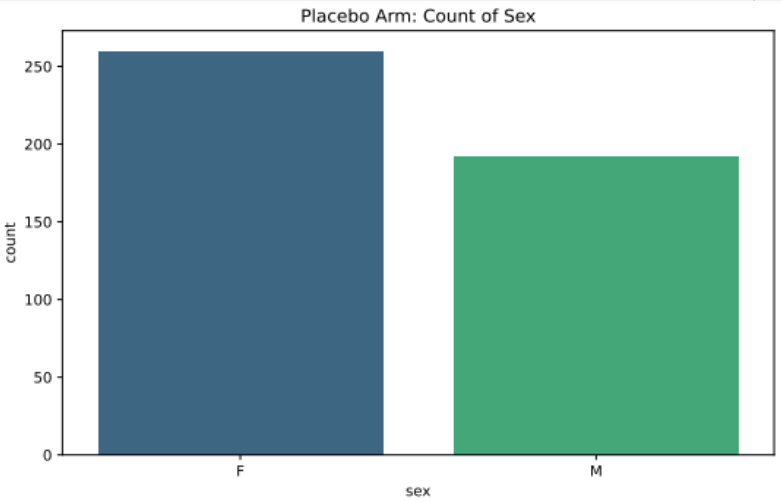
- *Demographics*

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

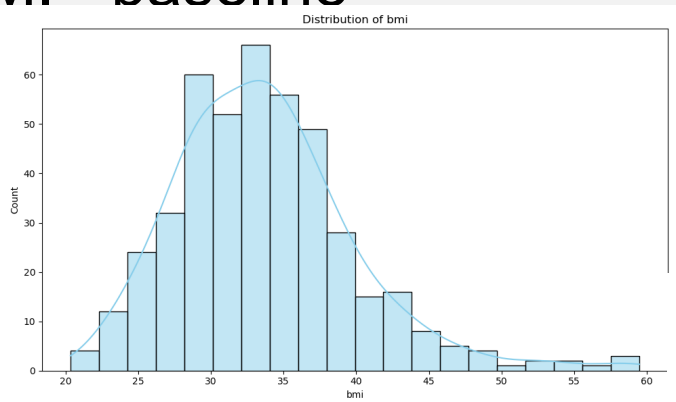
- Age



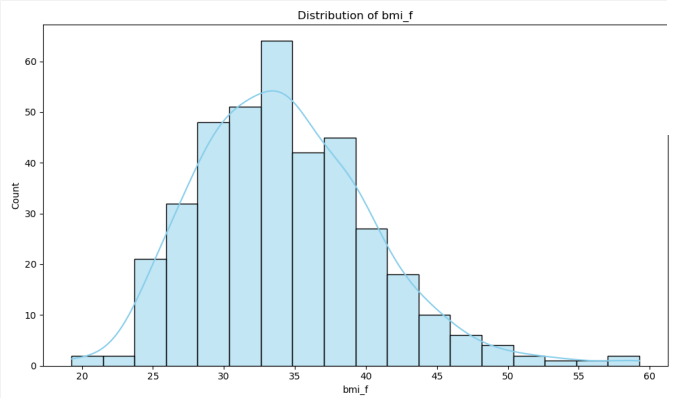
- Gender



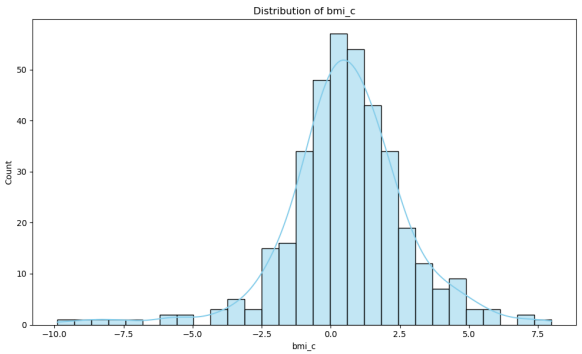
- BMI - baseline



- BMI - endpoint follow-up:



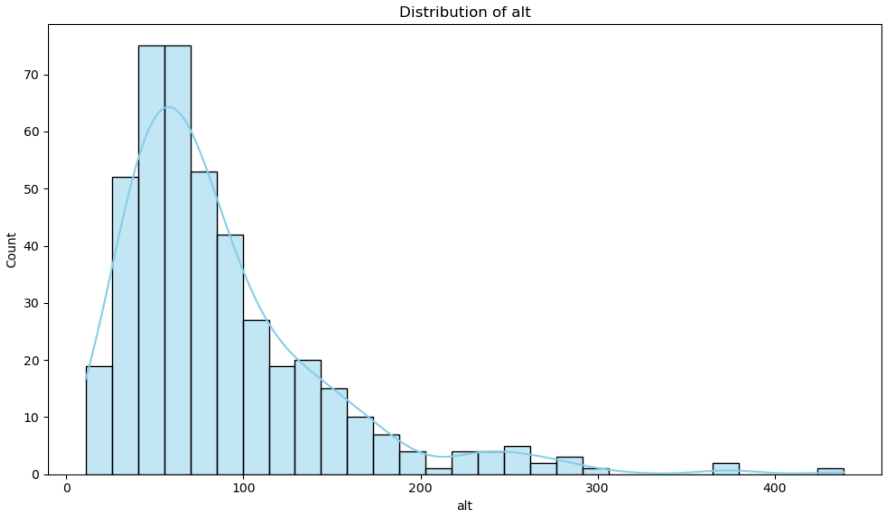
- BMI - changes:



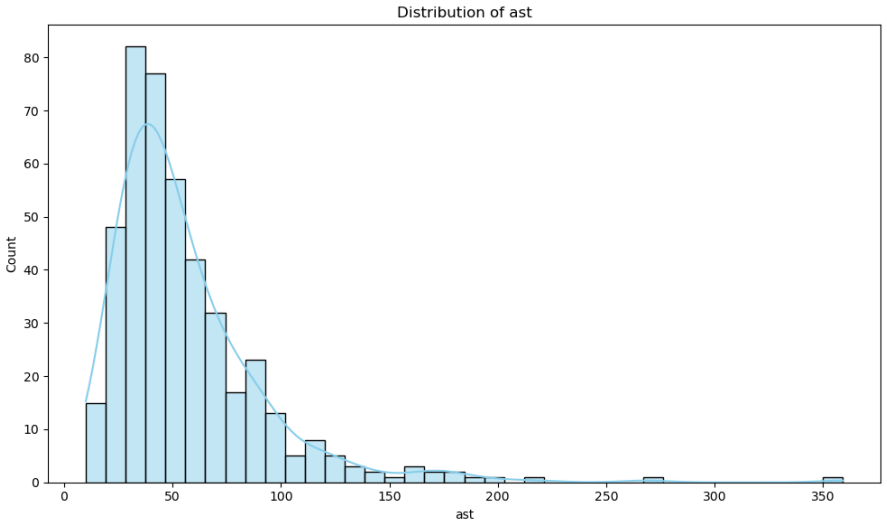
Routine Laboratory Biomarkers

- Baseline

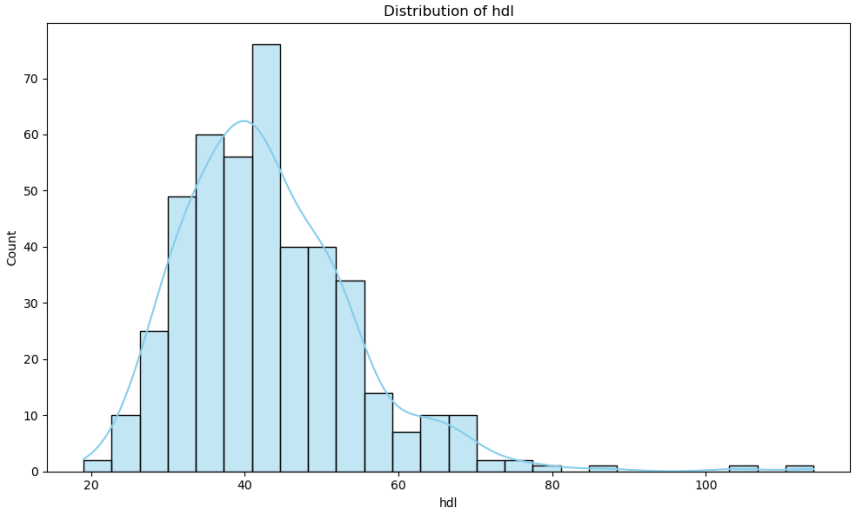
- ALT



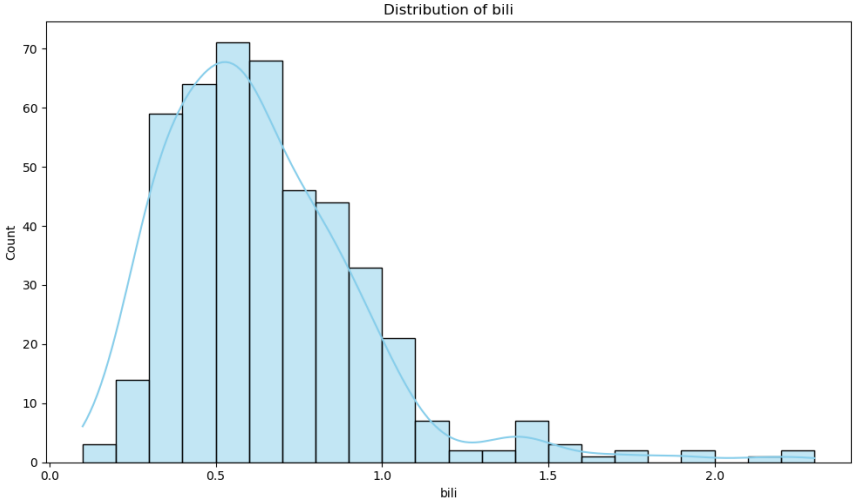
- AST



- HDL



- Bilirubin

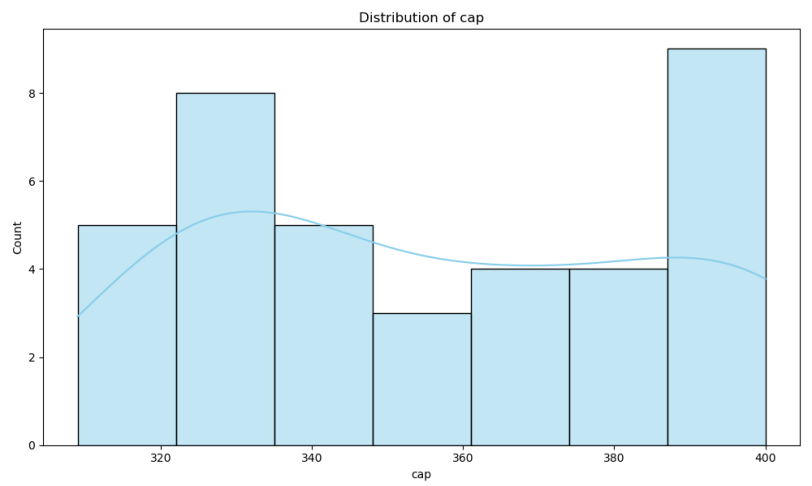


NITs

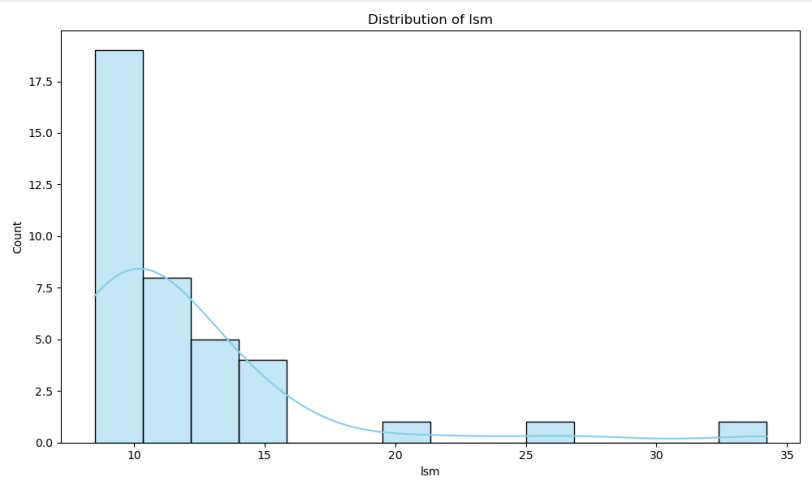
- Baseline

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

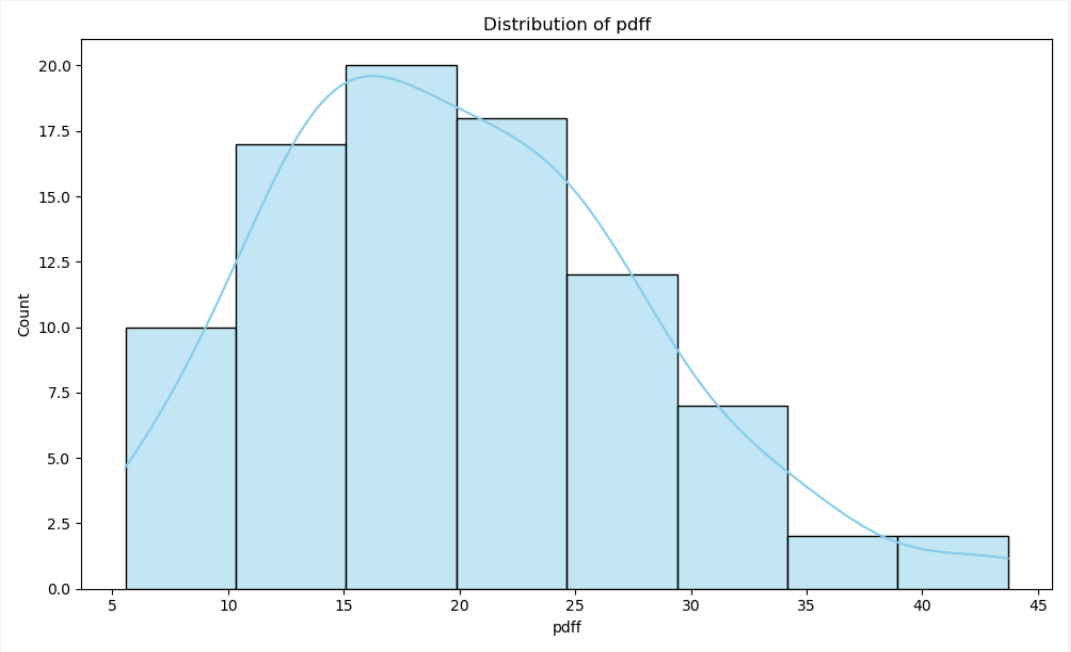
- CAP



- LSM



- MRI-PDFF

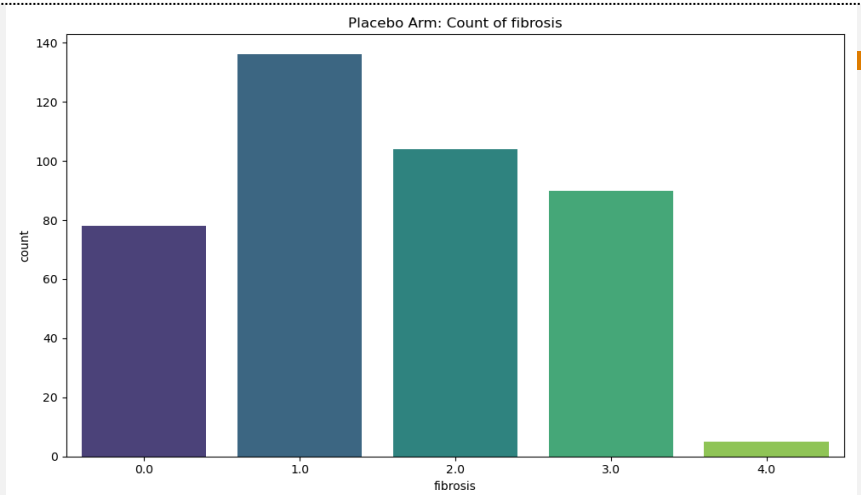


Liver Histology

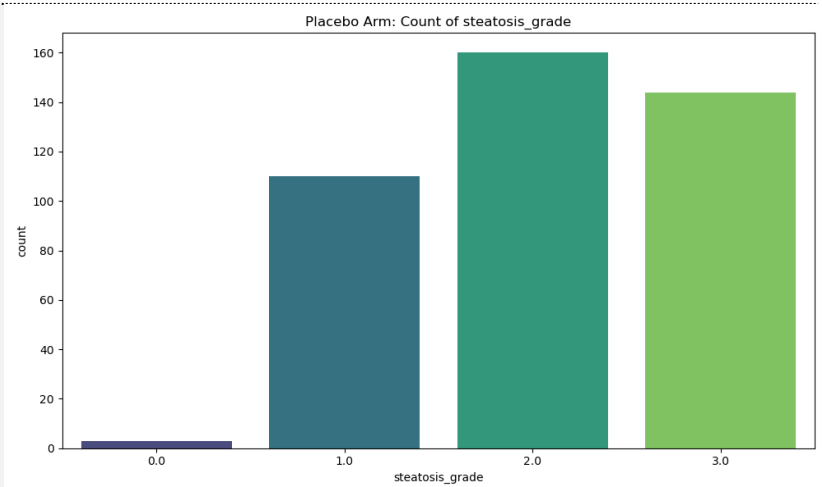
- Baseline

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

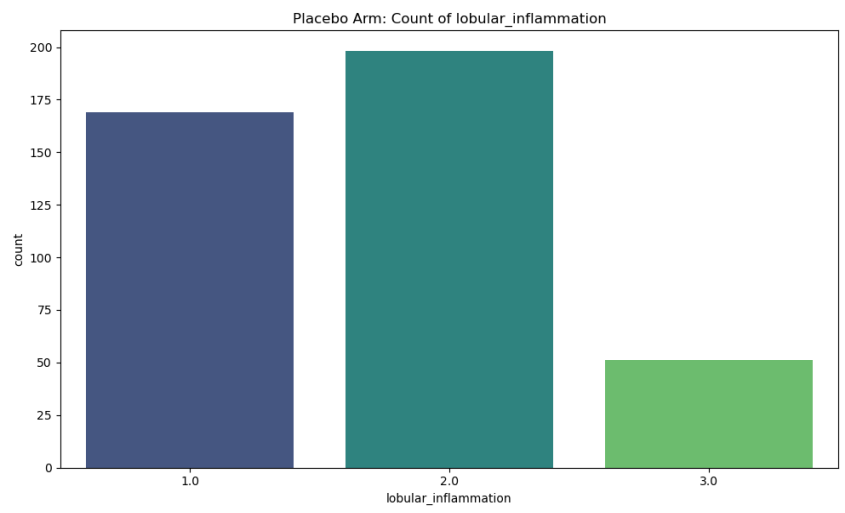
- Fibrosis



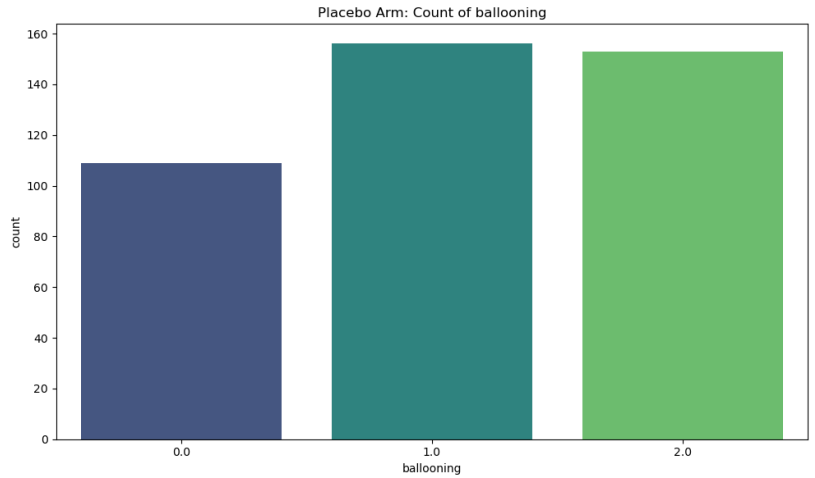
- Steatosis



- Lobular Inflammation



- Ballooning



Exploratory Analysis

t-SNE (t-distributed Stochastic Neighbor Embedding)

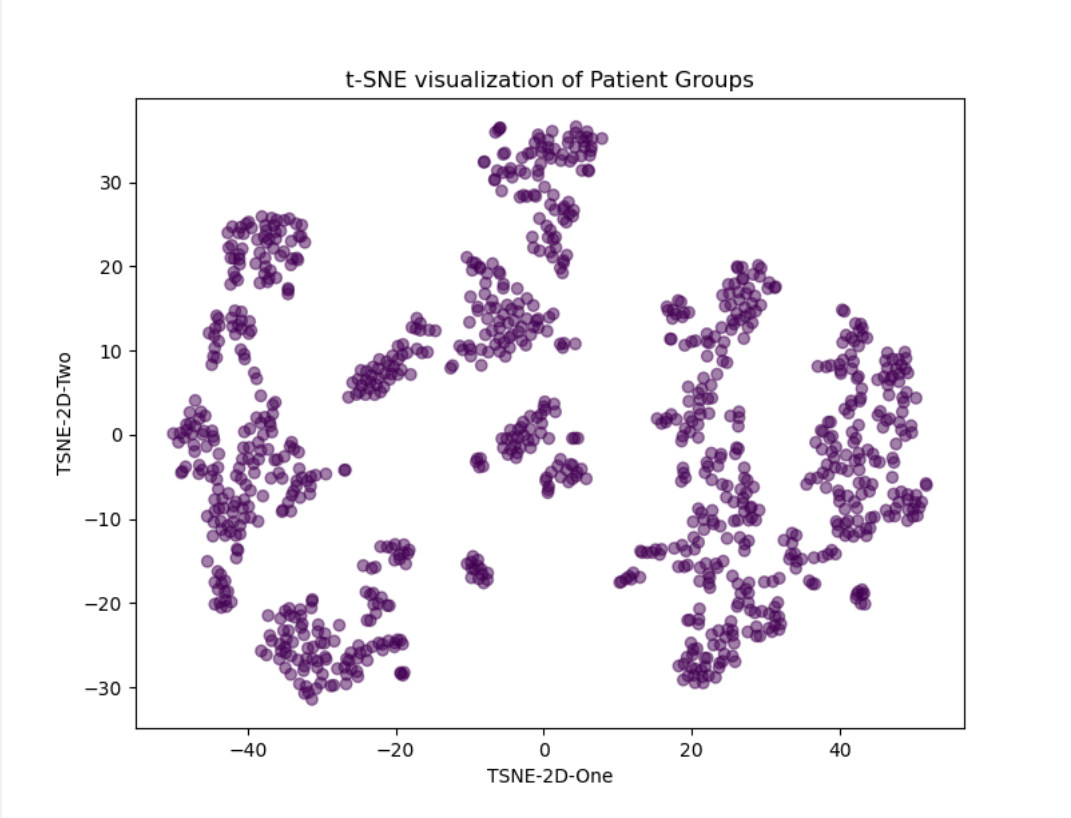
THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

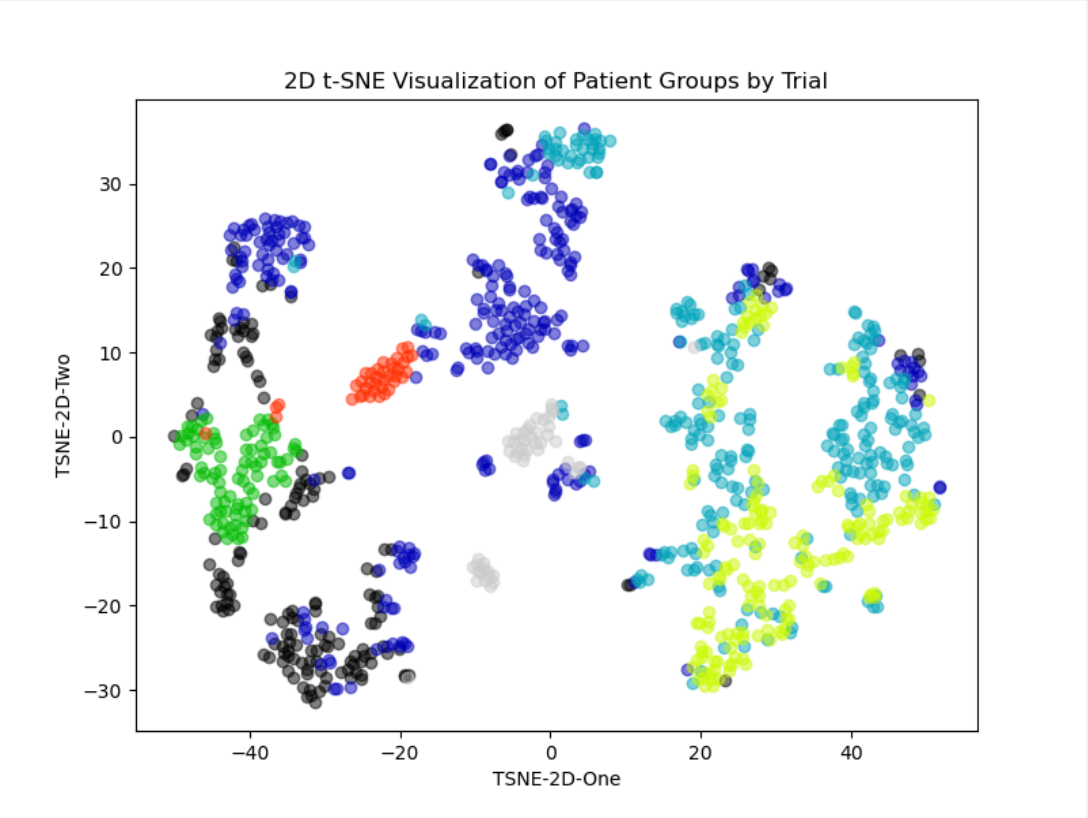
- t-SNE effectively reduces high dimensions non-linear relationships to 2D or 3D while preserving the local structure of the data, making it easier to visualize complex relationships.
- To identify groups of patients with similar conditions, responses to treatments, or other relevant medical characteristics.
- Data exploration:
  - understand the underlying structure of the data without making strong assumptions
  - discovering new patterns

Preliminary Analysis

THE FORUM
For Collaborative Research℠
Berkeley's Hub for Regulatory Science

- 2D





Emerging vision for the DAC

"A life science database for all hepatic indications"

Michael Cooreman

- Growing need & interest for data aggregation and analysis
  - Digital AI/ML Histology Working Group
    - Image analysis
  - NIT for RLSE Working Group
  - Radiology Working Group
  - Genetic and –omics data
  - PSC and PBC interest
  - Beyond placebo data: treatment data, baseline data …
  - Beyond clinical trial data: Real-world data and evidence (RWD/RWE)

Next generation data science at the intersection of regulatory science and public health

Discussion
&
Thank you