

**THE FORUM**

For Collaborative Research<sup>SM</sup>

*Berkeley's Hub for Regulatory Science*

# MASH Placebo-Arm Database Project Update

## - Data Harmonization & Proposed Analyses

Margot Yann, PhD, M.Eng.

Richard Haubrich, MD

Veronica Miller, PhD

September 10, 2025

# MASH PDB Project



**THE FORUM**  
For Collaborative Research<sup>SM</sup>

## MASH PDB Working Group - Open to all Liver Forum Members!

### Co-Chairs

- Manal Abdelmalek, Academic co-Chair
- Michael Cooreman, Industry co-Chair

### Executive Committees

- Quentin Anstee
- Bettina Hansen
- Veronica Miller
- Arun Sanyal
- Henry Chang

### Steering Committees

Data Contributors & Experts across stakeholder groups

**Thank you for the PDB Leadership!**

# The Data & Analysis Center (DAC) Team



**Veronica Miller, PhD**  
Director, Forum  
Principal Investigator



**Margot Yann, PhD, MEng**  
Director, DAC  
Principal Data Scientist



**Richard Haubrich, MD**  
Consultant



**Alice Kang, MPH, CPH**  
Sr. Project Manager



**Sunil Gupta**  
SAS, CDISC Consultant



**Zach Rooney, MSCS**  
IT System Analyst



**Debajyoti Debnath, MS**  
Research Associate

## Recruiting Now:

- Research Data Analyst
- Data Engineer

# The DAC Platform

A secure place for data sharing and analysis



- Data protection by design and by default
  - Approved at UC Berkeley for ePHI and highly sensitive data
  - Deep collaborations across UC Berkeley Offices - Privacy, Human Subjects, Information Security, VC Research Offices, UC Berkeley's Office of Intellectual Property and Industry Research Alliances (IPIRA)
  - Access restricted to DAC Team: access for regulatory agencies to support regulatory review
- Virtual machines, HPC cluster, and parallel file system storage
- Planned: Cloud-based Computing Environment for PDB analysis
- Bespoke Data Use Agreement (DUA) for LF PDB developed by UCB
  - Contact info: Alice Kang [akang@forumresearch.org](mailto:akang@forumresearch.org)

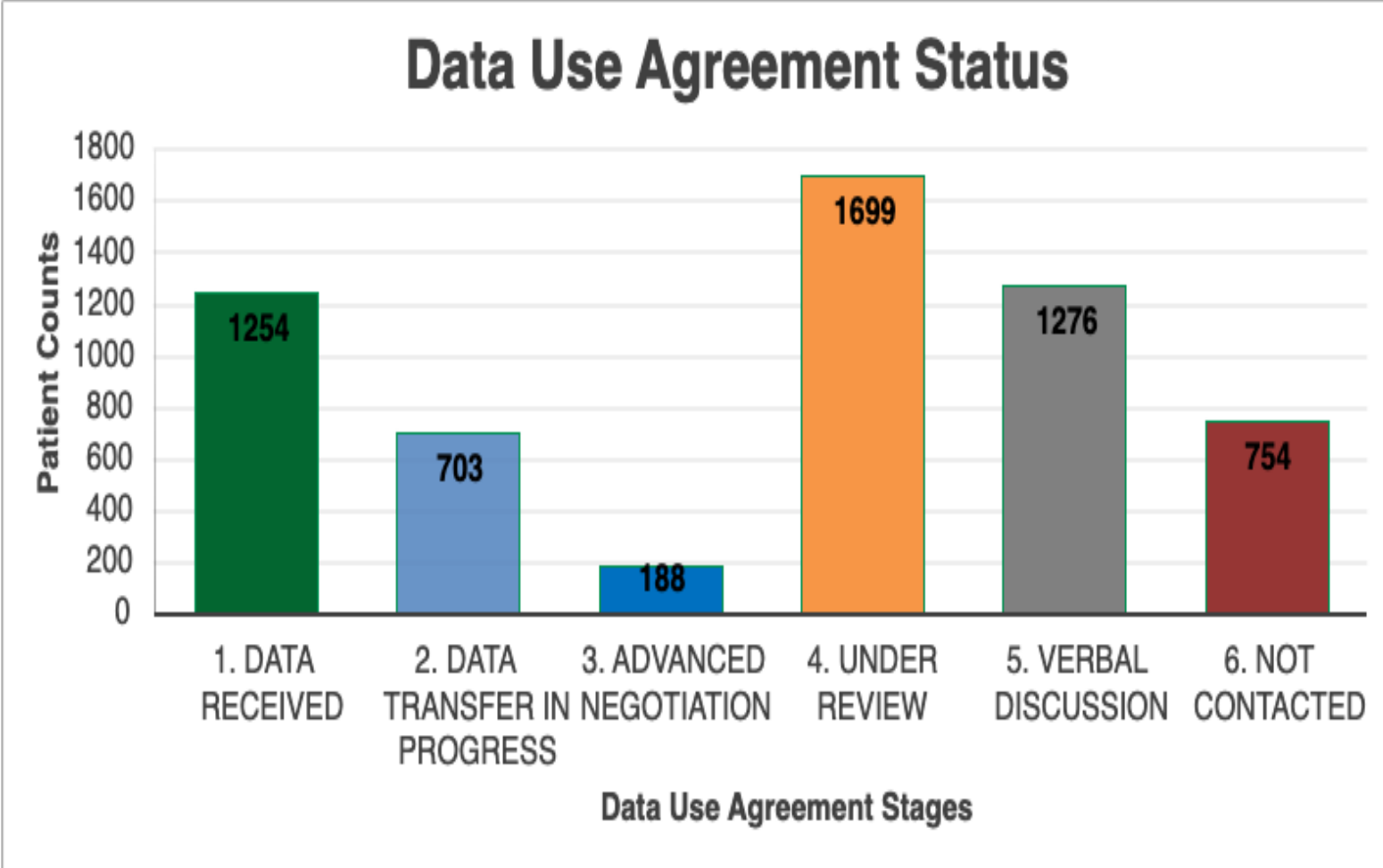
# Data Sources & Data Availability

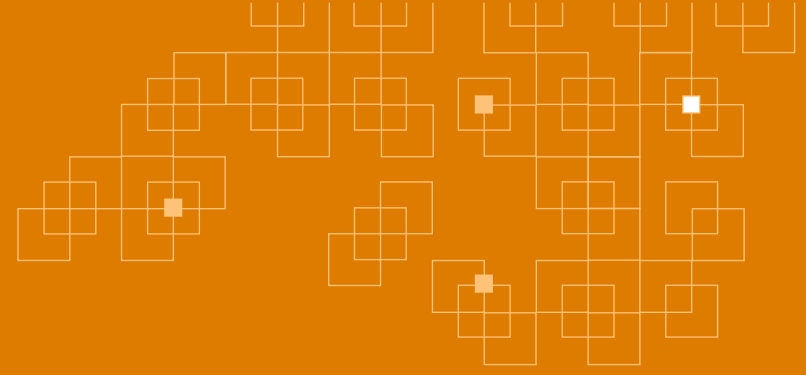
As of September 9, 2025



**THE FORUM**  
For Collaborative Research<sup>SM</sup>  
Berkeley's Hub for Regulatory Science

- Invited to participate
  - All completed phase 2 and phase 3 studies
- Potential # of patients: 5874
  1. Data Received: 1254  
4 companies & NIDDK
  2. Signed, data transfer in progress: 703  
2 companies
  3. Advanced negotiations: 188
  4. Under review: 1699
  5. Verbal discussion: 1276
  6. Not contacted: 754





# **MASH PDB: Data Harmonization Update**

# Harmonization Project Overview



- The **PDB harmonization project** involves integrating data from multiple studies into a standardized structure while ensuring **traceability, consistency, and regulatory compliance.**
- Challenges:
  - Harmonize all clinical trials in a consistent manner for clinical research
    - PDB project expects ~100 trials
  - FDA began requiring the use of CDISC (Clinical Data Interchange Standards Consortium) standards for regulatory submissions involving study data for studies initiated after December 17, 2016.
  - CDISC has developed and released more than 50 standards for clinical research, including data collection, analysis, and submission, as well as therapeutic areas such as oncology, infectious diseases, cardiovascular disease, and more – **not for MASH.**

# CDISC & FDA Guidelines



## Overview topics:

- **CDISC and MASH:** Explore how CDISC standards are currently applied in MASH clinical trials.
- **FDA Nonbinding Recommendations for MASH**
  - *“Technical Specifications for Submitting Clinical Trial Data Sets for Treatment of Noncirrhotic Nonalcoholic Steatohepatitis (NASH) - Guidance for Industry”, FDA CDER, January 2022.*
- **Engaging with CDISC, FDA, and Pharma Companies:** Discuss potential strategies for this collaboration in MASH and MASH Standard Development & how it could benefit MASH research and treatment

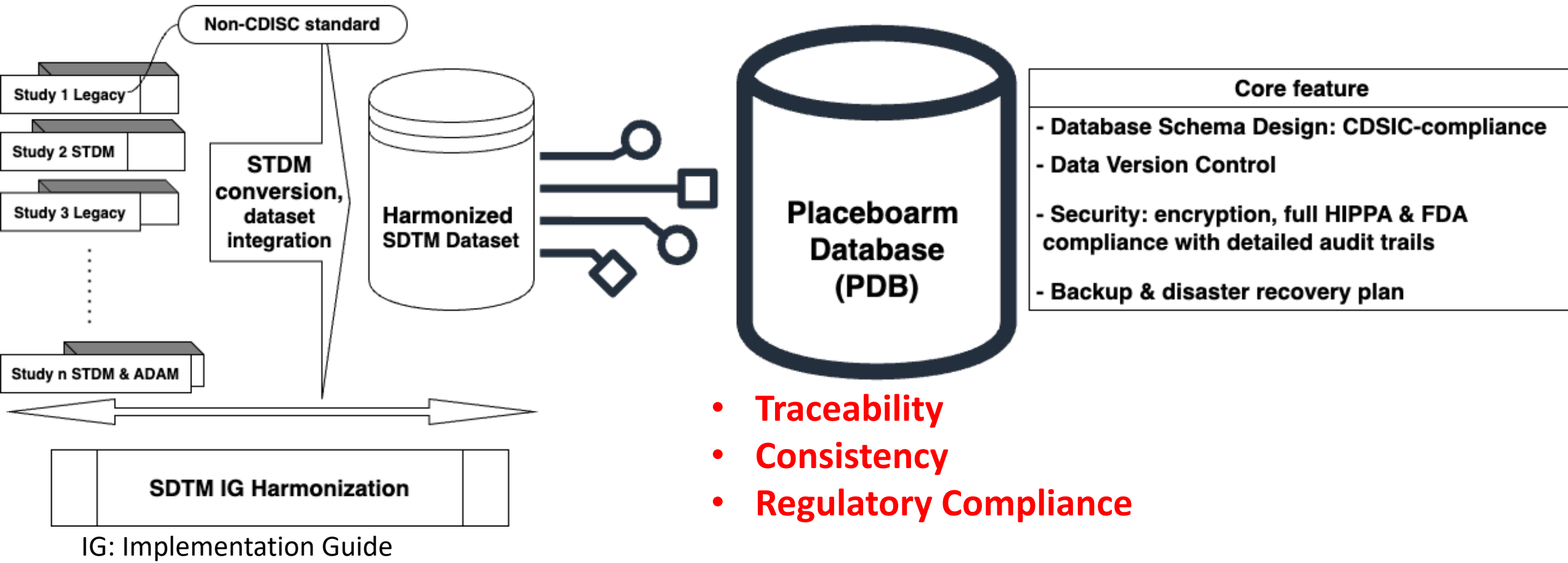
# Harmonization Flowchart



**THE FORUM**

For Collaborative Research<sup>SM</sup>

Berkeley's Hub for Regulatory Science



# Key Points



- **Upversioning of Dictionaries & Terminologies**
  - SDTM IG, CDISC Controlled Terminology, MedDRA
  - Assess the need for upversioning & document
- **Legacy Data Conversion**
- **Traceability of Data**
  - iDRG (Integrated Data Reviewer's Guide): the guide associated with integrated clinical study data or analysis data
  - Document the traceability process in the iDRG, ensuring all derivations and transformations are clearly explained

# Action Items



Harmonization overall process includes six modules:

- Harmonize Study Design
- Consistency Check Across Studies based on SDTM IG (3.4)
- Generate Comprehensive Codelist to Review value differences
- Harmonize SDTM Domains Mapping Specification Document
  - For each SDTM domain, document **mapping rules** for all variables and ensure they are harmonized across studies.
- Data Validation
  - Perform **comprehensive validation** on key variables and derivations
- Prepare iDRG

# Integrated Datasets



**THE FORUM**

For Collaborative Research<sup>SM</sup>

Berkeley's Hub for Regulatory Science

## ■ Source Study Data

Study Identifier (STUDYID)	Protocol Number	Source Data Standard	Cutoff Date or DBL Date / Study Status
EYP-001-202	EYP-001-202	SDTM v1.4 / IG v3.2, EudraCT number 2018-003119-22	2021-07-07/ Terminated
SHP-626-201	SHP-626-201	SDTM v1.4/ IG v3.1.2, SDTM CT version 2016-12-16	2018-07-27/ Completed

## ■ Trial Design

Study	Study Drug	Arm	Subject Enrollment	Treatment Period	Max Weeks	Dose frequency	Start Trial Date	End Trial Date
EYP001-002	Vonafexor / Placebo	7	120/39	12 weeks	18 weeks	Once daily / Twice daily	2019/1/30	2021/7/6
SHP626-201	SHP626 / Placebo	4	197/49	48 weeks	53 weeks	Once daily	2016/10/24	2018/7/27

# Overview - 1

## ■ Integrated Datasets – all STDM domains



**THE FORUM**

For Collaborative Research<sup>SM</sup>

Berkeley's Hub for Regulatory Science

Dataset	Domain Label	Dataset	Domain Label	Dataset	Domain Label
AE	Adverse Events	HO	Healthcare Encounters	SC	Subject Characteristics
CE	Clinical Events	IE	Inclusion/Exclusion Criteria Not Met	SE	Subject Elements
CM	Concomitant Medications	LB	Laboratory Test Results	SU	Substance Use
CO	Comments	MH	Medical History	SV	Subject Visits
DA	Drug Accountability	MI	Microscopic Findings	TA	Trial Arms
DM	Demographics	ML	Meal Data	TE	Trial Elements
DS	Disposition	MO	Morphology	TI	Trial Inclusion/Exclusion Criteria
DV	Protocol Deviations	PC	Pharmacokinetics Concentrations	TS	Trial Summary
EC	Exposure as Collected	PE	Physical Examination	TV	Trial Visits
EG	ECG Test Results	PR	Procedures	VE	Events
EX	Exposure	QS	Questionnaires	VS	Vital Signs
FA	Findings About Events or Interventions	RP	Reproductive System Findings	SUPP-	Supplemental Qualifiers

# Overview - 2



**THE FORUM**

For Collaborative Research<sup>SM</sup>

Berkeley's Hub for Regulatory Science

## Data Standards:

Standard or Dictionary	Versions Used
SDTM	SDTM v2.0 / IG v3.4
SDTM Controlled Terminology	CDISC Controlled Terminology 2024-03-29
Data Definitions	Define.xml
NASH / MASH	FDA: Technical Specifications for Submitting Clinical Trial Data Sets for Treatment of Noncirrhotic Nonalcoholic Steatohepatitis Guidance for Industry
Medical Events Dictionary	MedDRA 20
Pinnacle 21	Software 4.1.0; Validation Engine Name FDA 2304.3

## Harmonization Summary:

SDTMs (XX, RELREC, FA required)	Integrated SDTM IG	EYP-001-202 Only	Common in all studies	SHP-626-201 Only	Harmonized Datasets
	Domain Count / Variable Count				
Integrated IG	64 / 2195				
EYP-001-202 (total 27/465)		11 / 246			
SHP-626-201 & EYP-001-202 Common			16 / 219		
SHP-626-201 (total 24/366)				8 / 147	
Harmonized datasets (total 35/612)					35 / 612

- Notes: SUPPxx Unharmonized (stacked), total 14 domains / 136 variables

# Overview - 3

## ■ Issues Summary

- Six case scenarios listing the Pinnacle 21 report results: from comparing harmonized SDTM datasets with original studies' SDTM datasets.

<b>Six Cases (Total N = 487)</b>	<b>N (count)</b>	<b>Percentage (%)</b>	<b>Harmonized Dataset Issues</b>	<b>SHP-626-201's Original Issues</b>	<b>EYP-001-202's Original Issues</b>
<b>Case 1</b> Harmonized Issues are Consistent with issues from original studies	205	42%	2	1	1
<b>Case 2</b> Harmonized process corrected original SHP-626-201 Issues; no issues in Harmonized SDTMs	149	31%	0	1	0
<b>Case 3</b> Harmonized process corrected original EYP-001-202 Issues; no issues in Harmonized SDTMs	25	5%	0	0	1
<b>Case 4</b> Harmonized process corrected all original studies' Issues; no issues in Harmonized SDTMs	0	0	0	1	1
<b>Case 5</b> Harmonized process applied most current SDTMIG, CT and compliance checks when no issues existed in the original studies (False-Positive)	109	22%	2	0	0
<b>Case 6</b> Harmonized process applied most recent SDTMIG and compliance checks when similar issues exist in the original studies (False-Positive)	0	0	3	1	1

# Overview - 4

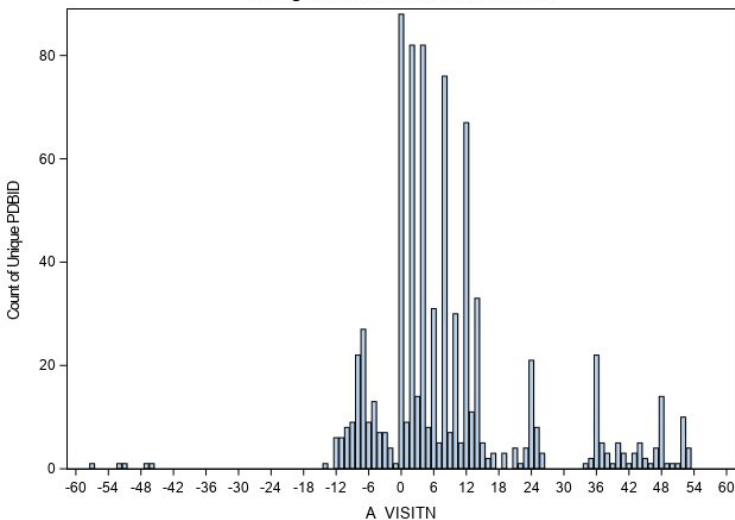


## Table List and Figure (TLFs) Output

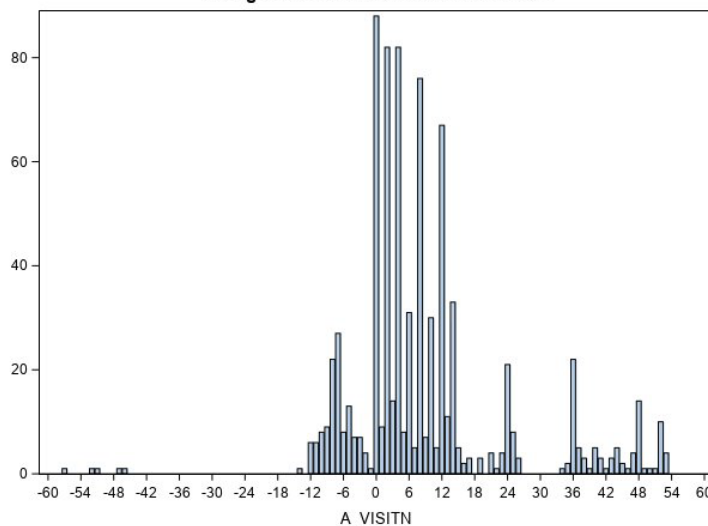
### Demographics and Baseline Characteristics

Age					Sex				Race							
YEARS					F		M		ASIAN		BLACK OR AFRICAN AMERICAN		WHITE		MISSING	
N	Mean	SD	Min	Max	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
88	54.41	11.96	22	80	52	59%	36	41%	4	5%	4	5%	41	47%	39	44%

Histogram of ALT Visit Week Number



Histogram of Bilirubin Visit Week Number

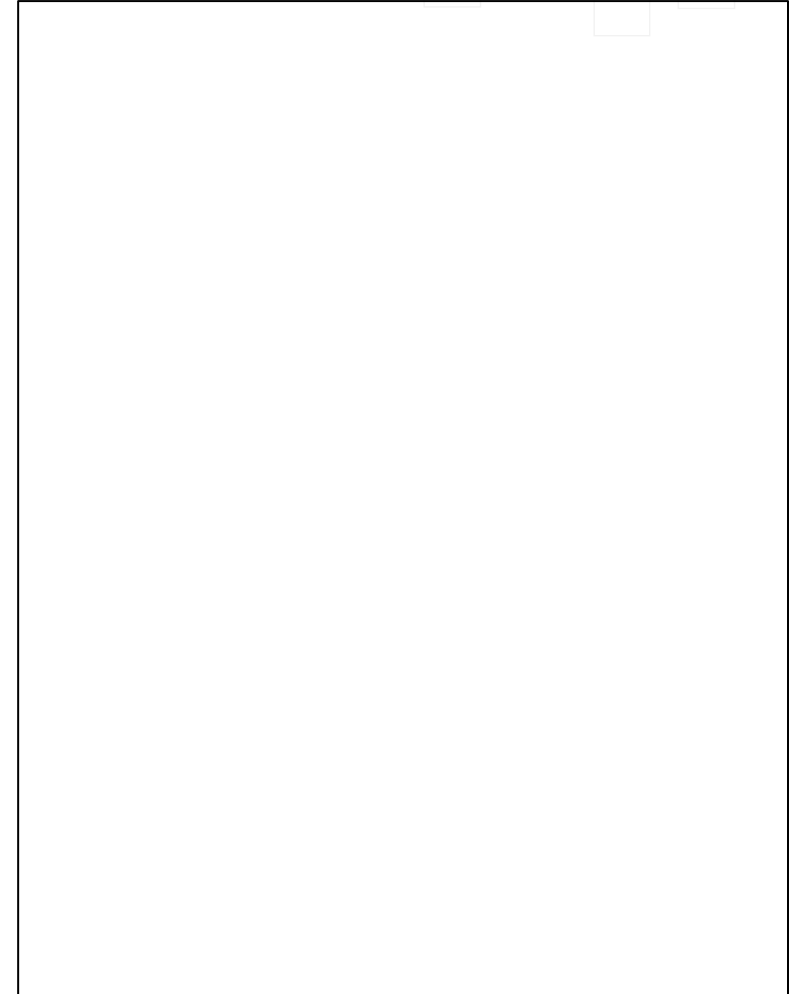


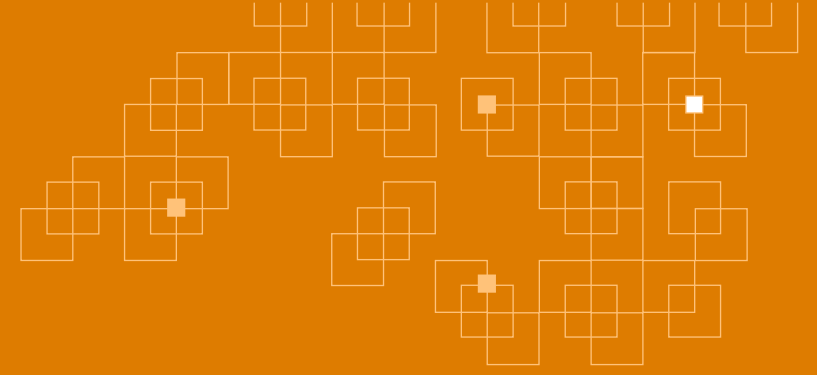
### Summary of Concomitant Medication

Preferred Medication Term	Placebo (N=88)	Preferred Medication Term	Placebo (N=88)
ANY MEDICATION	86 ( 97.7%)	MULTIVITAMINS, PLAIN	6 ( 6.8%)
METFORMIN	39 ( 44.3%)	OTHER VIRAL VACCINES	6 ( 6.8%)
ACETYLSALICYLIC ACID	23 ( 26.1%)	SERTRALINE	6 ( 6.8%)
LOSARTAN	17 ( 19.3%)	SIMVASTATIN	6 ( 6.8%)
OMEPRAZOLE	15 ( 17.0%)	ATORVASTATIN	5 ( 5.7%)
AMLODIPINE	14 ( 15.9%)	CETIRIZINE HYDROCHLORIDE	5 ( 5.7%)
COLECALCIFEROL	14 ( 15.9%)	CLONAZEPAM	5 ( 5.7%)
IBUPROFEN	12 ( 13.6%)	EMPAGLIFLOZIN	5 ( 5.7%)
LEVOTHYROXINE	12 ( 13.6%)	ESCITALOPRAM OXALATE	5 ( 5.7%)
PARACETAMOL	11 ( 12.5%)	FUROSEMIDE	5 ( 5.7%)
HYDROCHLOROTHIAZIDE	10 ( 11.4%)	HYDROCODONE	5 ( 5.7%)
VITAMIN D NOS	10 ( 11.4%)	LORATADINE	5 ( 5.7%)
ASCORBIC ACID	9 ( 10.2%)	METFORMIN HYDROCHLORIDE	5 ( 5.7%)
LISINAPRIL	9 ( 10.2%)	METOPROLOL	5 ( 5.7%)
SALBUTAMOL	8 ( 9.1%)	PRAVASTATIN	5 ( 5.7%)
AZITHROMYCIN	7 ( 8.0%)	ROSUVASTATIN	5 ( 5.7%)
FISH OIL	7 ( 8.0%)	TOPIRAMATE	5 ( 5.7%)
GABAPENTIN	7 ( 8.0%)	ALLOPURINOL	4 ( 4.5%)
IRBESARTAN	7 ( 8.0%)	CALCIUM	4 ( 4.5%)
LEVOTHYROXINE SODIUM	7 ( 8.0%)	DIAZEPAM	4 ( 4.5%)
VITAMINS NOS	7 ( 8.0%)	DICYCLOVERINE	4 ( 4.5%)
ALPRAZOLAM	6 ( 6.8%)	DULOXETINE HYDROCHLORIDE	4 ( 4.5%)
AMOXICILLIN	6 ( 6.8%)	ESTRADIOL	4 ( 4.5%)
ATENOLOL	6 ( 6.8%)	FEXOFENADINE	4 ( 4.5%)
GLIMEPIRIDE	6 ( 6.8%)	FLUOXETINE	4 ( 4.5%)

# FDA Submission Package

- Completed materials:
  - STDM dataset: 49 dataset files
  - SAS programs and SDTM Mapping Specs files
  - Defined XML
  - Table List and Figure Outputs (TLFs)
  - Integrated Clinical Study Data Reviewer's Guide
  - Protocols, aCRFs - original studies
- Next steps:
  - Research proposal, SAP – project specific





# MASH PDB: Proposed Future Analysis

# PDB Analysis - 1



## 1. Characterizing Placebo Response Rates in PDB

- Objective: Establish benchmark placebo response rates using harmonized PDB data
- FDA-defined endpoints:
  - Histologic resolution of MASH without worsening of fibrosis
  - Improvement in fibrosis without worsening of MASH
  - Associated non-invasive test (NIT) markers
- Rationale:
  - Published Randomized Clinical Trials report placebo response rates between 10–30% or more
  - Quantifying this will allow robust cross-trial comparisons & inform trial design

# Primary Objective & Endpoint



**THE FORUM**  
For Collaborative Research<sup>SM</sup>  
Berkeley's Hub for Regulatory Science

- Primary Objective:
  - Determine **the placebo response rate** observed **between two liver biopsies** obtained at least 16 weeks apart
- Primary Endpoint:
  - a **>= 1 stage improvement** in histologic fibrosis (based on the NASH CRN scoring system) **without worsening of steatohepatitis** in paired liver biopsies
  - Endpoint type: categorical (yes/no) or continuous
  - Potential variables to evaluate for association of a placebo response:
    - Age, race, sex, weight, BMI, comorbidities
    - Histologic fibrosis stage and NAS steatohepatitis score
    - LSM, ELF score, Fib 4, AST/ALT
    - Concomitant medications

# Secondary Objectives & Endpoints

- Secondary Objective 1
  - To determine **factors associated with placebo response**.
- Secondary Objective 2
  - To **determine the rate of placebo response in NITs** after at least 16 weeks on the placebo arm and **factors** associated with placebo response
- Secondary Endpoint 1
  - **fibrosis regression** as per the primary endpoint
- Secondary Endpoint 2
  - 30% reduction in LSM by VCTE
  - 0.5 reduction in ELF
  - 30% reduction in LSM by VCTE **and** 0.5 reduction in ELF
  - 19% reduction in LSM by MRE
  - Potential variables to evaluate association of placebo response defined by NITs:
    - Age, race, sex, weight, BMI, comorbidities
    - Histologic fibrosis stage and NAS steatohepatitis score
    - LSM, ELF score, Fib 4, AST/ALT
    - Concomitant medications



## ■ Study Population

- Analysis of data from the placebo arms of clinical trials in the Forum MASH PDB

## ■ Data Analysis – Statistical Approach

- Primary analysis:
  - The point estimate of the fibrosis response rate and its 95% confidence interval will be calculated
  - Assuming a theoretical placebo response rate of 20%, a sample size of 600 placebo recipients will provide for a 95% confidence interval with a margin of error of approximately +/- 4%
- Secondary Objective 1:
  - Multivariable logistic regression will be used to evaluate the factors associated with a response in fibrosis progression ( $\geq 1$  stage reduction) as a bivariate categorical endpoint
- Descriptive analyses of the population with no change in fibrosis

# Proposed Future Analysis - 2



## 2. Digital Twin (DT) Methodology: Feasibility & Validation

- Objective:

Explore the feasibility of digital twin methodology using the placebo-arm patients.

- Design: Internal Validation Without Treatment Arm

Since the PDB contains only placebo patients from harmonized MASH RCTs, we use a pseudo-treatment design to test whether DT matching can reliably reproduce disease trajectories.

# Proposed Future Analysis - 2



**THE FORUM**  
For Collaborative Research<sup>SM</sup>  
*Berkeley's Hub for Regulatory Science*

## ■ Approach:

### ■ Define baseline features for matching:

- Age, sex, BMI, fibrosis stage, NAS components, ALT, diabetes status, NIT scores, etc.

### ■ Stratified sampling to create two groups:

- Group A: Random sample of placebo patients (treated as “pseudo-treated”)
- Group B: Remaining placebo patients, used as the digital twin pool
- Sampling is stratified to ensure balance across key baseline characteristics

### ■ Match digital twins:

- For each patient in Group A, identify matched patient(s) from Group B using baseline data
- Matching methods may include propensity scores, Mahalanobis distance, or kNN

### ■ Evaluate predictive accuracy:

- Compare actual observed outcomes (e.g., fibrosis improvement, ALT, NITs) between Group A patients and their matched digital twins from Group B
- Assess concordance using Mean Absolute Error, correlation, or classification accuracy

# Proposed Future Analysis - 3



## 3. External Control Arms in RWD Settings

- Objective: Combine PDB and Digital Twin methods to construct external sources (e.g., real-world data or sponsor-submitted trials) to estimate counterfactual outcomes and simulate control arms.
- Rationale:
  - Regulatory interest in synthetic/external controls to augment single-arm or pragmatic trials
  - PDB offers a curated, harmonized source of control data
- Potential Use Cases:
  - Informing hybrid designs with both RCT and observational components
  - Contextualizing outcomes in early-phase RWD-based trials

# Contribution to Future RCTs

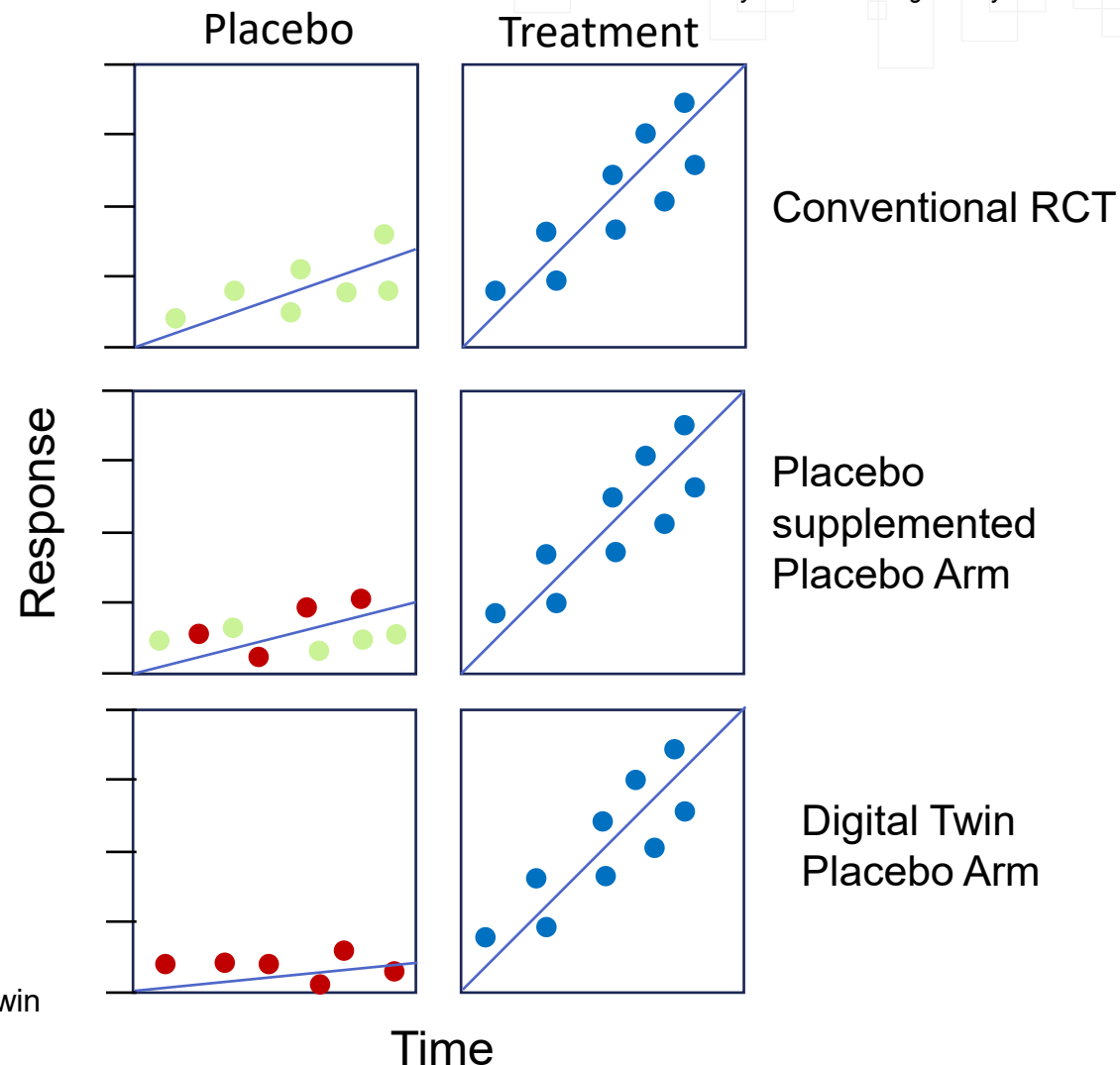
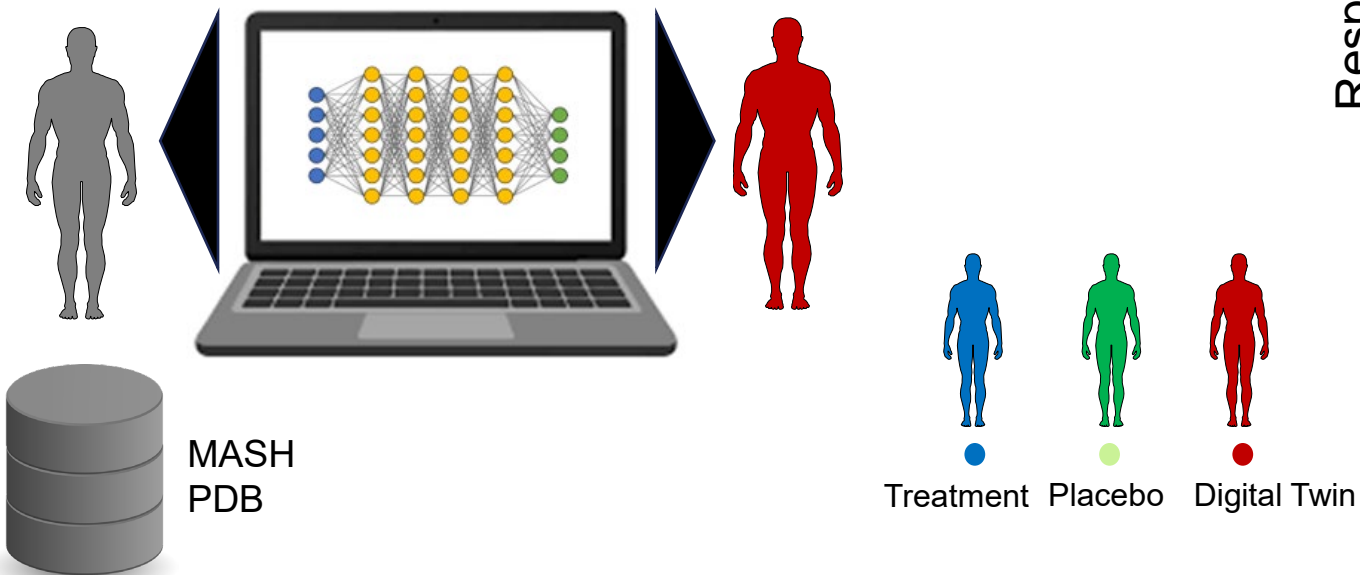


**THE FORUM**

For Collaborative Research<sup>SM</sup>

Berkeley's Hub for Regulatory Science

- Reduce size of the Placebo Arm to increase power and decrease time to recruitment
- More accurate representation of real-world non-treatment
- Replace the Placebo Arm in cases where placebo arm is not ethical



Questions  
&  
Thank you

