MULTI-REGIONAL CLINICAL TRIALS

THE MRCT CENTER OF BRIGHAM AND WOMEN'S HOSPITAL and HARVARD

AI Digital Twins and Synthetic Data: Application to Clinical Trials

DATE
September 30, 2025

TIME
11am–12:30pm ET

Disclaimer

The views expressed today do not imply endorsement or reflect the views or policies of Mass General Brigham, Harvard Medical School, or any affiliated organization or entity.

The MRCT Center is supported by voluntary contributions from foundations, corporations, international organizations, academic institutions, and government entities (see www.MRCTCenter.org), as well as by grants.

We are committed to autonomy in our research and to transparency in our relationships. The MRCT Center and its directors retain responsibility and final control of the content of any products, results, and deliverables.

Welcome!

Thank you for joining this webinar today!

Tips for today's session

- Use the Q&A for your questions – we will do our best to answer live.
- Feel free to use the Closed Captioning available on the Zoom toolbar.
- Most of the links in our presentations will be shared in the Chat.

The recording, slides, and any additional materials will be available next week.

The Multi-Regional Clinical Trials Center (MRCT Center)

The MRCT Center is a research and policy center focused on addressing the conduct, oversight, ethics and regulatory environment for clinical trials.

Our Vision

Improve the integrity, safety, and rigor of global clinical trials.

Our Mission

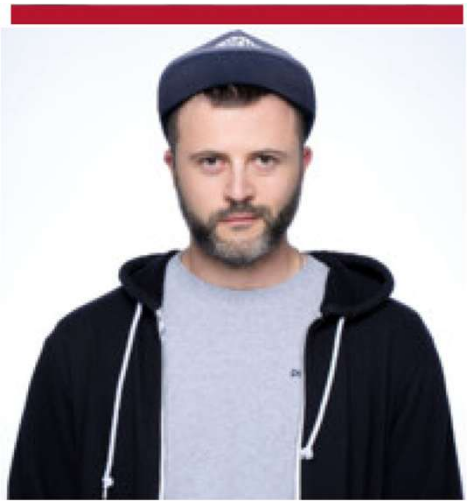
Engage diverse stakeholders to define emerging issues in global clinical trials and to create and implement ethical, actionable, and practical solutions.

AI Digital Twins and Synthetic Data: Application to Clinical Trials

MULTI-REGIONAL CLINICAL TRIALS
THE MRCT CENTER OF BRIGHAM AND WOMEN'S HOSPITAL and HARVARD

Barbara E. Bierer, MD

Professor of Medicine, Harvard Medical School
Faculty Director, MRCT Center of BWH & Harvard
bbierer@bwh.harvard.edu



Daniele Bertolini

Machine Learning Scientist

UNLEARN

30 September 2025

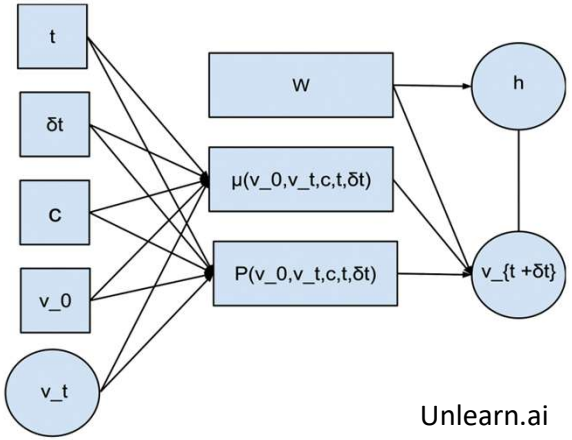
Digital Twins and Synthetic Data

Digital twins refer to simulated models of individual patients designed to predict disease trajectories and/or treatment responses, offering the potential to enhance statistical power, optimize trial design, and limit the number of participants assigned to a control arm.

Synthetic data are artificially generated datasets that mirror the statistical properties of verified clinical data.

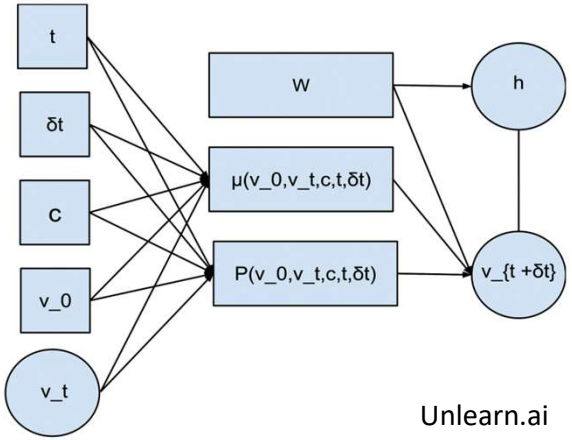
Both technologies are being developed for use in the design, conduct, and analysis of clinical trials, and with the intention of use for regulatory submissions.

My attempts to learn

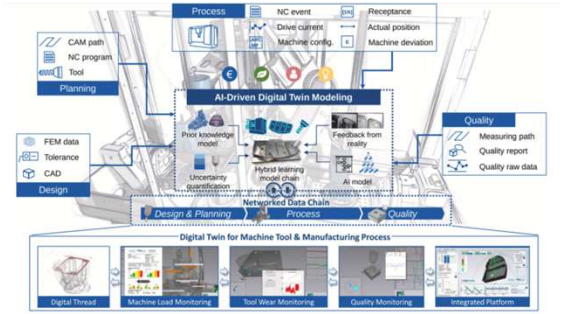


Unlearn.ai

My attempts to learn

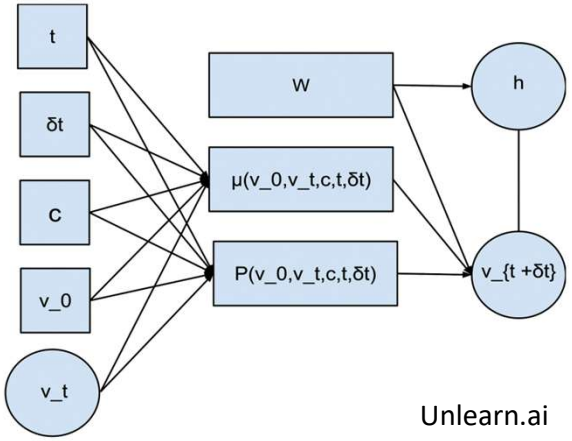


Unlearn.ai

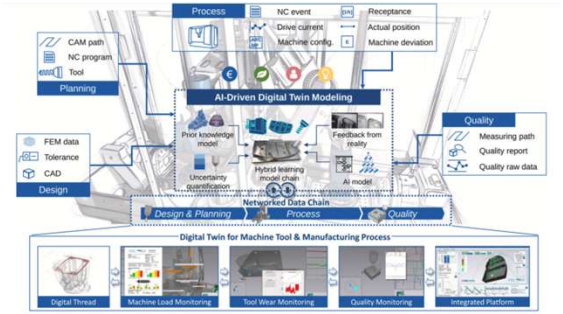


https://ars.els-cdn.com/content/image/1-s2.0-S0736584523000212-ga1_lrg.jpg

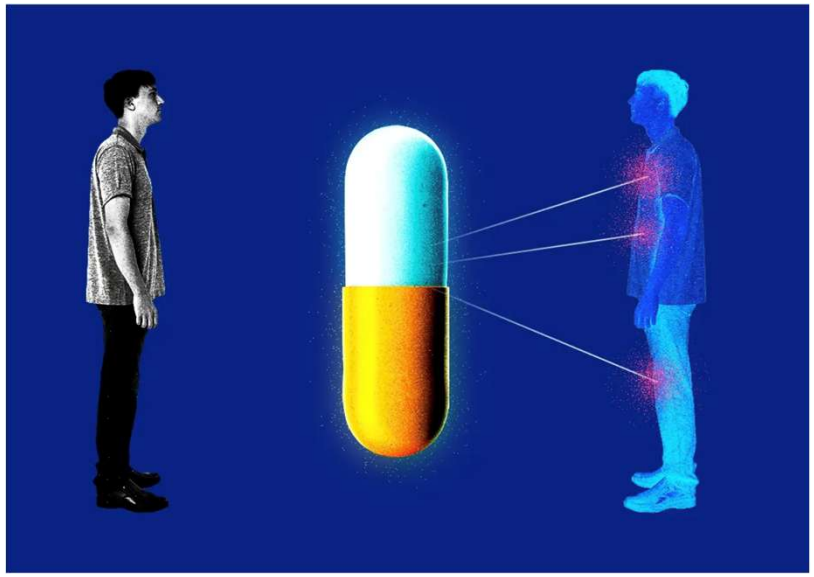
My attempts to learn



Unlearn.ai



https://ars.els-cdn.com/content/image/1-s2.0-S0736584523000212-ga1_lrg.jpg



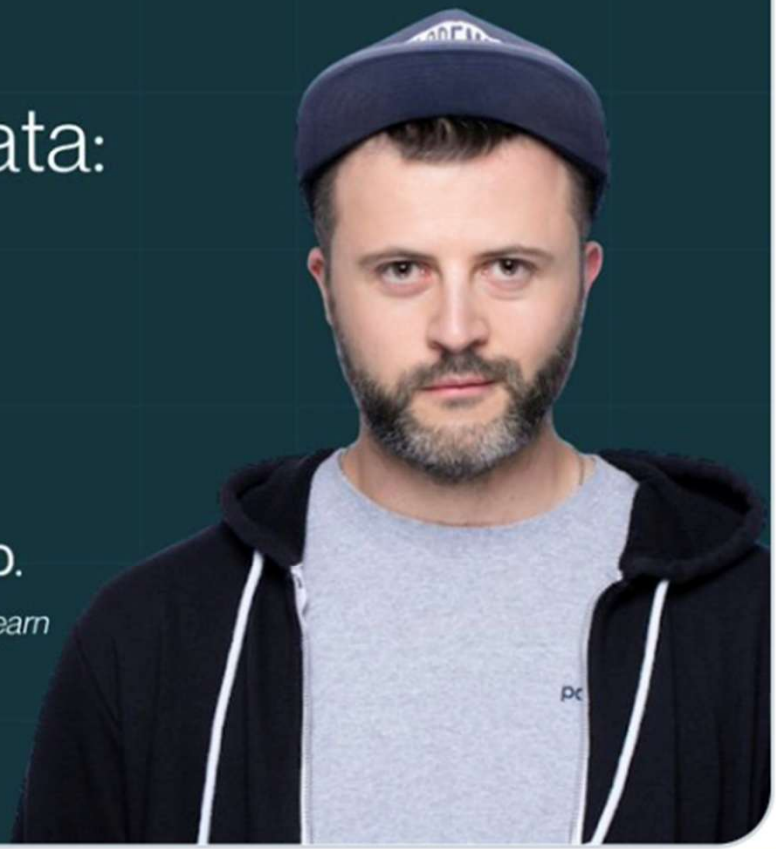
https://www.healthcare-brew.com/stories/2024/12/12/how-digital-twins-could-change-clinical-trials

MRCT Center | Webinar

Tuesday, September 30

11:00 AM ET

AI Digital Twins & Synthetic Data: Applications to Clinical Trials



Daniele Bertolini, Ph.D.

Machine Learning Scientist, *Unlearn*

UNLEARN

UNLEARN +

MULTI-REGIONAL CLINICAL TRIALS
THE mRCT CENTER of BRIGHAM AND WOMEN'S HOSPITAL and HARVARD

September 30, 2025

Introduction to Digital Twins

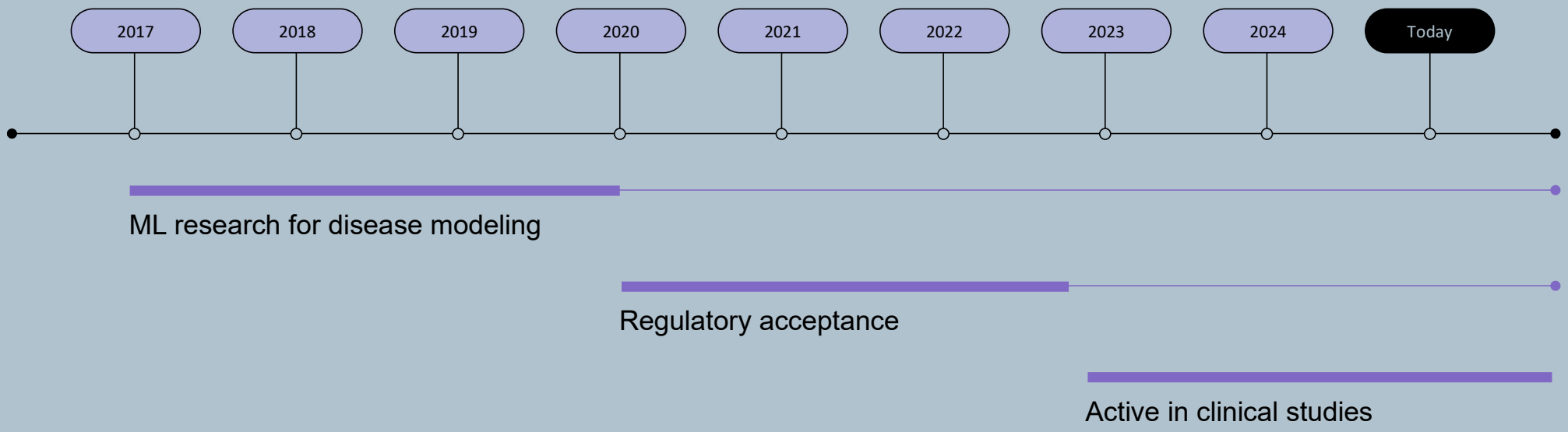
Daniele Bertolini
Unlearn.AI

Agenda

01	Digital Twins and Why They Are Useful
02	How to Build Digital Twins
03	How to Use Digital Twins

An introduction to Unlearn

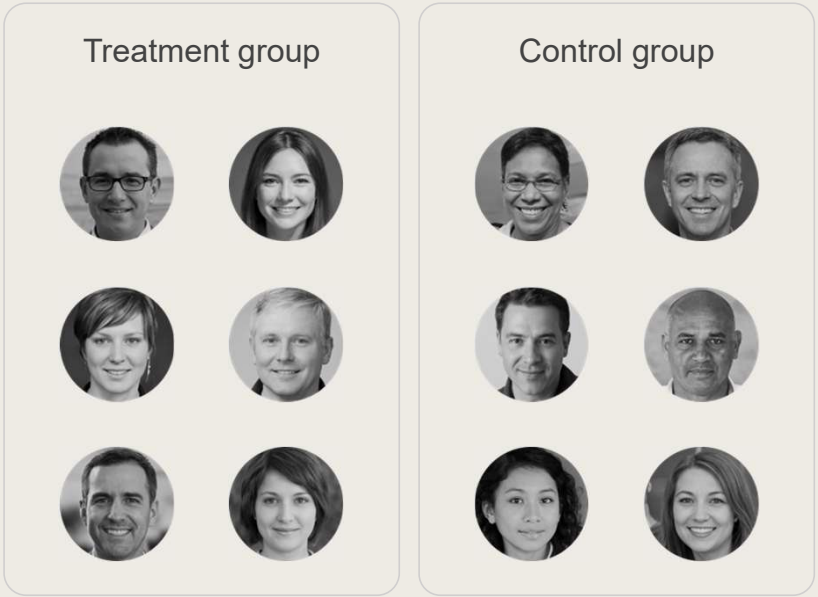
Unlearn has pioneered the use of Digital Twins in clinical trials.



Why run randomized controlled trials (RCTs)?

RCTs are the gold standard for evidence from a single experiment

- → Well-run experiments let sponsors make statistically robust statements about drug safety and effectiveness
- → Requires participants to receive standard of care; acceptable with equipoise
- → The community has developed decades of expertise in designing and running high quality **single** experiments



Why run so many RCTs?

Many different treatments, *very similar controls.*

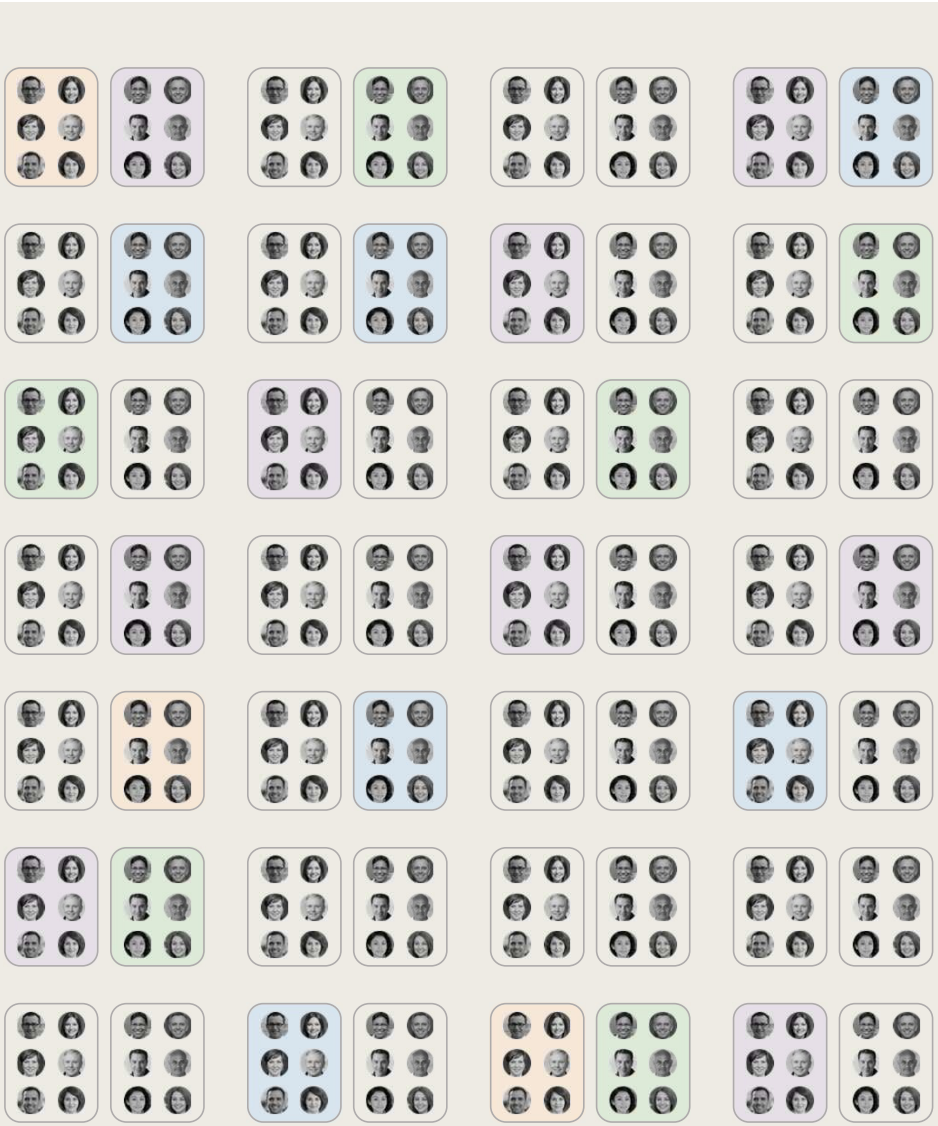
For many diseases, collectively we know *so much* about the indication from all the trials and studies we have run.

Why do we need to keep running RCTs? Why do we treat each trial like it's the first time we're evaluating a treatment in the indication?

It's so inefficient. This should make you mad.

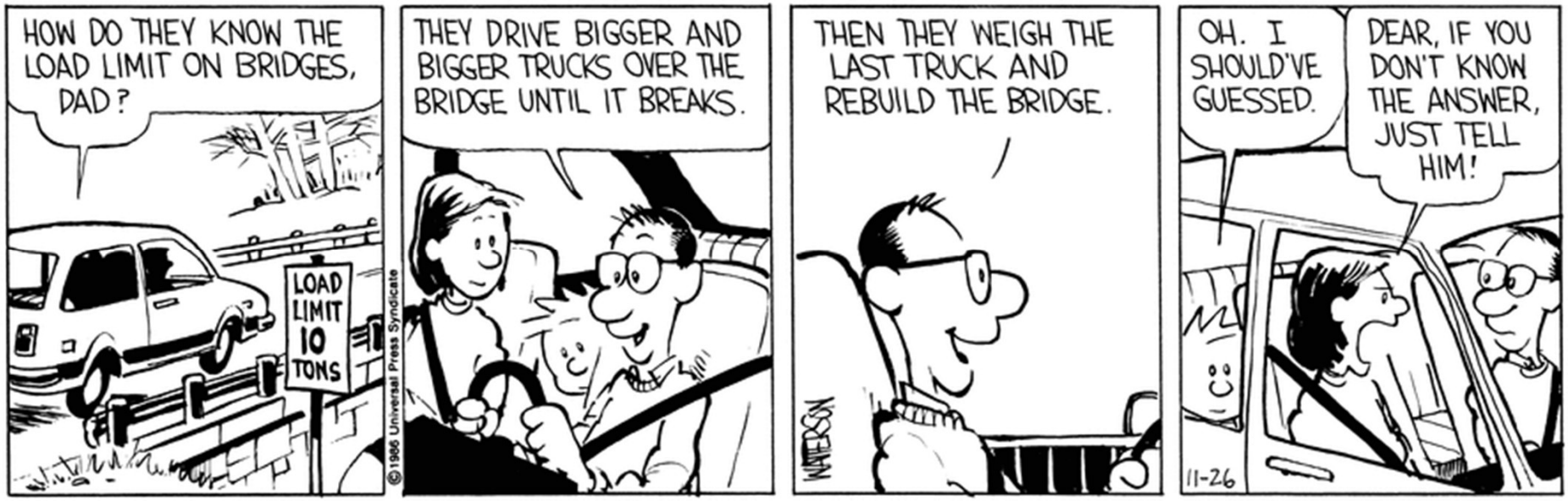
"People say they want placebo-controlled trials, but I always ask them would you be willing to die to give a p-value?"

Janet Woodcock, 2019



Digital Twins can help fix this (and do much more).

Digital Twins?



Calvin and Hobbes, Bill Watterson

Digital Twins?

A Digital Twin is a computational model of a physical object that generates data which are indistinguishable from its physical counterpart and can be dynamically updated.

- → The concept was first introduced by NASA in the ‘70s as a living model of the Apollo 13 and played a key role in understanding consequences and remedies of the oxygen tank explosion.
- → The term itself became popular in the early ‘00s in manufacturing and product development.
- → This model of a real-world entity permits virtual interactions that support forecasting, experimentation, and simulation involving that entity.

Digital Twin

It integrates data that paint a comprehensive picture of the bridge:

- → Response to traffic or other external stimuli on the roadway, support, and structural elements
- → Correlation between different places on the bridge and patterns of strain or wear
- → Extrapolation of impacts on the bridge based on data from other bridges, e.g. from major events

It is a tool to avoid running expensive, risky experiments on valuable infrastructure

It can be used to:

- → Anticipate wear and tear, gradual damage
- → Plan maintenance
- → Identify long term structural vulnerabilities
- → Anticipate responses to catastrophic events like earthquakes, fires

Digital Twin

A Digital Twin of a patient is comprehensive computational model of that individual which makes predictions of future outcomes and can be dynamically updated with data from their physical twin.

It is typically built from multimodal patient-level and population-level data:

- → Demographic
- → Disease history
- → Biomarkers
- → Lab tests
- → Genetic information
- → Clinical assessments
- → Images, etc

It can be used to:

- → Forecast disease progression and improve efficiency of RCTs
- → Evaluate interventions, identify long term health impacts, anticipate risks for significant health events, etc.

Example of Digital Twins for disease modeling: ALS

Ways disease is measured in ALS patients

- → Demographics and disease history
- → Assessments of ALS symptoms
- → Respiratory function (vital capacity)
- → Muscle strength
- → Quality of life measures
- → Biomarkers
- → Lab tests and vitals
- → Disease milestones

We model disease at the level of clinical data, not mechanistically or biologically.

Participant's Digital Twin in Amyotrophic Lateral Sclerosis

Age (years): 55
Sex: Male
Race: Caucasian

Diagnosis: ALS Possible
Site of Onset: Bulbar
Symptom Onset: 623 days

Medications: none
FVC norm: 4.25 Liters
Height: 167 cm

Time (months)	Baseline	1	2	3	4	5
						
Alanine Aminotransferase	41	41.8 ± 12.3	42.3 ± 13.9	41.7 ± 16.5	42.5 ± 20.6	41.5 ± 19.4
Albumin	42	42.2 ± 1.6	42.1 ± 2.1	42.1 ± 2.4	42.0 ± 2.5	41.9 ± 2.8
Alkaline Phosphatase	76	77.8 ± 11.0	77.6 ± 13.7	78.5 ± 17.5	79.4 ± 20.4	81.0 ± 21.6
ALSFRS Climbing	4	3.9 ± 0.4	3.7 ± 0.5	3.6 ± 0.6	3.5 ± 0.8	3.4 ± 0.8
ALSFRS Cutting	4	3.9 ± 0.2	3.9 ± 0.4	3.8 ± 0.5	3.7 ± 0.6	3.6 ± 0.7
ALSFRS Dyspnea	3	2.9 ± 0.7	2.9 ± 0.8	2.9 ± 0.9	2.9 ± 1.0	2.8 ± 1.0
ALSFRS Handwriting	4	3.9 ± 0.3	3.8 ± 0.4	3.7 ± 0.5	3.7 ± 0.6	3.7 ± 0.6
ALSFRS Hygiene	4	3.9 ± 0.3	3.8 ± 0.5	3.7 ± 0.6	3.6 ± 0.6	3.5 ± 0.8
ALSFRS Insufficiency	4	4.0 ± 0.1	3.9 ± 0.2	3.9 ± 0.3	3.8 ± 0.4	3.8 ± 0.5
ALSFRS Orthopnea	4	4.0 ± 0.2	3.9 ± 0.3	3.8 ± 0.4	3.7 ± 0.5	3.7 ± 0.6
ALSFRS Salivation	3	2.8 ± 0.7	2.8 ± 0.8	2.7 ± 0.9	2.7 ± 1.0	2.6 ± 1.0
ALSFRS Speech	2	1.9 ± 0.6	1.9 ± 0.8	1.8 ± 0.9	1.9 ± 1.0	1.8 ± 1.0
ALSFRS Swallowing	3	2.9 ± 0.6	2.8 ± 0.8	2.8 ± 0.9	2.7 ± 0.9	2.8 ± 1.0
ALSFRS Turning	4	3.9 ± 0.3	3.8 ± 0.4	3.8 ± 0.4	3.7 ± 0.5	3.7 ± 0.7
ALSFRS Walking	4	3.9 ± 0.3	3.8 ± 0.4	3.8 ± 0.5	3.7 ± 0.6	3.6 ± 0.6
Aspartate Aminotransferase	23	24.0 ± 5.2	24.3 ± 7.0	25.2 ± 8.4	26.1 ± 8.9	26.8 ± 10.0
Basophils	0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
Blood Urea Nitrogen	5.7	5.6 ± 1.0	5.6 ± 1.2	5.8 ± 1.5	5.8 ± 1.5	5.8 ± 1.7
Calcium	2.3	2.3 ± 0.1	2.3 ± 0.1	2.3 ± 0.1	2.3 ± 0.1	2.3 ± 0.1
Chloride	102	102.7 ± 2.3	102.6 ± 2.9	102.7 ± 3.1	102.5 ± 3.2	102.5 ± 3.5
Cholesterol	5	5.0 ± 0.5	5.0 ± 0.6	5.1 ± 0.7	5.1 ± 0.8	5.1 ± 0.8
Creatine Kinase	139	146.8 ± 52.7	156.2 ± 77.2	169.0 ± 105.6	173.4 ± 115.8	175.1 ± 122.4
Creatinine	0.7	0.7 ± 0.1	0.7 ± 0.1	0.7 ± 0.1	0.7 ± 0.1	0.7 ± 0.2
Diastolic Blood Pressure	94	92.5 ± 8.2	91.8 ± 10.8	89.9 ± 11.6	89.7 ± 10.7	89.1 ± 11.9
Eosinophils	0.5	0.6 ± 0.3	0.6 ± 0.4	0.6 ± 0.5	0.6 ± 0.5	0.6 ± 0.6
FVC Liters	3.1	3.0 ± 0.3	3.0 ± 0.5	2.9 ± 0.5	2.9 ± 0.6	2.8 ± 0.7
Gamma Glutamyl Transferase	-	25.4 ± 6.9	26.2 ± 8.9	27.4 ± 11.2	27.5 ± 11.9	28.6 ± 13.9

Digital Twins and Synthetic Data

→ Synthetic data: artificially generated data that mimics the statistical properties of real data.

→ Digital Twins are a form of synthetic data. They are generated from observations of actual patients. Predictions can be updated if new information about their physical twins becomes available. They can be used for:

- → Forecast disease progression and improve efficiency of RCTs
- → Evaluate interventions, identify long term health impacts, anticipate risks for significant health events, etc.

→ Purely synthetic subjects do not take observed patient's data as input (e.g., they could represent hypothetical patient trajectories satisfying certain I/E criteria). Purely synthetic subjects have also interesting applications:

- → Simulate future trial
- → Augmentation for model training
- → As a tool to anonymize sensitive data

How to build Digital Twins

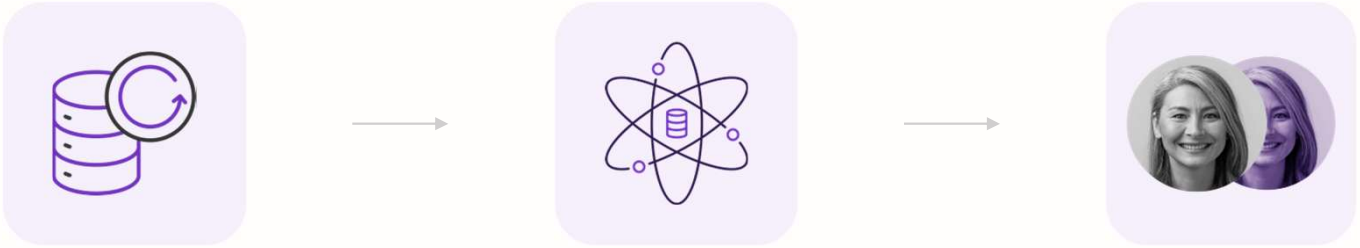
Machine Learning is an effective approach to solve this problem

Snapshot in time (e.g., baseline data)

Comprehensive predictions of future outcomes

What is Machine Learning?

Machine Learning is a branch of AI. It develops algorithms that learn patterns from data (instead of being explicitly programmed) and can make predictions for unseen scenarios. A typical ML workflow:



Why Machine Learning?

→ We could technically use mechanistic models and use training data to learn the parameters.

→ While this would make models apparently simpler, it limits flexibility and most likely does not provide a good fit to the true underlying process.

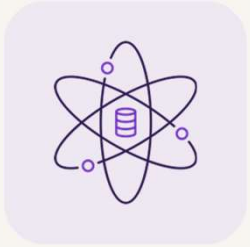
→ With machine learning, we let the data “speak” for itself and let the model (under proper training and regularization procedures) learn patterns from data. This leverages recent advances in AI and frees you from making risky assumptions (especially when modeling complex processes across many variables and time like for the case of a Digital Twins of a patient).

→ It works!

How to build Digital Twins

The context of use will inform data requirements and specific choices of model architecture, optimization, and validation.

Process Data



Train Model

Generate Predictions

Some unique features of clinical data

We focus on models trained on historical clinical data.
This is an example of Unlearn's dataset for Alzheimer's Disease.

Source	Size	Type	Disease Severity	Visit Cadence	ADAS and MMSE	CDR-SB	Biomarkers	Labs
CPAD	36 trials 15,000 subjects	AD trials control arms	mostly mild to moderate AD	3-months	✓	✓	x	✓
ADNI, NACC, OASIS, EPAD	7 studies 20,000 subjects	observational studies	1/3 MCI, 2/3 AD	6-months	✓	✓	✓	x

- → **Multisource**
  - E.g., past clinical trials, observational studies, registries, real world data
- → **"Small" and "wide"**
  - A large dataset has o(10K) subjects and typically o(100) variables
- → **Heterogeneous populations, visit cadence, standard of care**
  - E.g., MCI vs moderate AD
- → **Sparse**
  - E.g., not all studies measure the same variables
- → **Multimodal**
  - Continuous, binary, categorical, events, images, etc.

Some unique features of clinical data

We focus on models trained on historical clinical data.
This is an example of Unlearn's dataset for Alzheimer's Disease.

Source	Size	Type	Disease Severity	Visit Cadence	ADAS and MMSE	CDR-SB	Biomarkers	Labs
CPAD	36 trials 15,000 subjects	AD trials control arms	mostly mild to moderate AD	3-months	✓	✓	x	✓
ADNI, NACC, OASIS, EPAD	7 studies 20,000 subjects	observational studies	1/3 MCI, 2/3 AD	6-months	✓	✓	✓	x

- → **Multisource**
  - E.g., past clinical trials, observational studies, registries, real world data
- → **"Small" and "wide"**
  - A large dataset has o(10K) subjects and typically o(100) variables
- → **Heterogeneous populations, visit cadence, standard of care**
  - E.g., MCI vs moderate AD
- → **Sparse**
  - E.g., not all studies measure the same variables
- → **Multimodal**
  - Continuous, binary, categorical, events, images, etc.

Some unique features of clinical data

We focus on models trained on historical clinical data.
This is an example of Unlearn's dataset for Alzheimer's Disease.

Source	Size	Type	Disease Severity	Visit Cadence	ADAS and MMSE	CDR-SB	Biomarkers	Labs
CPAD	36 trials 15,000 subjects	AD trials control arms	mostly mild to moderate AD	3-months	✓	✓	x	✓
ADNI, NACC, OASIS, EPAD	7 studies 20,000 subjects	observational studies	1/3 MCI, 2/3 AD	6-months	✓	✓	✓	x

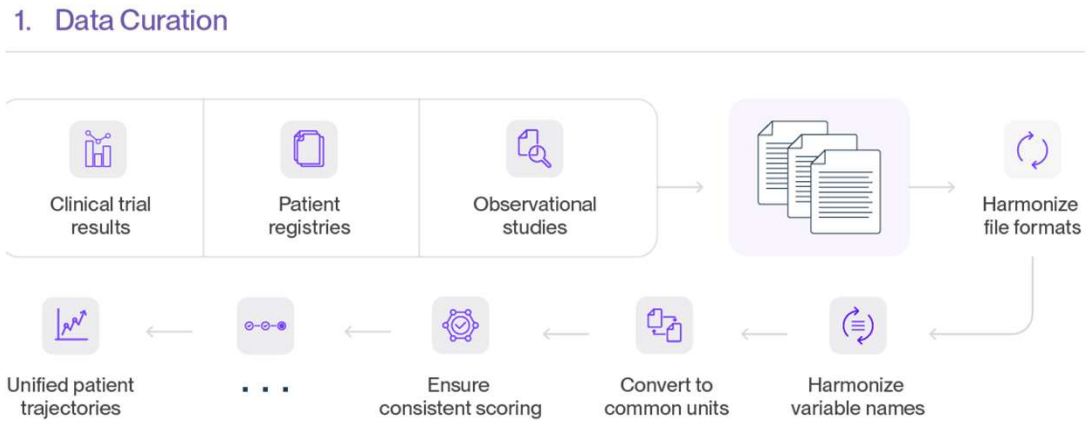
- → **Multisource**
  - E.g., past clinical trials, observational studies, registries, real world data
- → **"Small" and "wide"**
  - A large dataset has o(10K) subjects and typically o(100) variables
- → **Heterogeneous populations, visit cadence, standard of care**
  - E.g., MCI vs moderate AD
- → **Sparse**
  - E.g., not all studies measure the same variables
- → **Multimodal**
  - Continuous, binary, categorical, events, images, etc.

Data curation

Data curation is critical to build high quality resources for modeling.

- → Harmonization across datasets (trials, observational studies, registries)
- → Quality control to curate data suitable for creating models of disease
- → This step tends to be the most resource intensive

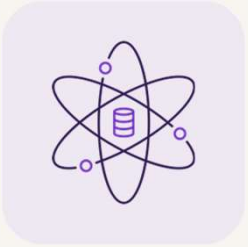
A typical data curation workflow.



How to build Digital Twins

The context of use will inform data requirements and specific choices of model architecture, optimization, and validation.

Process Data

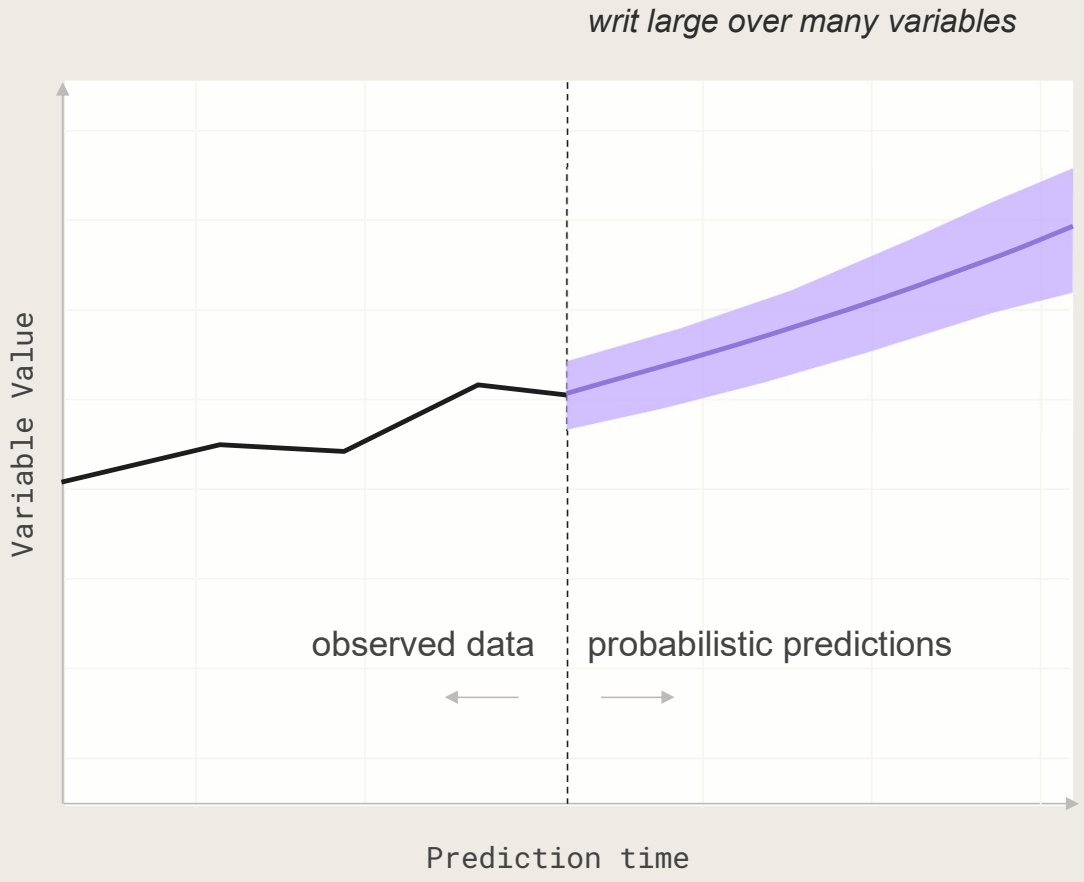


Train Model

Generate Predictions

Basic model requirements

- → Models should be able to make predictions across relevant variables and times
- → Models should handle some of the key data characteristics: sparse, multimodal, small
- → Preferably, predictions should be probabilistic (generative model)



Why generative models?

A generative model learns the full probability distribution over variables of interest and time.

- → The underlying disease progression is a stochastic process, knowing the probability distribution captures the associated uncertainty
- → Can sample conditionally (or even unconditionally) to simulate different scenarios
- → Samples are realistic (ideally, “statistically indistinguishable” from real trajectories)

$$p(X_{\text{static}}, X_{\text{longitudinal}}(t), X_{\text{events}}(t))$$

How can we use the joint distribution?

$$p(X_{\text{longitudinal}}(t > 0), X_{\text{events}}(t) \mid X_{\text{baseline}})$$

Conditioning on subject's baseline yields a model of their future health (their Digital Twin)

$$p(X_{\text{longitudinal}}(t > 0), X_{\text{events}}(t) | X_{\text{counterfactual}})$$

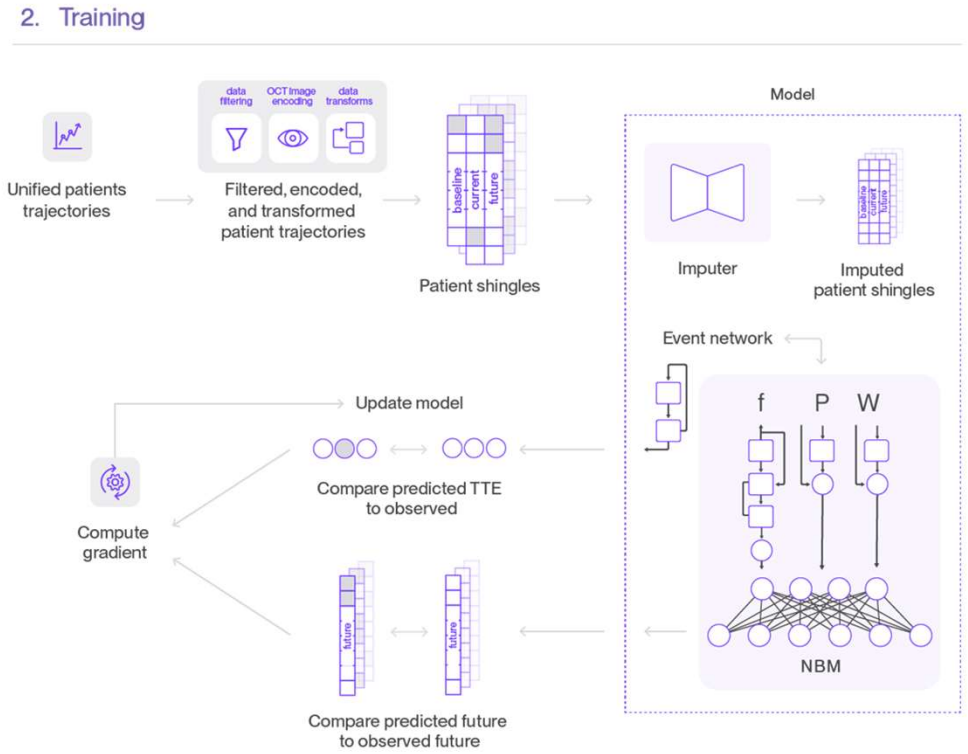
Could also explore counterfactuals: e.g., what would the progression of that same subject look like had they taken a different standard of care?

$$p(X_{\text{longitudinal}}(t > 0), X_{\text{events}}(t) | X_{\text{baseline}} \in \text{IE criteria})$$

Instead of conditioning on a specific subject baseline characteristics, we can sample populations that meet pre-set inclusion/exclusion criteria – this is useful to simulate different trial designs

Model training

- → Data is split into a training portion and an a hold-out portion.
- → Training optimizes the model for sample quality, with the goal that model generated data matches observed data.
- → Metrics like correlation and bias are evaluated on the hold-out portion for many model candidates and the best model is selected.
- → Specific optimization metrics should be chosen based on the context of use.



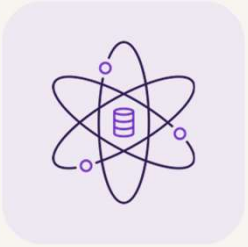
An example architecture using Neural Boltzmann Machines*. Training divides trajectories into (baseline, current, future) shingles and maximizes an approximate likelihood. Separate neural networks control mean and covariance of the joint distribution.

*Alam, Nameyeh, et al. "Digital Twin Generators for Disease Modeling." arXiv: 2405.01488; Lang, Alex, et al. "Neural Boltzmann Machines." arXiv:2305.08337.

How to build Digital Twins

The context of use will inform data requirements and specific choices of model architecture, optimization, and validation.

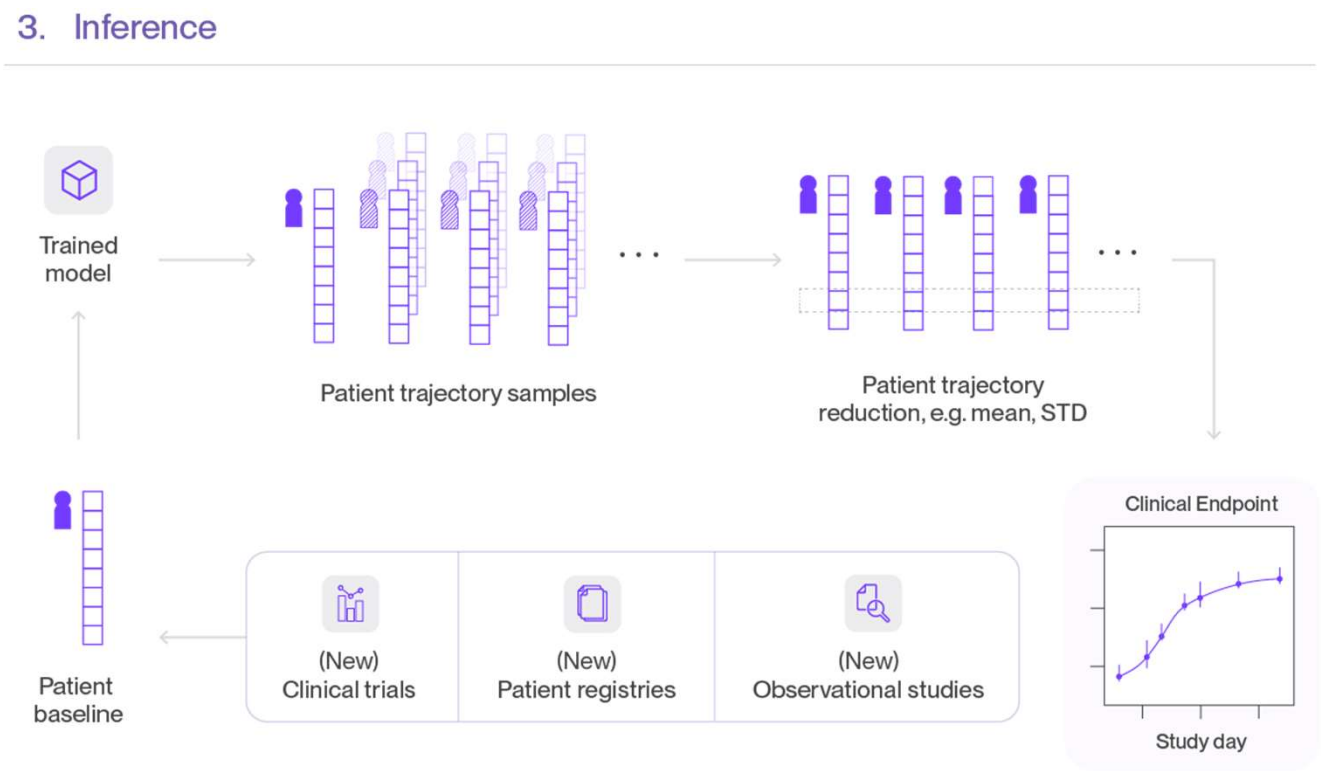
Process Data



Train Model

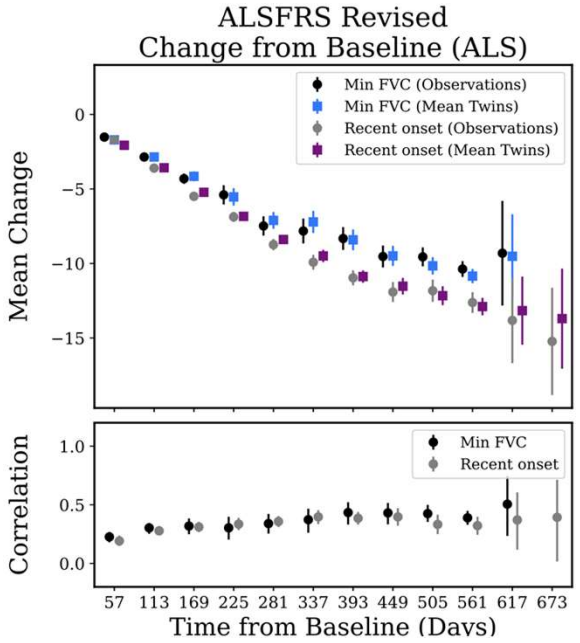
Generate Predictions

Generate Digital Twins

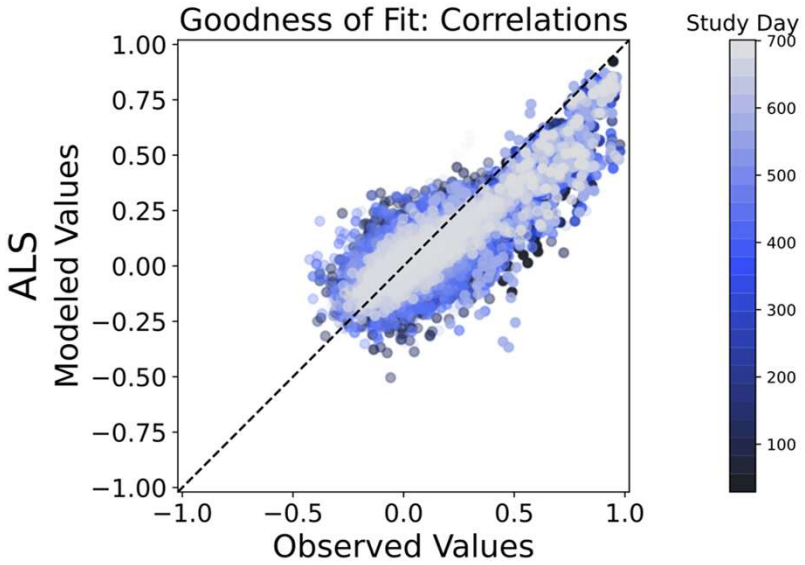


Example: ALS Digital Twins

Model predicts >50 variables, including clinical outcomes, labs, vitals, biomarkers.



ALSFRS-R measures progression of ALS. Correlation between observed and predicted outcome is a key metric to boost RCT efficiency.



The model captures pairwise correlations across variables. This is a key metric for design simulations.

Back to RCT efficiency.

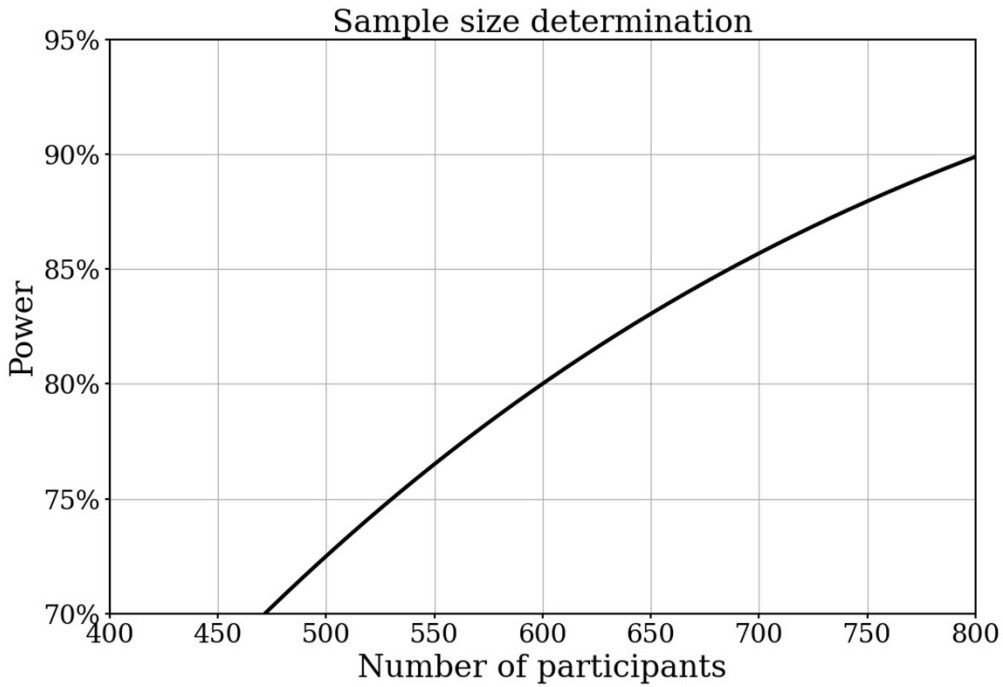
How can we improve trial efficiency?

Sample size is set to target a power threshold for one or more endpoints, based on:

- → The expected/target treatment effect size
- → The expected variance of the treatment effect
- → And other parameters (expected dropout rate, significance level)

Power curves are fairly steep. Example shown:

- → 600 participants: power = 80%
- → 800 participants: power = 90%

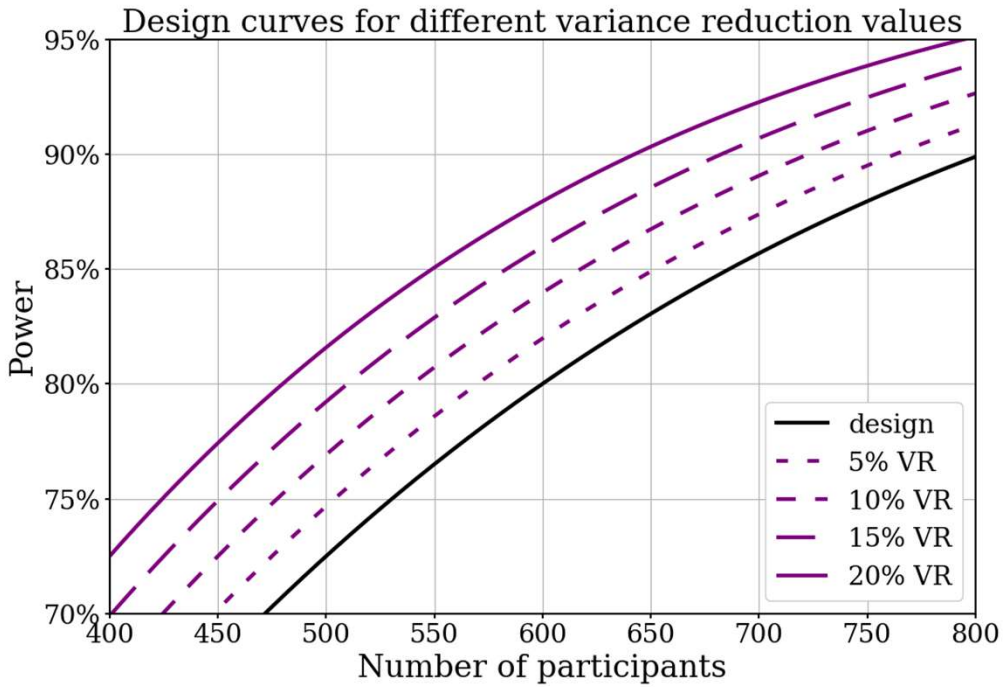


How can we improve trial efficiency?

Decreasing the variance of the treatment effect estimate even by small amounts leads to big gains in power.

How could we do this?

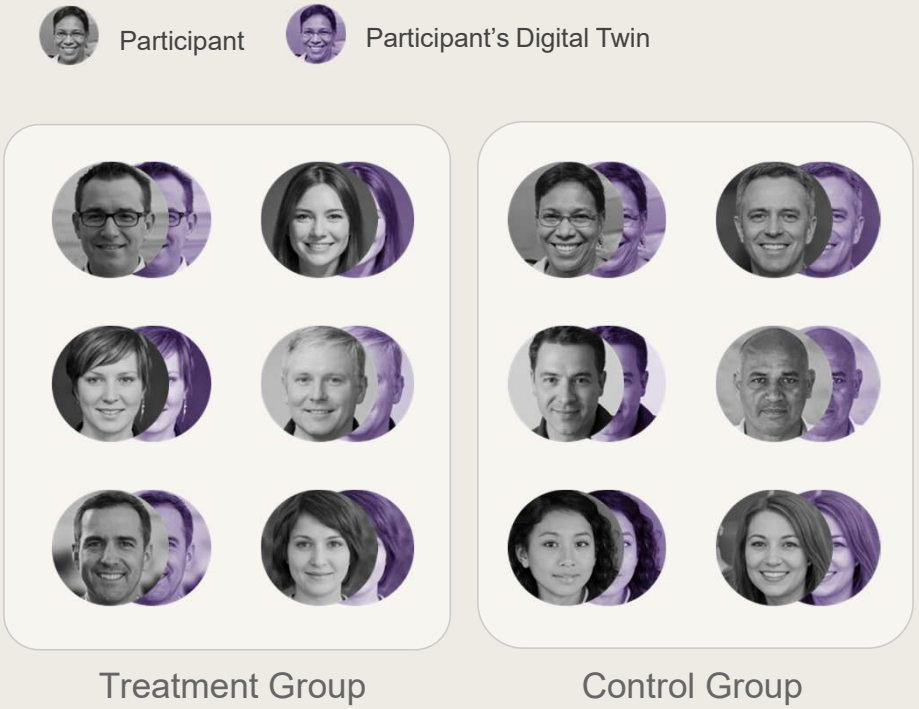
- → Enroll populations with a smaller variability (and same/larger treatment effect)
- → Use endpoints that have less variability
- → **Increase the efficiency of the statistical analysis**



How can we improve trial efficiency?

Digital Twins can be used as covariates in a covariate-adjusted analysis. Instead of adjusting for, e.g., sex, age or other baseline covariates, you can adjust for the predicted outcome of each subject. This super-covariate can maximize efficiency. The method is called **PROCOVA* (Prognostic Covariate Adjustment)**.

In this application, Digital Twins are *not* new participants. They are *more information* about the participants in the study. They make every participant's data more rich by bringing information from past clinical trials and observational studies into the analysis.



*Schuler, Alejandro, et al. "Increasing the efficiency of randomized trial estimates via linear adjustment for a prognostic score." The International Journal of Biostatistics 18.2 (2021): 329-356.

PROCOVA boosts trial efficiency

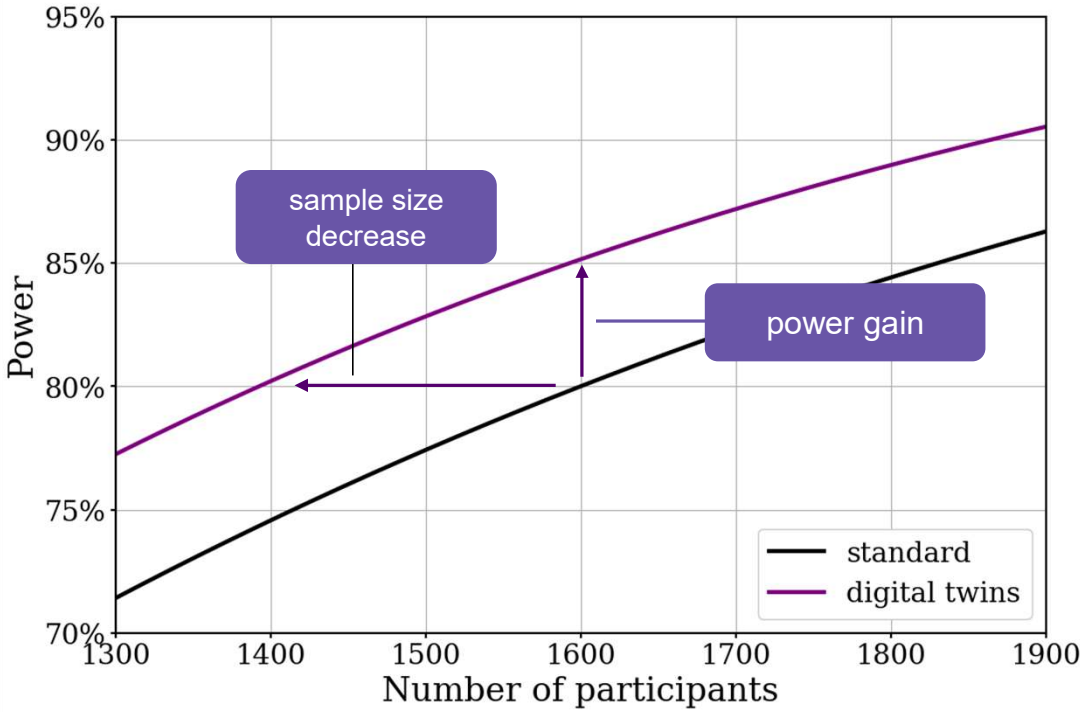
The prognostic value of the covariate can be used for:

Power add

→ Digital twins added at analysis time, this boosts the effective sample size, keeping enrollment constant

Sample size reduction

→ Expected prognostic value incorporated in the sample size calculation. Can decrease the placebo arm or both arms equally



Planning for future trials with Digital Twins

Hypothetical Alzheimer's Disease trial,
with expected variance reduction ~12%.

	Standard design	More power	Fewer participants
# participants	1600	1600 **(1787 ESS)**	1433 **(-167)**
# per arm	800:800	800:800	800:633
Power	80%	84.6% **(+4.6%)**	80%
Enrollment time	21 months	21 months	19 months **(-2 months)**
Cost	$560M	$560M	$501M **(-$59M)**

Positive interactions with FDA and EMA on the use of PROCOVA in phase 2 and 3 clinical trials

- → EMA qualified PROCOVA
- → FDA concurred with EMA's qualification
- → Growing interest and adoption in the statistics community

There are many applications of Digital Twins and Synthetic Data in drug development

We will discuss some of these applications, as well as details of PROCOVA in the next sessions of this webinar.

Using Digital Twins to boost power or reduce sample size is just one application. Digital Twins or synthetic subjects can also be used for, e.g.,

Trial Simulations

Trial Design Optimization

Subgroup Discovery

Early Phase Decision Making

Market Access

There are many applications of Digital Twins and Synthetic Data in drug development

All of these applications are fundamentally enabled ***in practice*** by operational capabilities:

High quality, ML ready datasets

Tools to analyze datasets and train and validate models

Applications to support users

(e.g. clinical development teams)

Support for model deployment and maintenance

Regulatory compliance, data security, and traceability

(see FDA's draft guidance *Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products*)

Summary

- Digital Twins of patients are computational models that provide comprehensive forecasts of outcomes over time and can be dynamically updated with data from their physical twins.
- Machine learning models trained on carefully harmonised data (e.g., from past trial, observational studies, registries) are an effective way to create Digital Twins.
- Key considerations in developing Digital Twins models include data availability and curation, context of use and optimization.
- We introduced a key application: PROCOVA. Digital Twins can be used as super covariates to reduce sample size with same power or boost power with same sample size.
- Next Topics include: PROCOVA deep-dive, synthetic controls for single arm trial, Bayesian analyses, design optimization and simulations, regulatory considerations.

UNLEARN +

MULTI-REGIONAL CLINICAL TRIALS
THE MRCT CENTER of BRIGHAM AND WOMEN'S HOSPITAL and HARVARD

Thank you.