UNLEARN

Digital twins for disease modeling and drug development

Applications for smarter, faster clinical trials

Jonathan Walsh

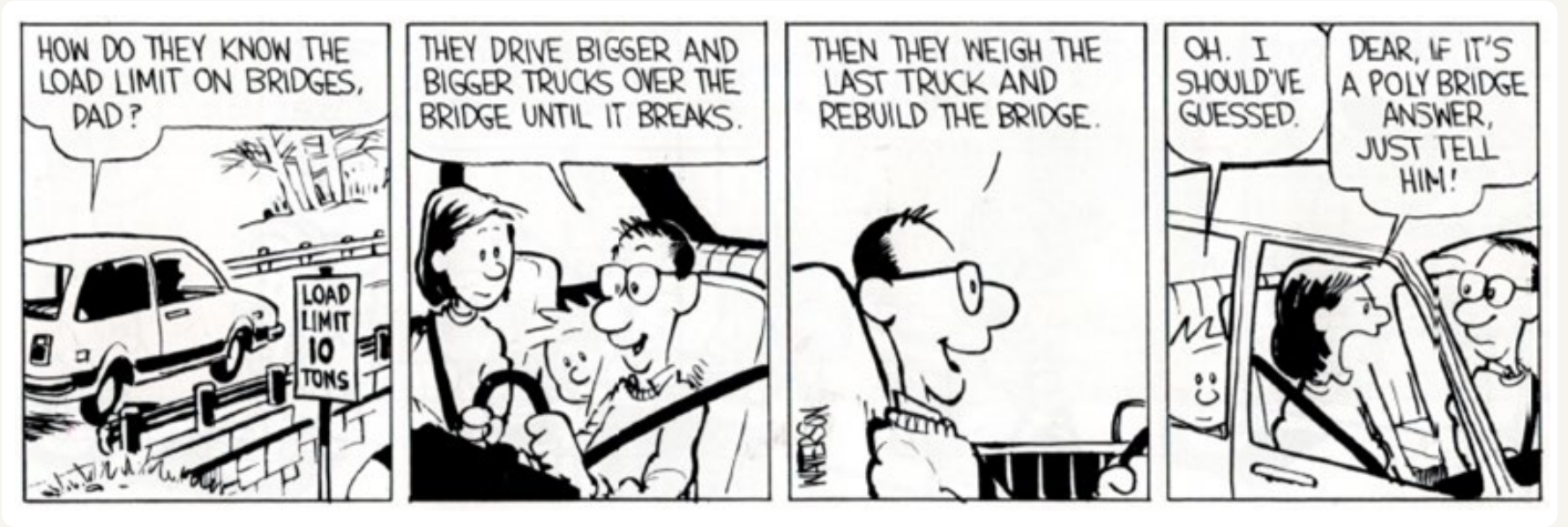
Agenda

1 ..... Why Clinical Trials?

2 ..... Digital Twins for Disease Modeling

3 ..... How to use Digital Twins in Clinical Trials

4 ..... Case Studies



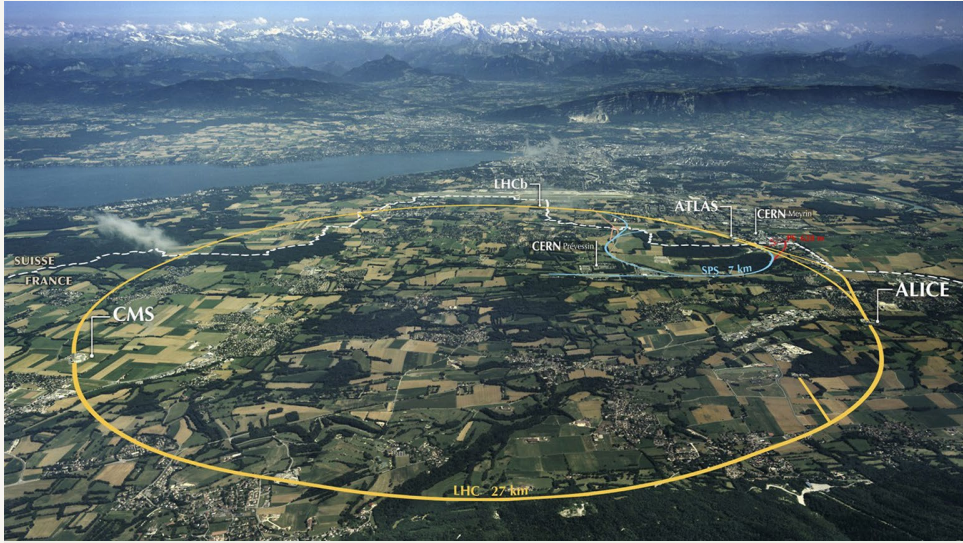
Calvin and Hobbes, Bill Watterson

In high energy physics, simulations are a tool to make experiments more sensitive

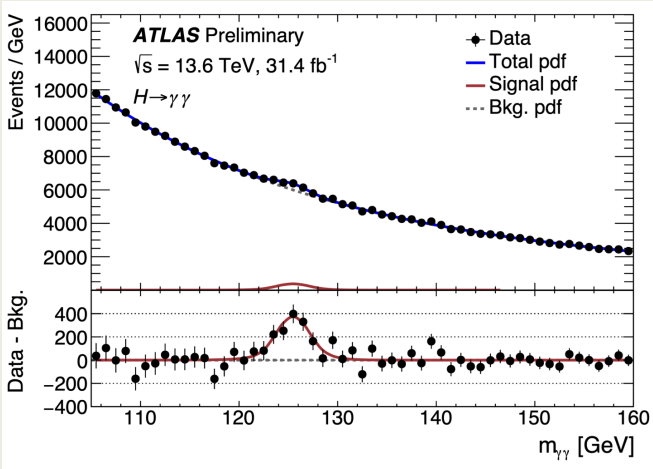
The complexity of the experiment requires modeling to understand it

- → Backgrounds are enormous
- → The measurement instruments are insanely complex
- → Systematic errors are critical to understand

Simulations are a tool to get the most out of expensive experiments (~$1B / year)



CERN



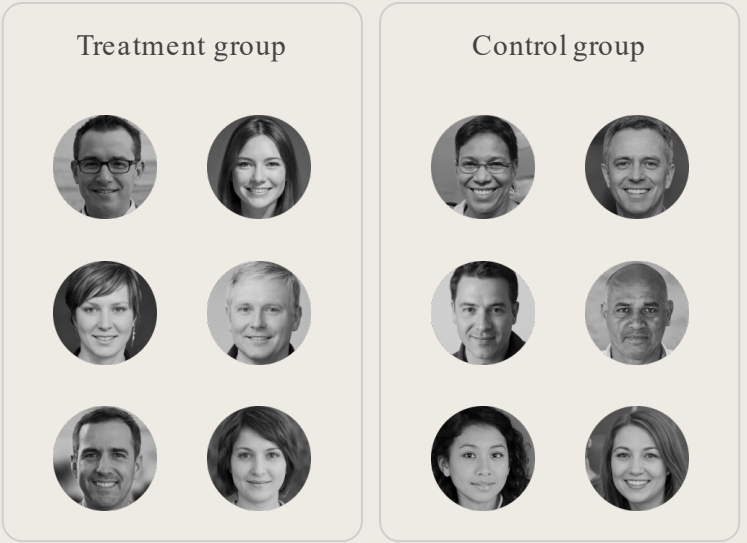
ATLAS

Clinical trials are a vastly more complex and expensive set of experiments

Clinical trials establish the safety and efficacy of drugs through experiments on people

- → Phase 1 trials test safety and mechanism of action for drug
- → Phase 2 trials trials establish safety and early efficacy signals
- → Phase 3 trials demonstrate regulatory-grade safety and efficacy

Total clinical trial industry expenditures are **~$80B / year**

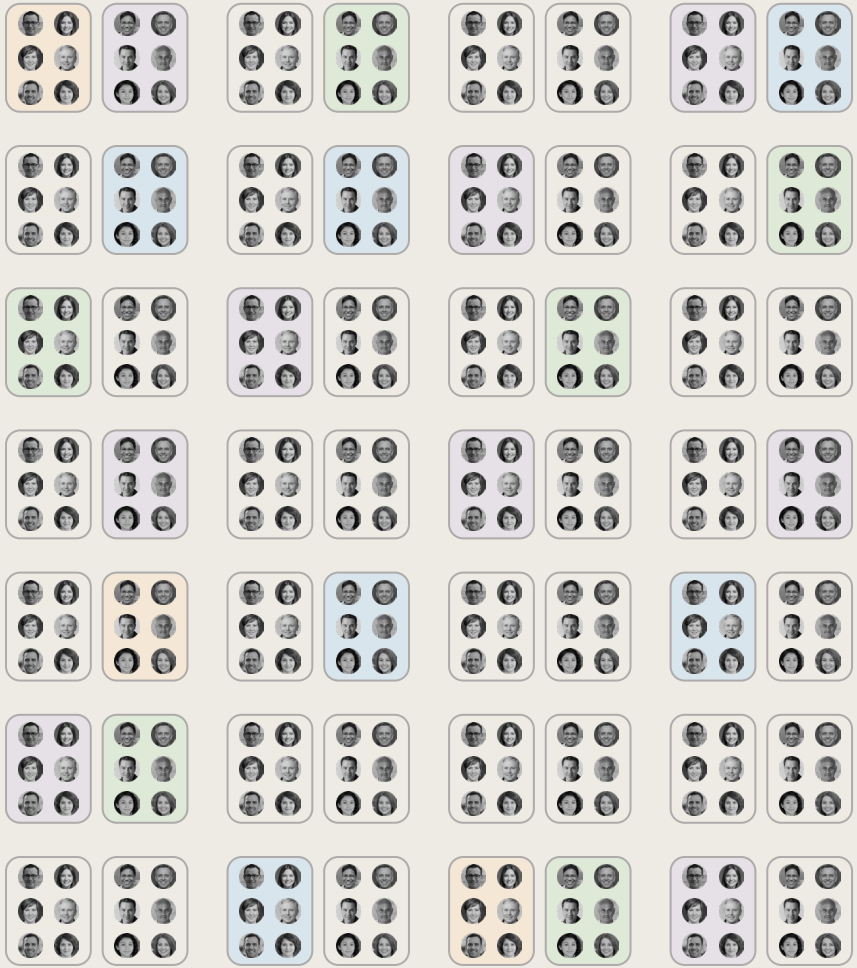


Why run randomized controlled trials (RCTs)?

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 does the field 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.

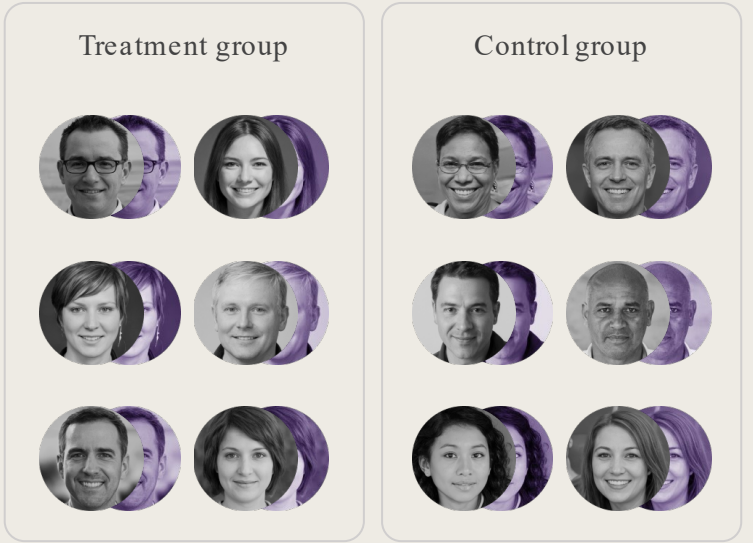
One approach is to maximize the utility of data in RCTs. Digital twins are a great tool to do this.

Digital twins are*not* new participants

They are *more information* about the participants in the study. They make every participant's data more useful.

We can use digital twins to bring the learnings from past clinical trials and observational studies into a new clinical trial.

The trick is to do this safely, so that we don't compromise the integrity of the RCT.



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 ALS

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

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

Time (months)	Baseline	1	2	3
				
Alanine Aminotransferase	41	41.8 ± 12.3	42.3 ± 13.9	41.7 ± 16.5
Albumin	42	42.2 ± 1.6	42.1 ± 2.1	42.1 ± 2.4
Alkaline Phosphatase	76	77.8 ± 11.0	77.6 ± 13.7	78.5 ± 17.5
ALSFRS Climbing	4	3.9 ± 0.4	3.7 ± 0.5	3.6 ± 0.6
ALSFRS Cutting	4	3.9 ± 0.2	3.9 ± 0.4	3.8 ± 0.5
ALSFRS Dyspnea	3	2.9 ± 0.7	2.9 ± 0.8	2.9 ± 0.9
ALSFRS Handwriting	4	3.9 ± 0.3	3.8 ± 0.4	3.7 ± 0.5
ALSFRS Hygiene	4	3.9 ± 0.3	3.8 ± 0.5	3.7 ± 0.6
ALSFRS Insufficiency	4	4.0 ± 0.1	3.9 ± 0.2	3.9 ± 0.3
ALSFRS Orthopnea	4	4.0 ± 0.2	3.9 ± 0.3	3.8 ± 0.4
ALSFRS Salivation	3	2.8 ± 0.7	2.8 ± 0.8	2.7 ± 0.9
ALSFRS Speech	2	1.9 ± 0.6	1.9 ± 0.8	1.8 ± 0.9
ALSFRS Swallowing	3	2.9 ± 0.6	2.8 ± 0.8	2.8 ± 0.9
ALSFRS Turning	4	3.9 ± 0.3	3.8 ± 0.4	3.8 ± 0.4
ALSFRS Walking	4	3.9 ± 0.3	3.8 ± 0.4	3.8 ± 0.5
Aspartate Aminotransferase	23	24.0 ± 5.2	24.3 ± 7.0	25.2 ± 8.4
Basophils	0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0

Think of this setup as a basic modeling problem

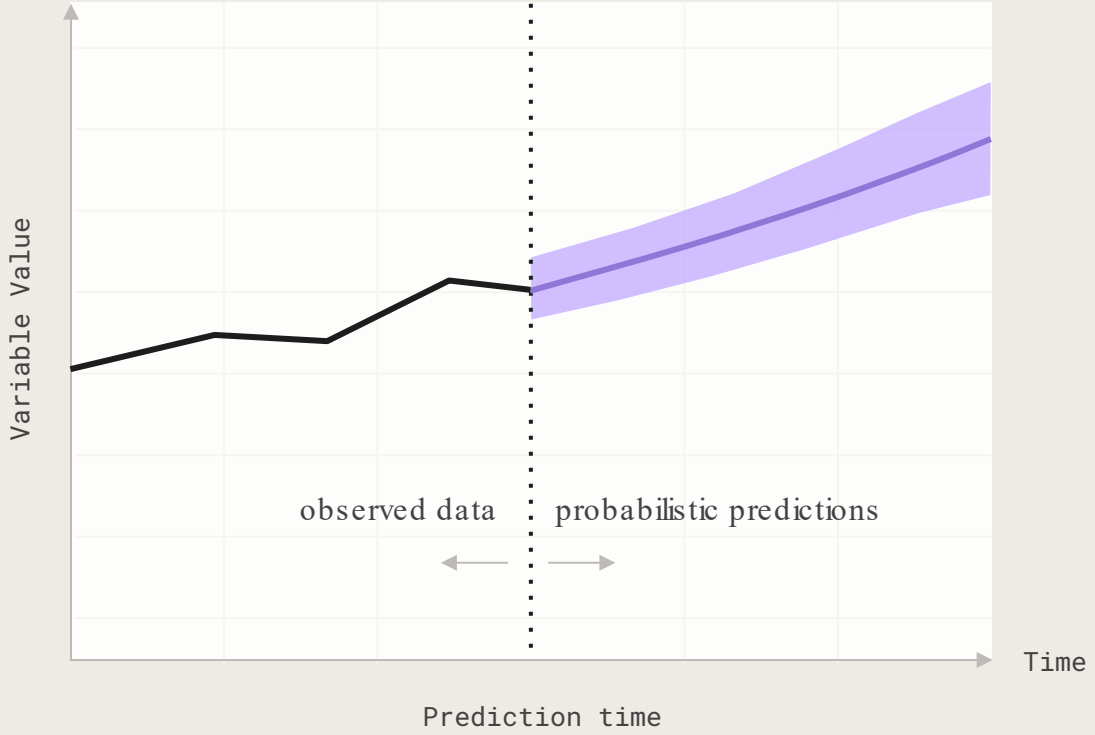
Data about a patient's health

A probabilistic model of longitudinal health data for patients with the disease

Comprehensive, probabilistic predictions of future outcomes

Think of this setup as a basic modeling problem

writ large over many variables



We want to learn the joint distribution over a dataset

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

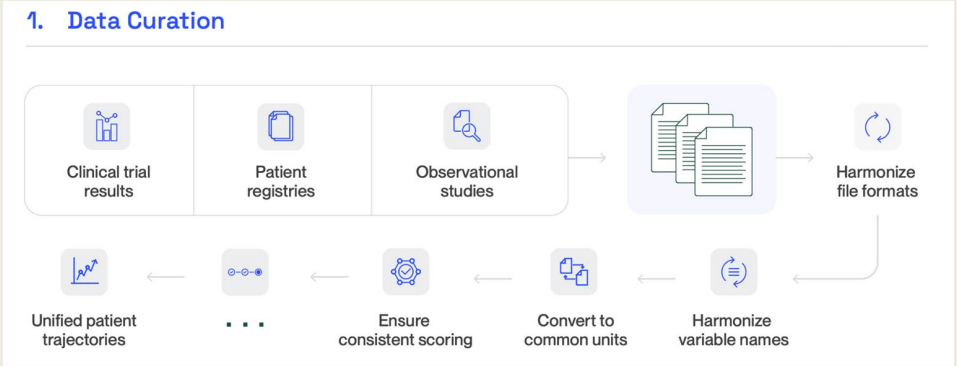
Conditioning on a set of data from a patient yields a model of their future health – their digital twin

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

Processes for building models: data

Data curation is critical to build high quality resources for modeling

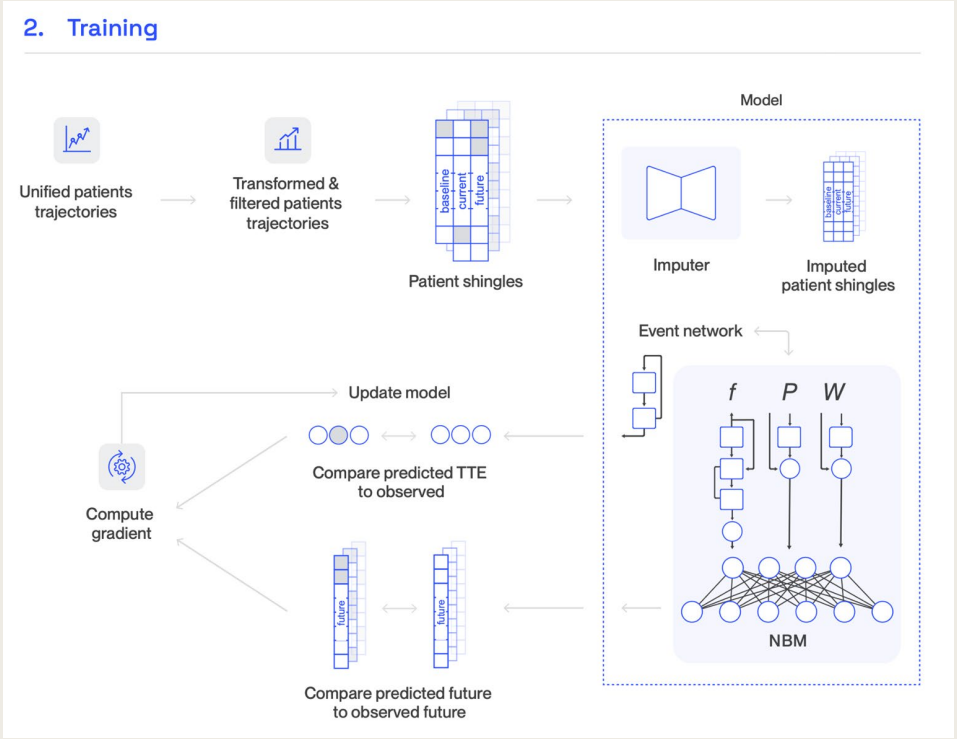
- → Harmonization across datasets (trials, observational studies, registries)
- → Quality control to curate data suitable for creating models of disease
- → This step tends to be the most resource intensive



Processes for building models: training

Architectures are designed to handle the realities of clinical data

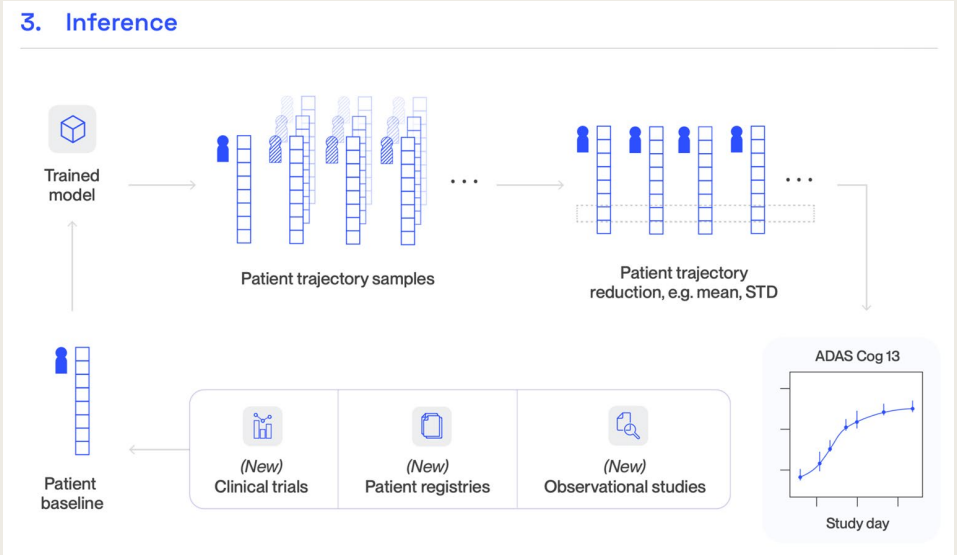
- → Models are Markovian, training divides trajectories into shingles
- → Optimization is for sample quality, with the goal that model generated data matches observed data
- → Validation along many dimensions of performance



Processes for building models: inference

Prediction follows a Markov process

- → Sample-level trajectories are created from the model
- → Mean and variance moments give the expectation and uncertainty for individual patient predictions
- → Generalization to new studies is a key dimension of performance



We have invested a fair amount of effort in the core machine learning problem

↗ Generative models for joint distribution

↗ Tabular time series datasets

↗ Multimodal

↗ Missingness is common

↗ Not large (~1000 patients, ~10 visits)

↓

Architectures for this class of problem

Alam et al, "*Digital Twin Generators for Disease Modeling*", arXiv:2405.01488

Digital twins can be incorporated in clinical trials via covariate adjustment

Example:

- → An endpoint is the change in a cognitive test score over the trial duration
- → The predicted endpoint value is extracted from the digital twin for each participant, to be used in covariate adjustment
- → This is the expected standard of care outcome and does not require modeling the novel treatment

> *"Covariate adjustment leads to efficiency gains when the covariates are prognostic for the outcome of interest in the trial. Therefore, FDA recommends that sponsors adjust for covariates that are anticipated to be most strongly associated with the outcome of interest."*
>
> – FDA guidance on covariate adjustment

Covariate adjustment is commonly used to increase power and reduce variability in clinical trial analyses.

outcome = intercept + slope x **(baseline covariate)** + (treatment effect) x (treatment indicator)

Forecast control outcomes from participants' digital twins are a "super covariate" that provides the maximum increase in power[1].

outcome = intercept + slope x **(forecast control outcome)** + (treatment effect) x (treatment indicator)

U.S. Patents: Systems and Methods for Supplementing Data with Generative Models (US20210057108A1) 2023 Status: *Pending*

[1] 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.
[2] Reference to Bayesian PROCOVA 2.0 arXiv paper to go here

EMA and FDA support the use of digital twins to run smaller, faster and highly powered RCTs*

→ EMA qualified our methodologies

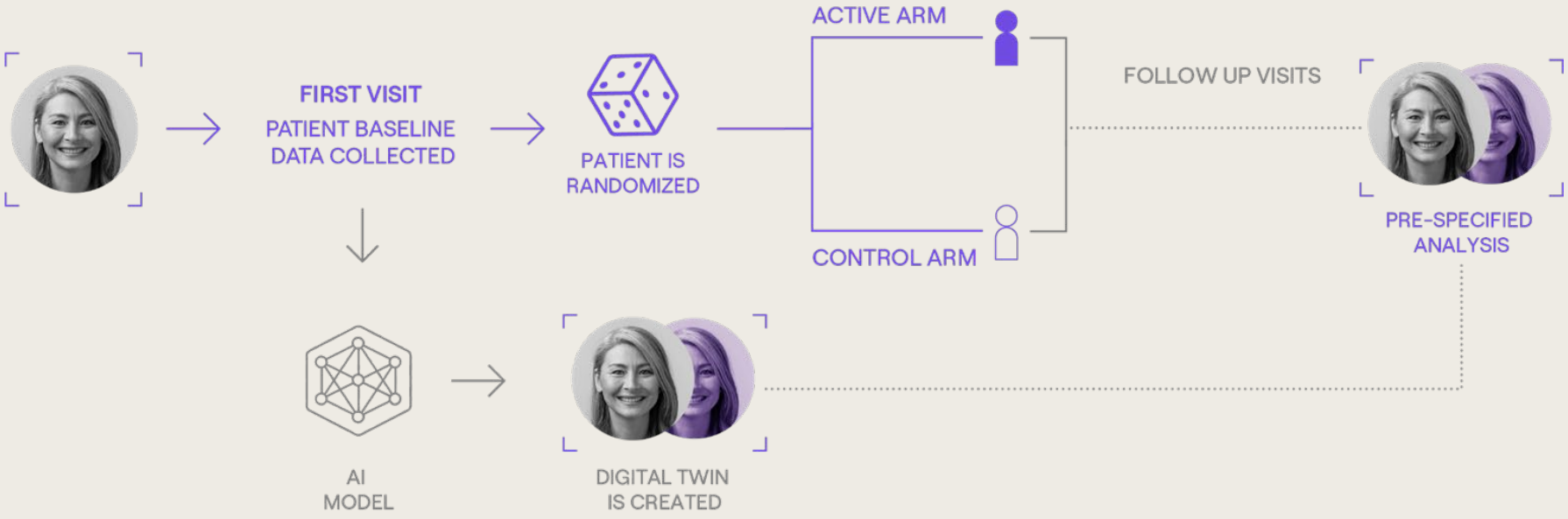
→ FDA concurred with EMA's qualification

→ Proven successful Type C meetings in Neuroscience

→ Regulatory policy is a consistent area of focus as the field advances in the use of AI

*Unlearn's EMA Qualification Opinion. FDA's Concurring Statements

Operationally including *digital twins is simple*



Case study: Alzheimer's disease

Retrospective analysis of a phase 2 trial of tilavonemab in early Alzheimer's disease (AD)

Specification of the AI model

Generation of digital twins

Reanalysis of the tilavonemab study

Planning for future AD studies

Key content taken from a poster from the Alzheimer's Association International Conference (AAIC) 2024, jointly presented with AbbVie

Summary tilavonemab phase 2 trial

- → Phase 2 randomized trial
- → 453 participants
- → Conducted from 2017-2021

Patients included in AWARE were randomized to receive either placebo or one of three doses (300, 1000, and 2000 mg) of tilavonemab in a 1:1:1:1 ratio, over a 96-week treatment period

CDR-SB, ADAS-Cog 14, FAQ, and MMSE scores were key outcomes in the trial

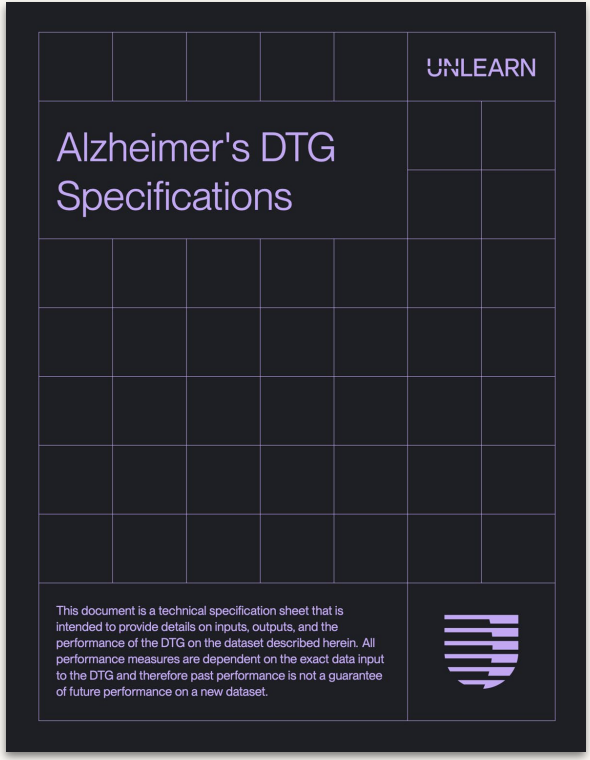
Patients who met the following criteria were eligible for AWARE

- ☑ Aged 55–85 years
- ☑ Diagnosed with early AD (met NIA-AA clinical criteria for mild cognitive impairment or probable AD)
- ☑ CDR global score of 0.5 at screening
- ☑ MMSE score of 22 to 30 at screening
- ☑ RBANS-DMI score ≤85
- ☑ Positive amyloid PET scan
- ☑ Modified Hachinski Ischemic Scale score ≤4

Specification of the AI model

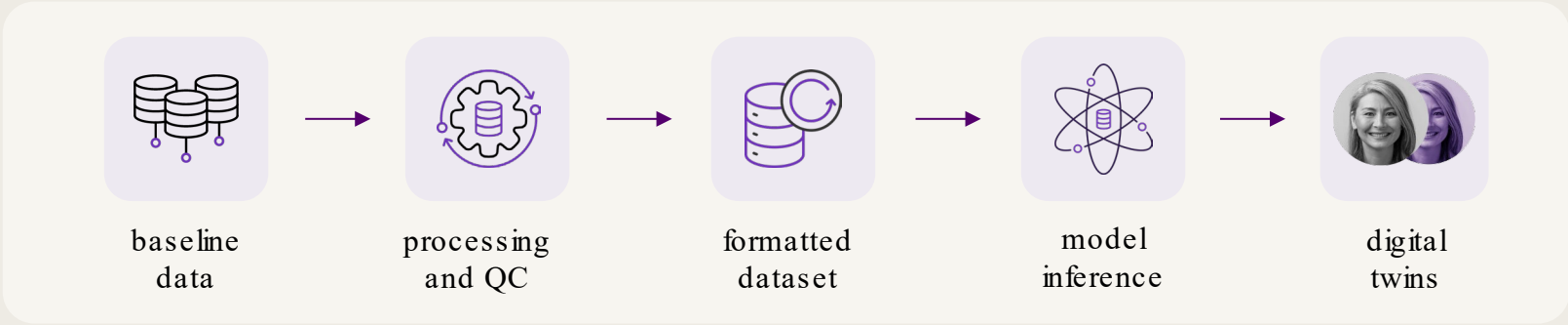
Unlearn's Alzheimer's disease Digital Twin Generator (DTG) model:

- → Trained on data from more than 8500* participants, from the control arms of 29 clinical trials and data from 4 observational studies from early MC to severe AD
- → Approximately 60 input variables (demographics, biomarkers, baseline disease severity measures, labs and vitals)
- → Outputs include item-level predictions of multiple composite scores, biomarkers, labs and vitals



Our model is regularly updated to include new endpoints and new data (now over 25k patients)

Generation of digital twins



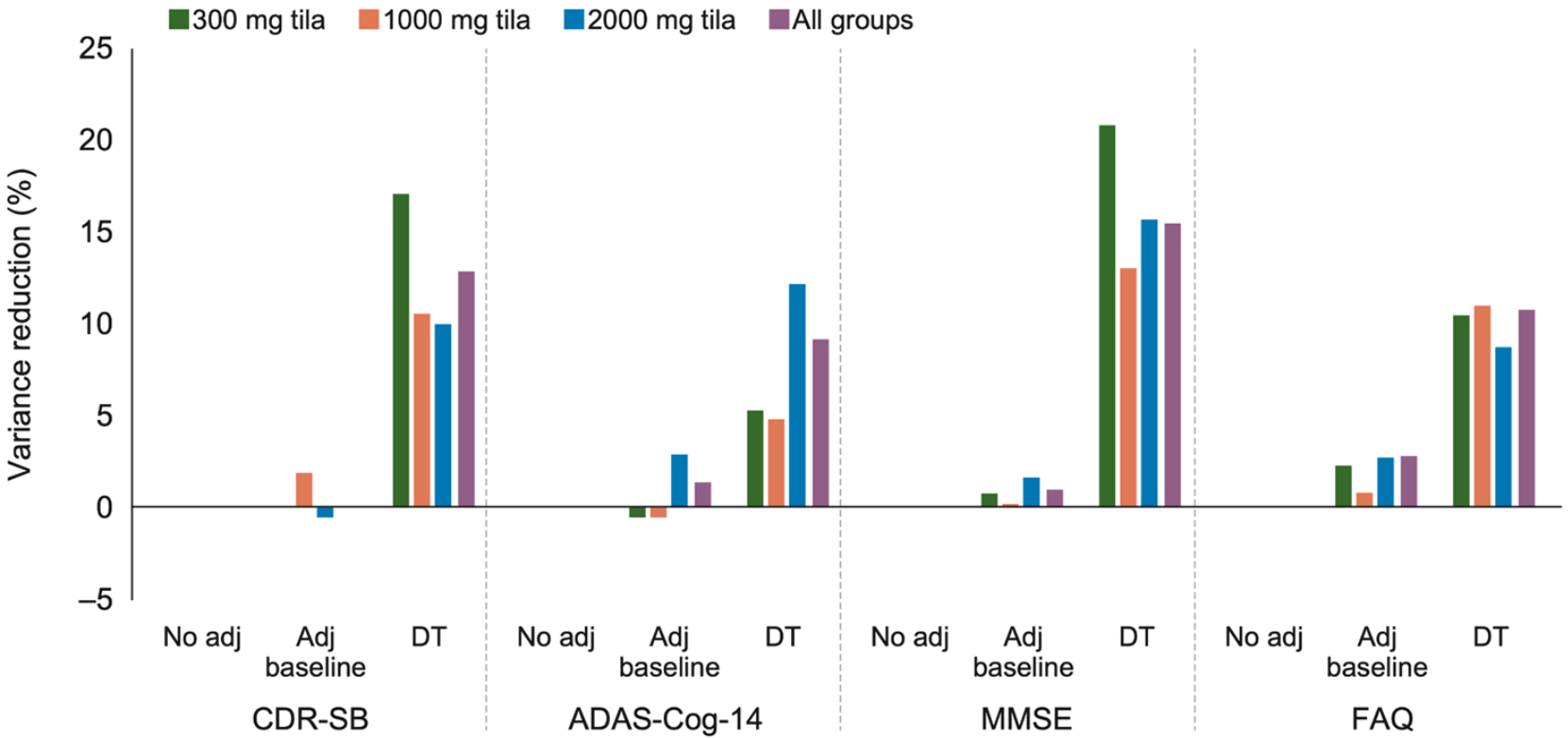
Baseline data provided in a loose standard format for trial data.

Processing and QC steps ensure the baseline data is formatted correctly for model inputs.

Inference is done via Monte Carlo sampling, to produce digital twins.

Moments (mean, variance) are provided for analysis.

Results from reanalysis



Interpreting reanalysis results

We can directly relate variance decrease to future study design. The prognostic value of the covariate can be used for:

Power add

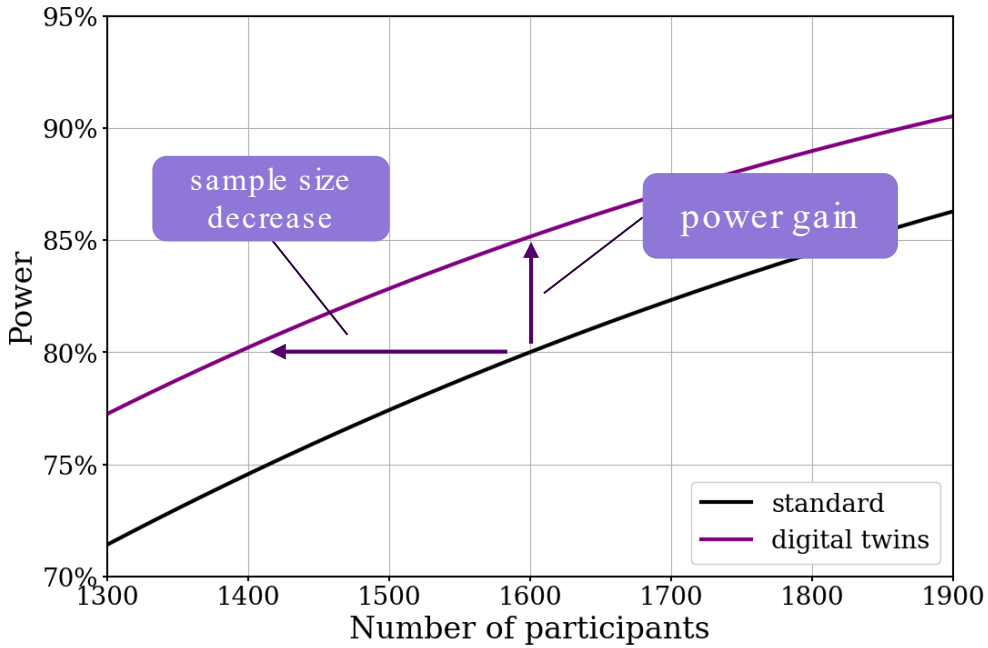
→ Digital twins added at analysis time

→ 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

**Assuming the variance reduction in the primary will be 90% of observed (13%)*

	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)

UNLEARN

Thank you