Machine learning generated digital twins as an external control in non-randomized trials

Introduction

- → Non-randomized trials are common in neurodegenerative disease development, particularly in ALS and HD trials.
- → The absence of a concurrent control group complicates interpretation of efficacy signals.
- → External controls derived from historical data are frequently used to address this challenge.
- → Machine learning enables individualized prediction of disease progression ("digital twins") as an alternative to propensity methods.

Methods: External Control Arms

Historical Data Sources

Amyotrophic Lateral Sclerosis (ALS)

13,500+ Patients | >30 RCTs | PRO-ACT, NEALS, PROACE, ANSWERALS, APST

Huntington's Disease (HD)

32,000+ Patients | 4 RCTs | CHDI: ENROLL-HD, PREDICT, TRACK, CRESTE

Propensity Score Methods

Calculate Propensity Score

ALS Covariates
- Age
- Sex
- Diagnosis
- ALSFRS-R/FVC Preslope
- Onset location
- Baseline ALSFRS-R/FVC

HD Covariates
- Age
- Sex
- CAG Repeats
- CAP Score
- cUHDRS (TMS, TFC, SDMT, SWRT)
- PIN Score, HD ISS Stage

Propensity Score Methods

Inverse Probability Weighting (IPW)
Weights all historical data based on PS

Propensity Score Matching (PSM)
1:1 matching to historical data

Purpose
Propensity score methods attempt to construct a similar population in the historical data to the trial

Propensity Model
- → Propensity scores estimated via logistic regression (5-fold CV)
- → The propensity score represents the probability of trial assignment
- → Among the selected covariates, average standardized mean differences after weighting were <0.1

VS

Digital Twin Methods

ML-Generated Digital Twins
- → ML model trained on historical untreated data using a non-parametric random forest that learns slopes of progression
- → Features: HD > 25, ALS > 30
- → Digital twins generated for each participant of the target trial conditional on their unique baseline measures

1:1 — Trial Participant — Digital Twin

DT Methods

Counterfactual Prediction (Plug-in)
Direct comparator to the treated

Augmented Inverse Probability Weighting (AIPW)
Counterfactual prediction with model correction

Purpose
Digital twins predict disease progression unique to each trial participant based on their baseline measures

Machine Learning Model
- → Model trained using nested CV for hyperparameter selection
- → Digital twin generated for each participant; exact match based on baseline features
- → Digital twins used for direct counterfactual prediction and AIPW estimation

Target Trials

PS Methods | DT Methods

ALS: Celecoxib Trial — N = 300

Primary Endpoint: ALSFRS-R at 12 months

I/E Criteria: Age ≥18; Disease Onset <5 years; Vital Capacity >60%; Stable riluzole

Outcomes: ALSFRS-R; FVC

HD: 2CARE Trial (Coenzyme Q10) — N = 609

Primary Endpoint: TFC at 48 months

I/E Criteria: Age ≥16; TFC ≥9; CAG repeat ≥36

Outcomes: TFC; FA; IS; TMS

Primary Evaluation Metrics

Treatment Effect Bias
Accuracy of estimated treatment effect

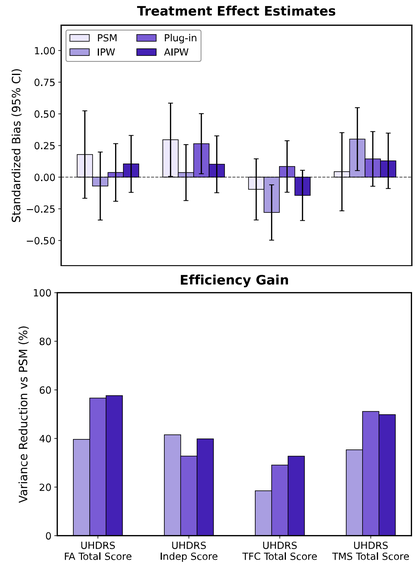
Standard Error
Precision and variability of treatment effect estimates

Statistical Power
Powers for propensity score and digital twin methods

Results

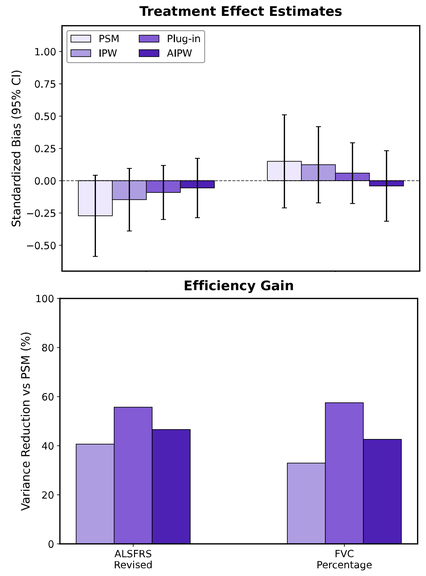
2CARE Coenzyme Q10 in HD

- → Digital twin based estimators (plug-in and AIPW) produced lower standardized bias (bias divided by SD) across some UHDRS endpoints compared with PSM/IPW.
- → Digital twin methods demonstrated efficiency gains, with ~30–60% variance reduction relative to PSM across endpoints.
- → A standardized bias of 0.25 indicates that the bias is ¼ of the standard deviation relative to the outcome.



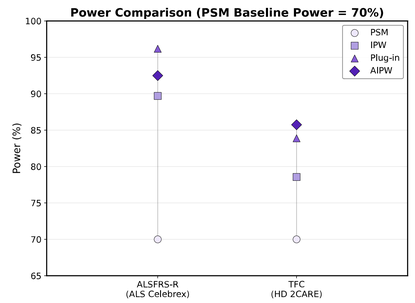
Celecoxib in ALS

- → AIPW digital twin based estimators demonstrate close to zero bias, while the digital twin plug-in estimate also exhibit low bias.
- → Across endpoints, plug-in and AIPW estimators outperformed PS methods, with AIPW providing the most consistent balance of low bias and high efficiency.



Statistical Power

- → Digital twin methods increased power up to 96% compared to the reference (70% for PSM)
- → Digital twins accounts for the correlation structure between trial data and predictions, increasing statistical power.



Conclusions

Machine learning generated digital twins represent a promising external control methodology for non-randomized trials, offering improved efficiency and reduced bias to establish causal relationships.

Strengths of Digital Twins

- → Machine learning models learns from all available data using the totality of the data to predict individualized, model-based counterfactual outcomes.
- → Potential to reduce bias and variability of treatment effect estimates compared to PS methods.
- → Digital twin predictions are not susceptible to covariate imbalance or difficulties with matching.
- → Machine learning models are validated and locked making them suitable for prespecified analyses following FDA's model credibility assessment.

Limitations

- → Model performance depends on the representativeness, quality, and completeness of historical training data.
- → Development of machine learning models are more complex than propensity score methods.
- → Valid causal inference for all four methods presented rely on important assumptions including no unmeasured confounders.

Authors

Run Zhuang, Daniele Bertolini, and Jonathan Walsh

Data Acknowledgements

The Pooled Resource Open-Access ALS Clinical Trials Consortium, NEALS, CHDI Foundation and ENROLL-HD

Disclosures

All authors are equity-holding employees of Unlearn.

Unlearn.AI, San Francisco, United States of America

UNLEARN