Biogen

ISCTM 2026 Scientific Meeting
Roland Brown, Biogen

PROGNOSTIC ADJUSTMENT & OTHER DIGITAL TWIN METHODS FOR IMPROVING TRIAL EFFICIENCY

DISCLOSURES

- Dr. Brown has received personal compensation as an employee of Biogen, Inc., and received stock or an ownership interest from Biogen

AGENDA

1 **What are Participant Digital Twins?**
How can they help improve trial efficiency?

2 **Example in Alzheimer's Disease**
How we developed an in-house prognostic score model for Alzheimer's

3 **Implications for Other Disease Areas & Higher Risk Approaches**

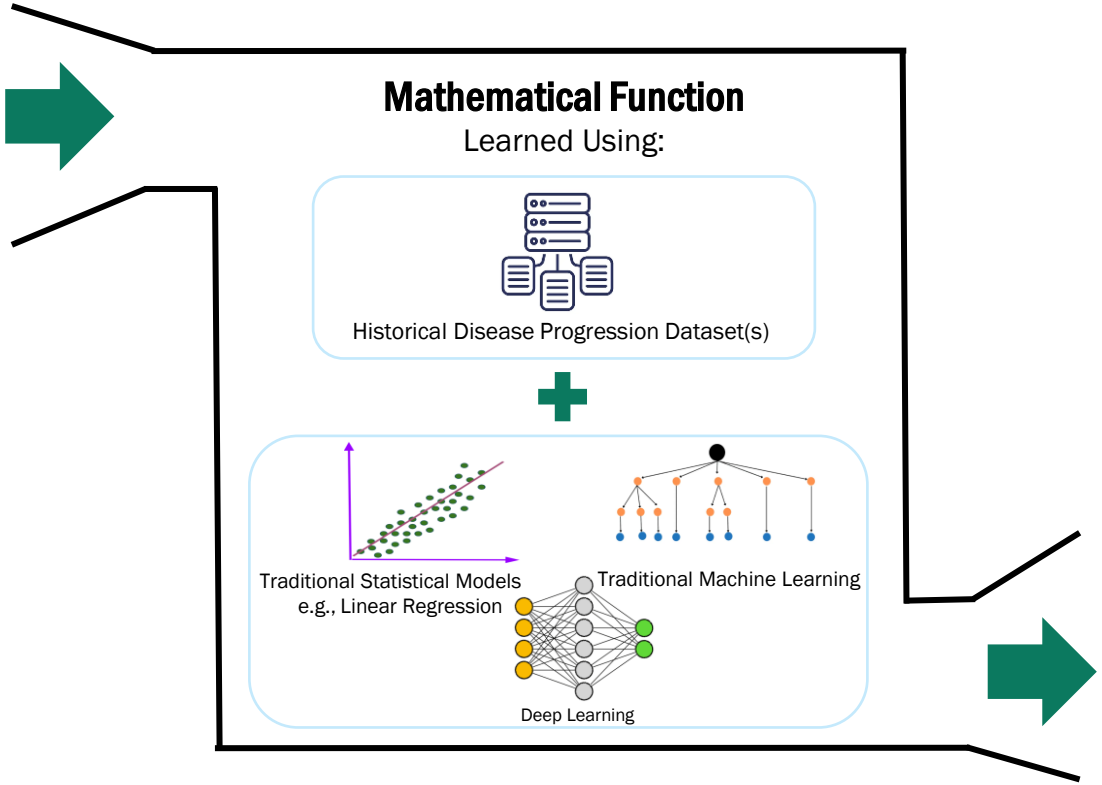
WHAT ARE PARTICIPANT DIGITAL TWINS?

A participant's digital twin (DT) is a function of their baseline characteristics that predicts their likely progression of disease.

Baseline Information

- Demographics
- Baseline Disease Status
- Biomarkers
- Imaging

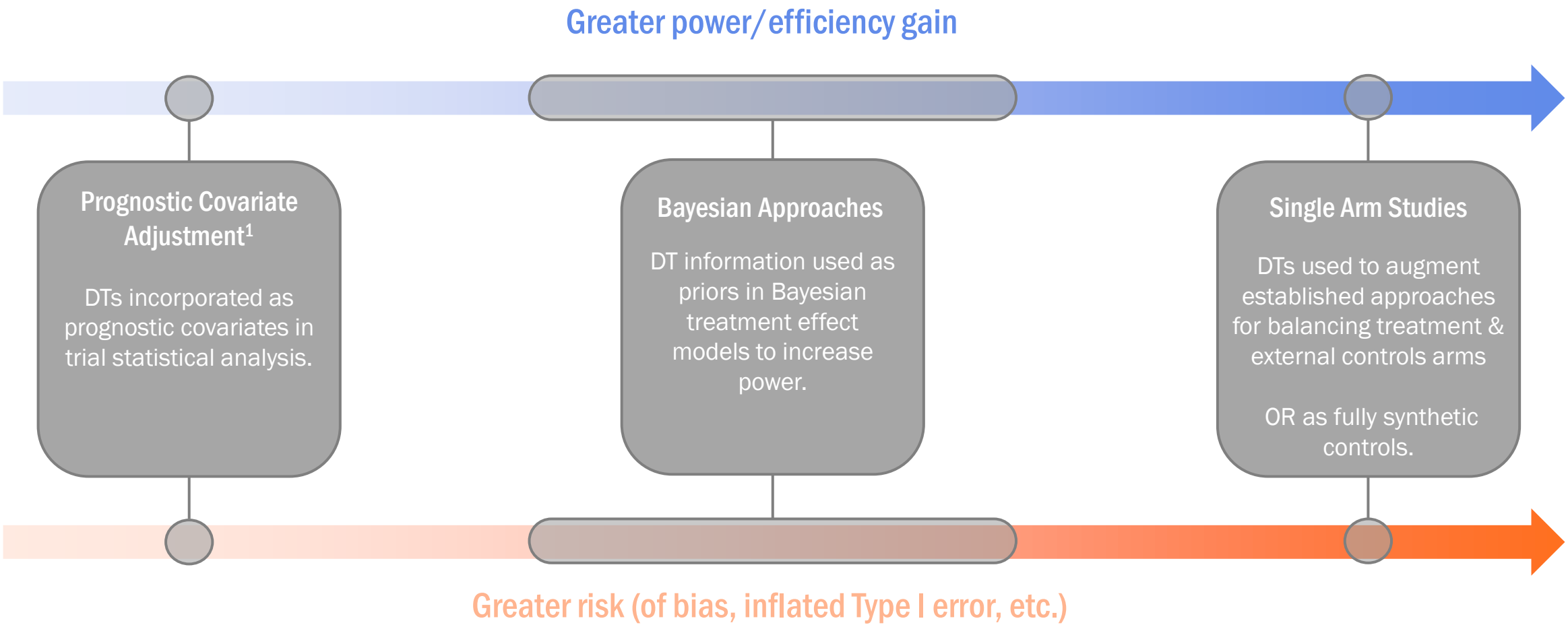
Anything else we can measure....



Digital Twin

- Prediction of future natural disease progression (in clinical measures, biomarkers, etc.)
- For RCT use, typically focused on key trial endpoints
- These digital twins are also called prognostic scores

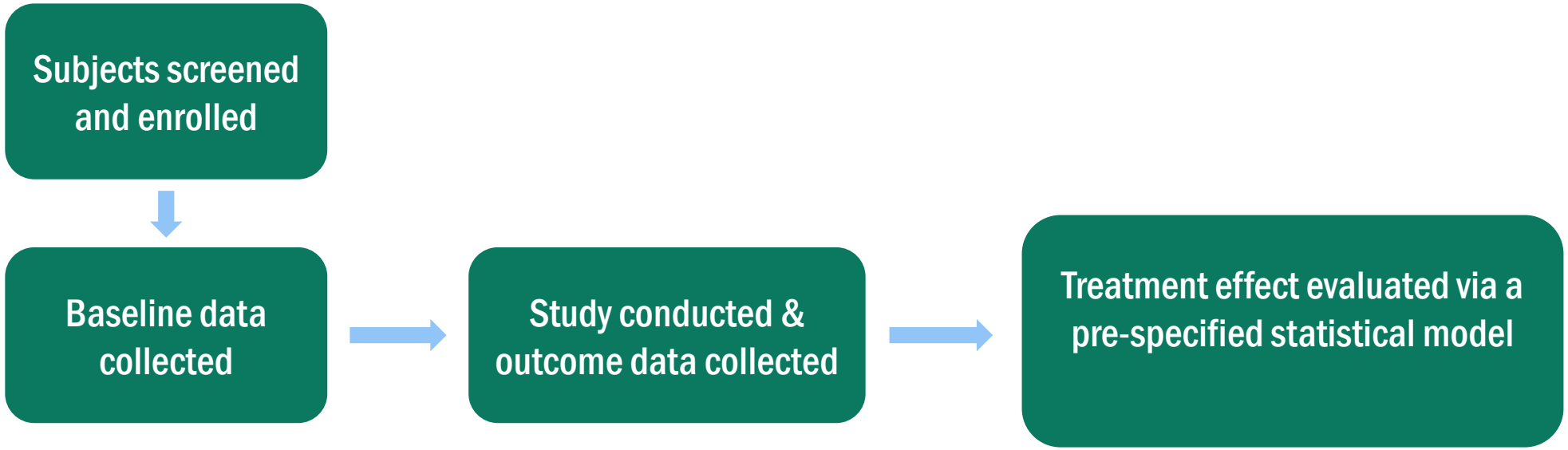
DIGITAL TWINS CAN ENABLE EFFICIENCY GAIN ACROSS A RISK-REWARD CONTINUUM



[1] Encouraged or accepted as a methodological approach in pivotal studies by major regulators (FDA, EMA).

HOW PROGNOSTIC SCORES ENHANCE TRIAL ANALYSES

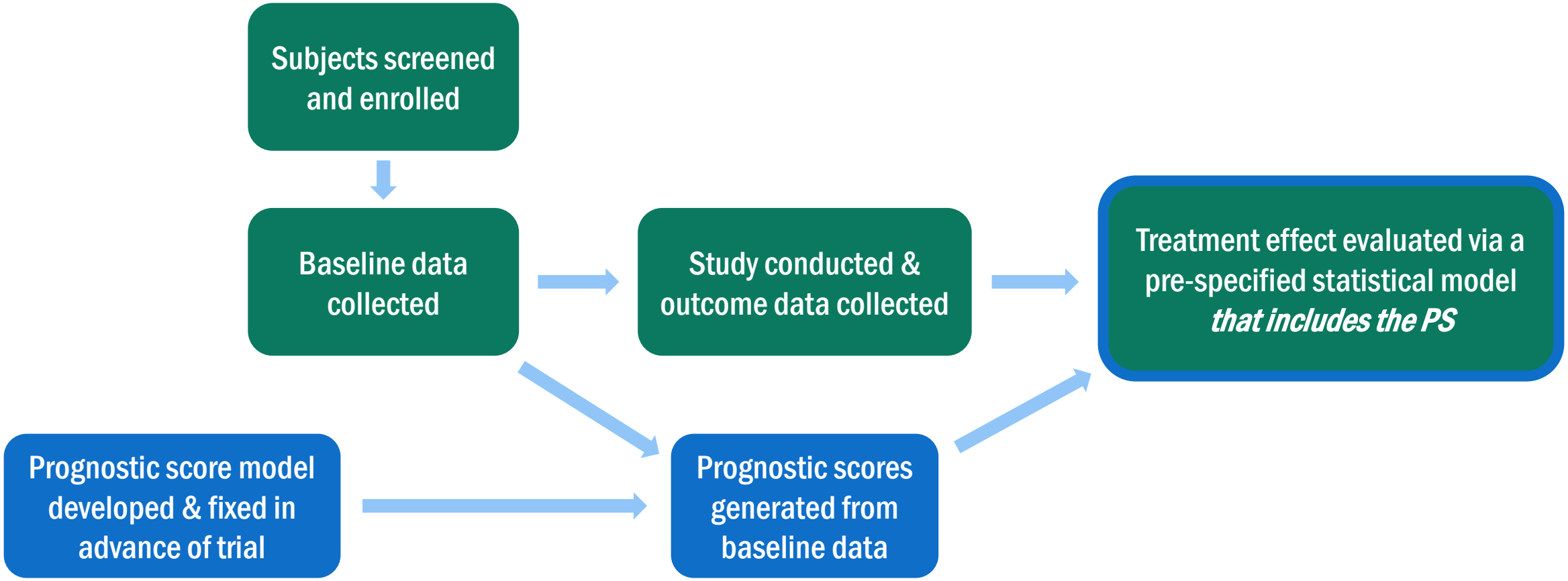
A typical clinical trial analysis



HOW PROGNOSTIC SCORES ENHANCE TRIAL ANALYSES

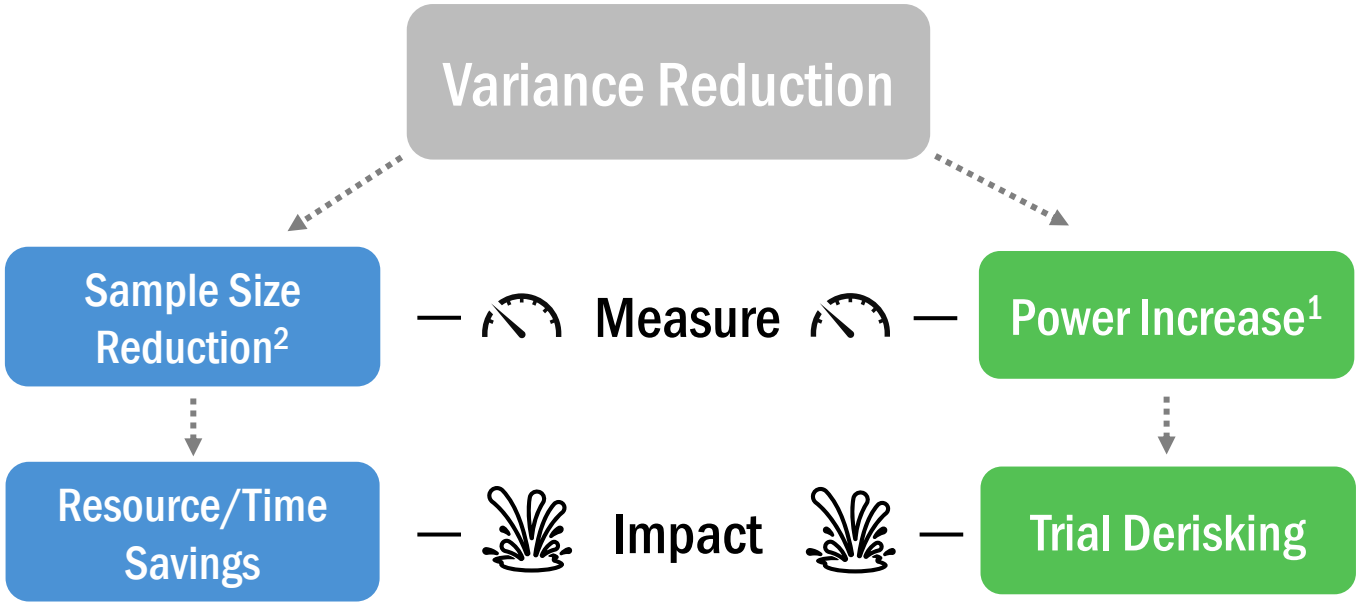
A clinical trial analysis with prognostic scores

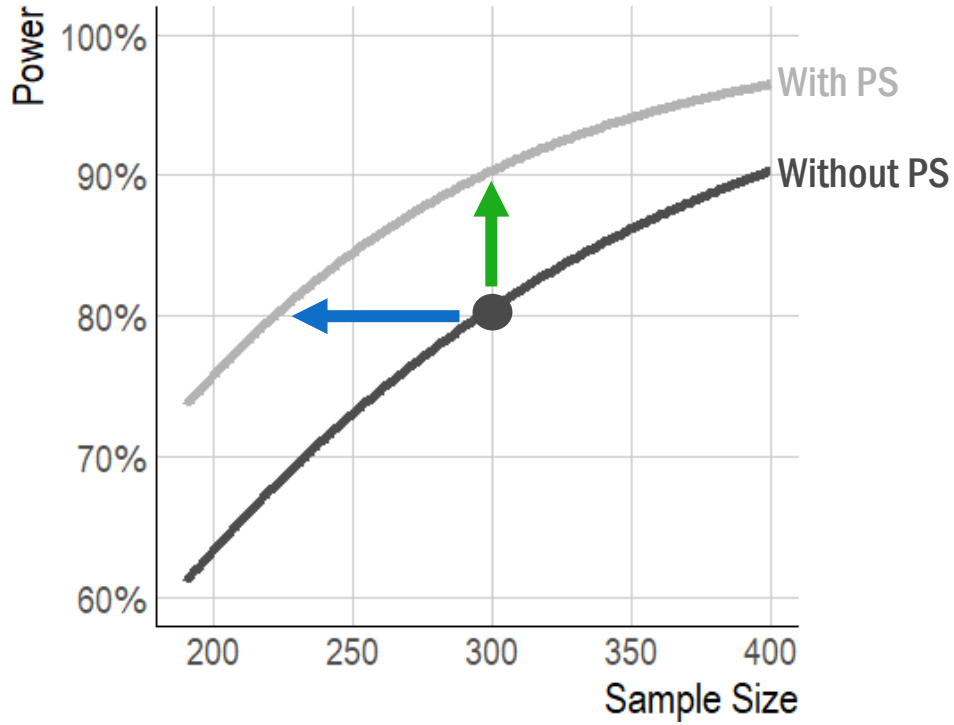
Prognostic scores enable increased study efficiency by reducing unexplained variance via ***covariate adjustment***



PROGNOSTIC COVARIATE ADJUSTMENT

Impact & Risks





Risks are minimal if DT is prespecified & only incorporates baseline information

- Power can be slightly reduced **IF** score has no prognostic value
- Risk of underpowered study **IF** planned sample size is reduced and score underperforms

[1] Power increase comes from keeping sample size/costs the same combined with variance reduction due to PS adjustment.
[2] Planned sample size is reduced while maintaining study power/probability of success.

PROJECT OVERVIEW

Project Goal

Build an Alzheimer's disease prognostic score model and evaluate its ability to increase clinical trial efficiency.

Targeted Use

Biogen Ph2 and Ph3 AD Clinical trials

Phase 1: Develop

Develop and implement training and validation of AD prognostic score model

Data Access & Harmonization

Combine historical Biogen trial data with real world datasets

Model Building & Validation

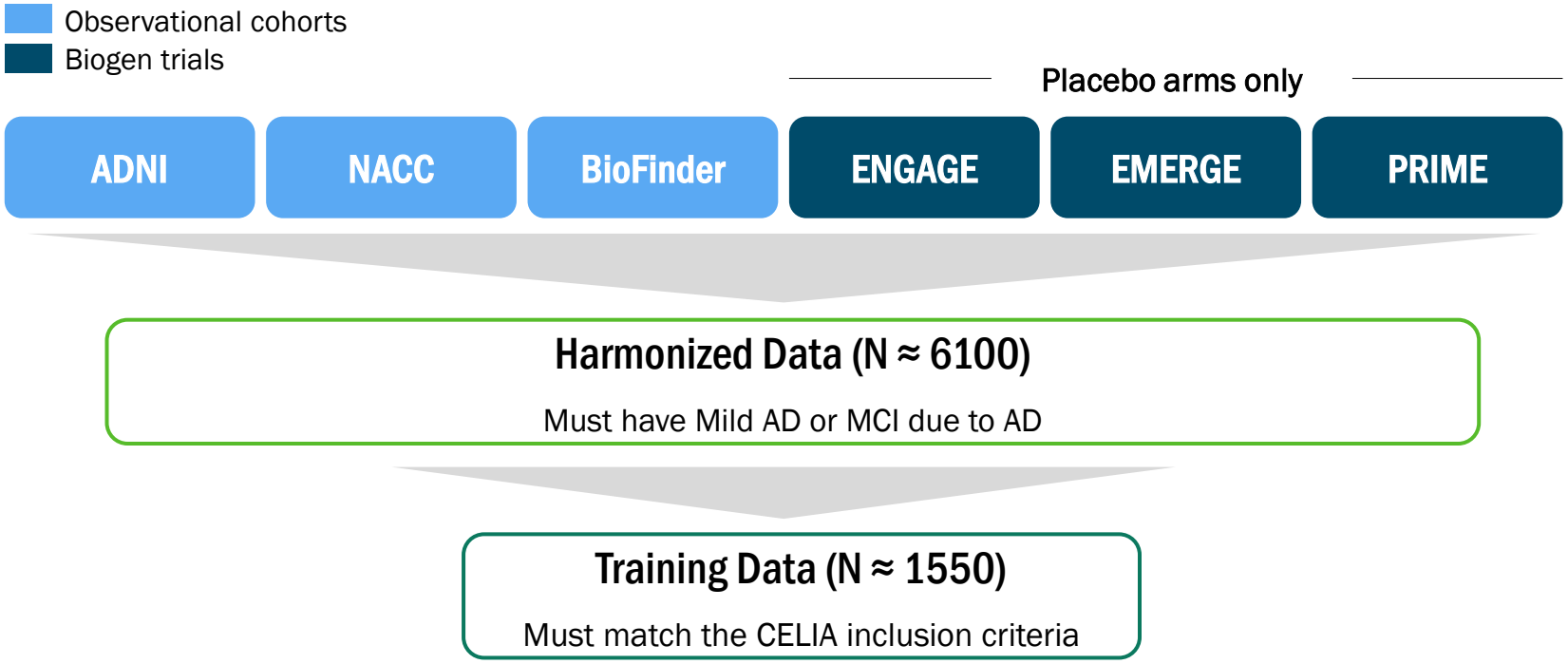
Build models (statistical or AI/ML) to predict likely disease progression

Phase 2: Evaluate

Evaluate effects of the PS on study power/sample size savings in held out trial

DATA SOURCES

Six data sources harmonized to train models with a held-out evaluation trial



Independent Evaluation

TANGO
N = 650

VARIABLES

Our goal was to leverage baseline information to predict month 18 CDR Sum of Boxes (CDR-SB) change

Outcomes

Primary outcome (focus)

- CDR-SB change from baseline at 18 months

Other outcomes

- ADAS-Cog change
- MMSE change
- Etc.

67 Predictors

"Standard" predictors (22)

- Amyloid +/- and APOE genotype
- Baseline clinical composites & subscores (e.g., CDR-SB, MMSE, ADAS-Cog)
- Baseline Dx (i.e., Mild AD vs MCI due to AD)
- Basic health (e.g., BMI, BP)
- Demographics (e.g., age, sex)
- Medical history (e.g., diabetes, hypertension)

Baseline clinical sub-scores (44)

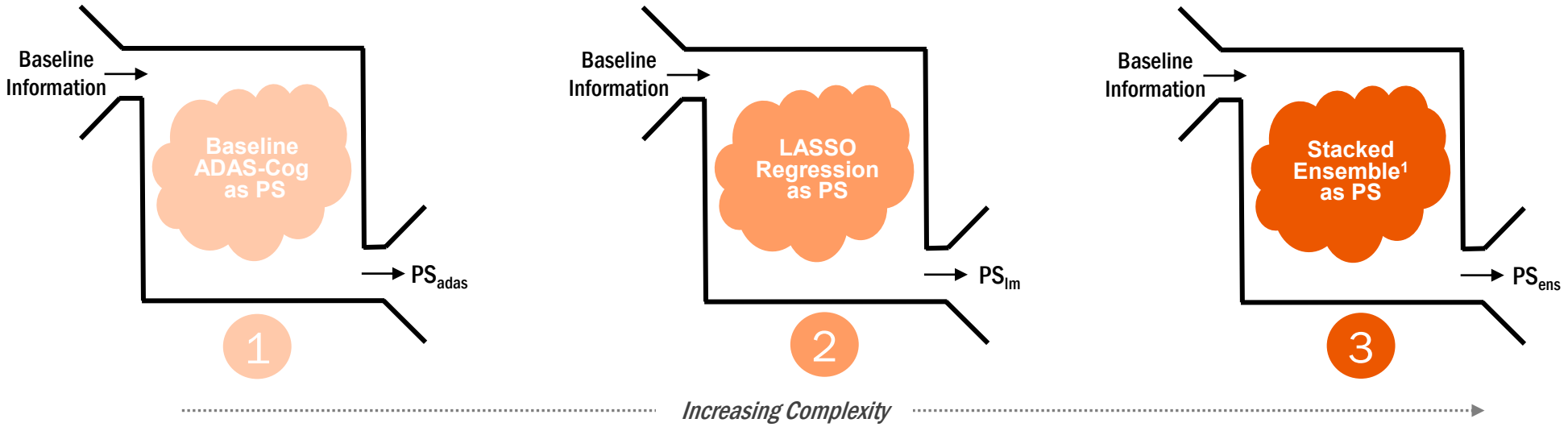
- CDR, MMSE, ADAS-Cog

MRI brain region volumes (12)

- E.g., Hippocampus volume

CANDIDATE PROGNOSTIC SCORES

For evaluation in TANGO



- Option 1 uses baseline ADAS-Cog as the prognostic score (which hasn't historically been adjusted for)
- Option 2 and Option 3 learn a relationship between baseline data and CDR-SB change via the harmonized dataset
- Candidates span a wide range of analytical/implementation complexity

[1] Ensemble stacks the following models: xgBoost, random forest, support vector regression, k-nearest neighbors, elastic net regression, ordinary linear regression

IMPACT OF DIGITAL TWIN COVARIATE ADJUSTMENT

The ensemble DT is most performant, but there are effective simpler options.

Independent evaluation in TANGO

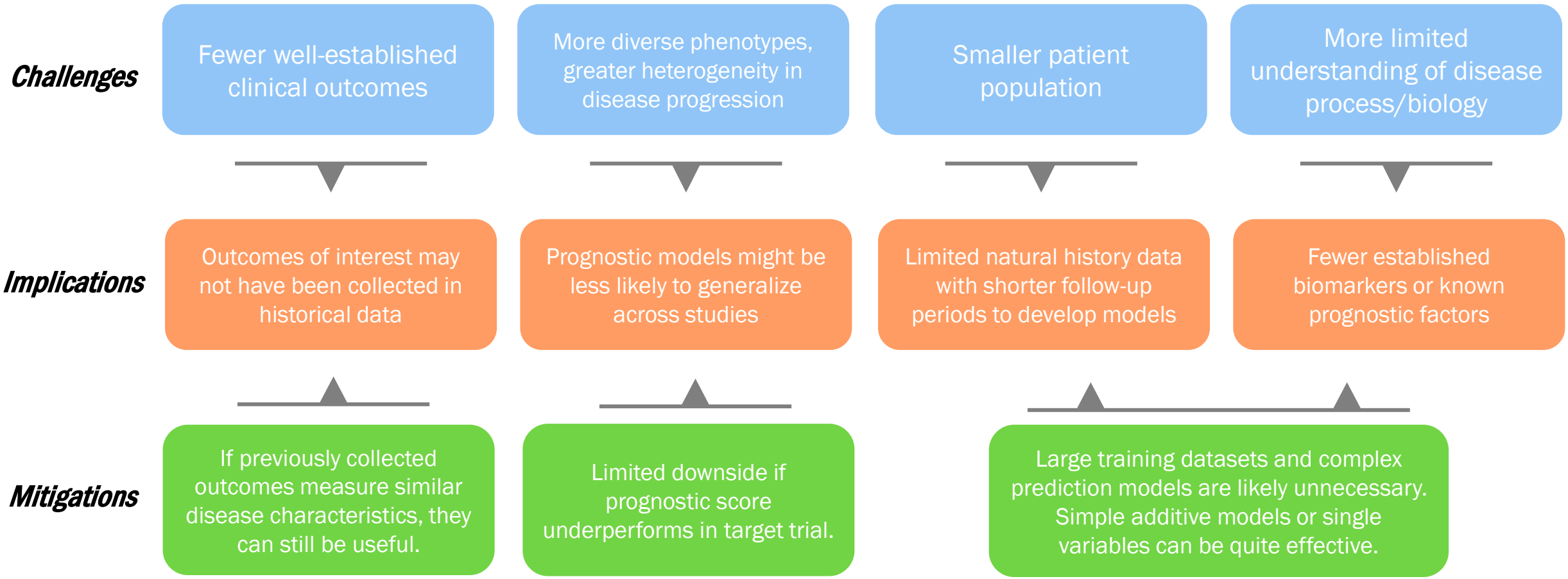
Measure[1]	1. ADAS-Cog	2. LASSO DT	3. Ensemble DT
Sample Size Reduction	14.2%	18.6%	19.7%
Power Change from 80%	+5.7%	+7.4%	+7.9%
Power Change from 90%	+3.8%	+4.9%	+5.1%

Increasing Complexity ▶

[1] All values reported here are over and above standard adjustment as specified in the TANGO primary analysis.

DIGITAL TWINS IN LESS COMMON DISEASES

Opportunity for increasing efficiency in resource-constrained environment, but with unique challenges



HIGHER RISK DIGITAL TWIN APPROACHES

	Same-Study Prognostic Covariate Adjustment	Bayesian Approaches	Single-Arm Studies
Overview	▪ Develop digital twin model and use it for covariate adjustment in the same study	Digital twins incorporated as prior information in the treatment effect model	Digital twins augment existing methods for external controls OR used as fully synthetic controls
Benefits	▪ Useful when little or no historical data available ▪ Statistical theory indicates approach can be used with similarly low risk of introducing bias/Type I error inflation	▪ Potential for greater efficiency gains than prognostic adjustment alone ▪ Risk-reward tradeoff can be calibrated by up-/down-weighting influence of prior digital twin information	Enable better comparisons to a treatment-naïve group where it would be otherwise infeasible/unethical ▪ Rare diseases ▪ LTE analyses where all participants move to active treatment
Risks	▪ Proper implementation requires very careful planning and pre-specification of analysis ▪ Difficult/impossible to predict degree of precision gain (or lack thereof)	▪ Greater efficiency gains come with greater risk of bias/Type I error inflation	▪ DT use cannot mitigate well-known risks of single-arm studies ▪ Using DTs as synthetic controls requires making potentially difficult-to-verify assumptions

TAKEAWAYS & LESSONS LEARNED

Prognostic adjustment is a low-risk use for digital twins

- Taking power gain without reducing sample size is essentially risk free
- Increased risk if planned sample size is reduced
- Degree of efficiency gain dependent on disease area (known prognostic factors, availability of historical data to train models, etc.)

Methods: bigger not always better

- LASSO regression was nearly as good as the best AI/ML model
- Large deep neural net model (not shown) underperformed less complex options
- Single prognostic variables can also meaningfully increase precision
- Simpler models offer interpretability, easier implementation, explainability to regulators & other stakeholders, possibility of development using much smaller datasets

Higher risk uses of digital twins

Greater potential upside, offset by higher risk

- Same-study digital twin/prognostic score development
- Bayesian approaches
- Single arm studies

Currently best for earlier phase trials, or for internal decision making, or where randomized placebo arm difficult or impossible