Biogen

NERDS 2025
Roland Brown, Biogen

ENHANCING CLINICAL TRIALS THROUGH PROGNOSTIC SCORE COVARIATE ADJUSTMENT

Opportunities & challenges in rare disease

AGENDA

1 **What are Prognostic Scores?**
How do they help us with trial efficiency?

2 **Increasing Trial Efficiency**
How we developed an in-house prognostic score model for Alzheimer's

3 **Prognostic Scores in Rare Disease**
Challenges & Opportunities

Biogen

CLINICAL DEVELOPMENT IS RESOURCE INTENSIVE

We should always be searching for any opportunity to increase trial efficiency & return on investment

50%

Clinical trial spending as share of total R&D spending, Pharma

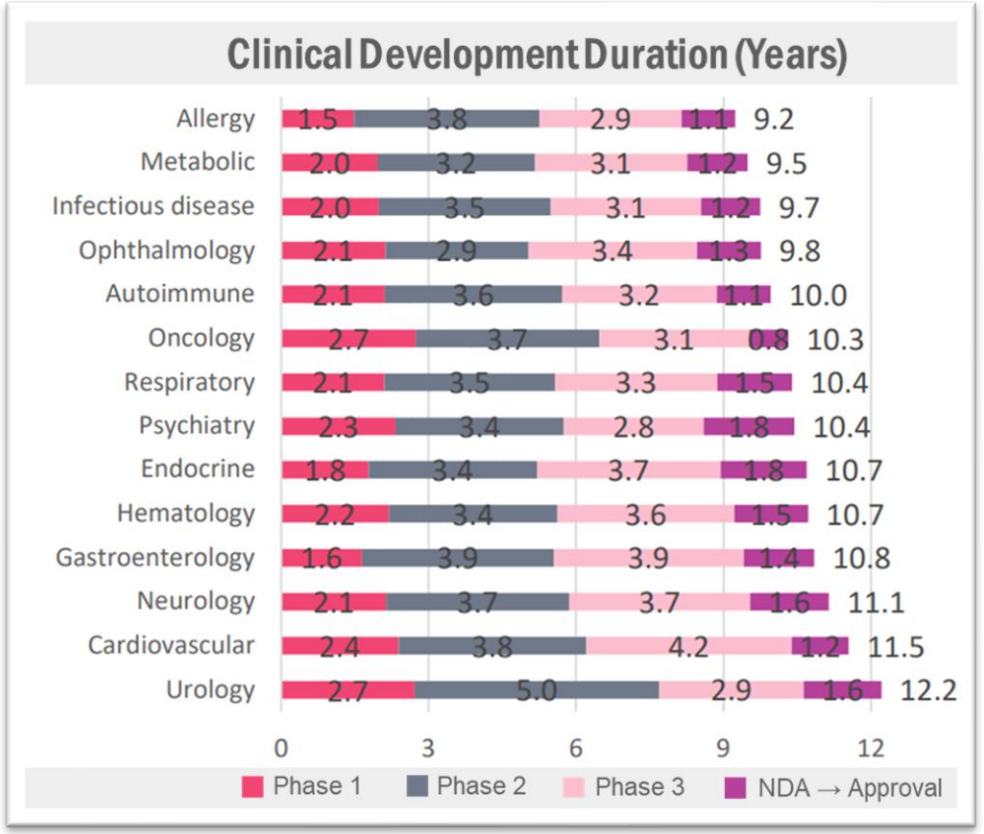
10%

Clinical trial spending as share of total revenue, Pharma

vs

4%

R&D spending as share of total revenue, SP500 Companies



Biogen

WHAT ARE PROGNOSTIC SCORES?

A participant's prognostic score (PS) is a function of their baseline characteristics that predicts their likely disease outcome.

Baseline Information

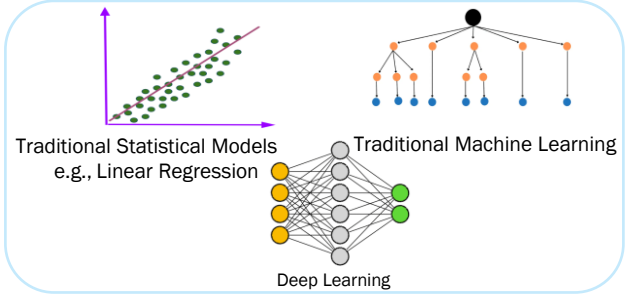
- Demographics
- Baseline Disease Status
- Biomarkers
- Imaging

Anything else we can measure....

Some Mathematical Function

Learned Using:

Historical Disease Progression Dataset(s)



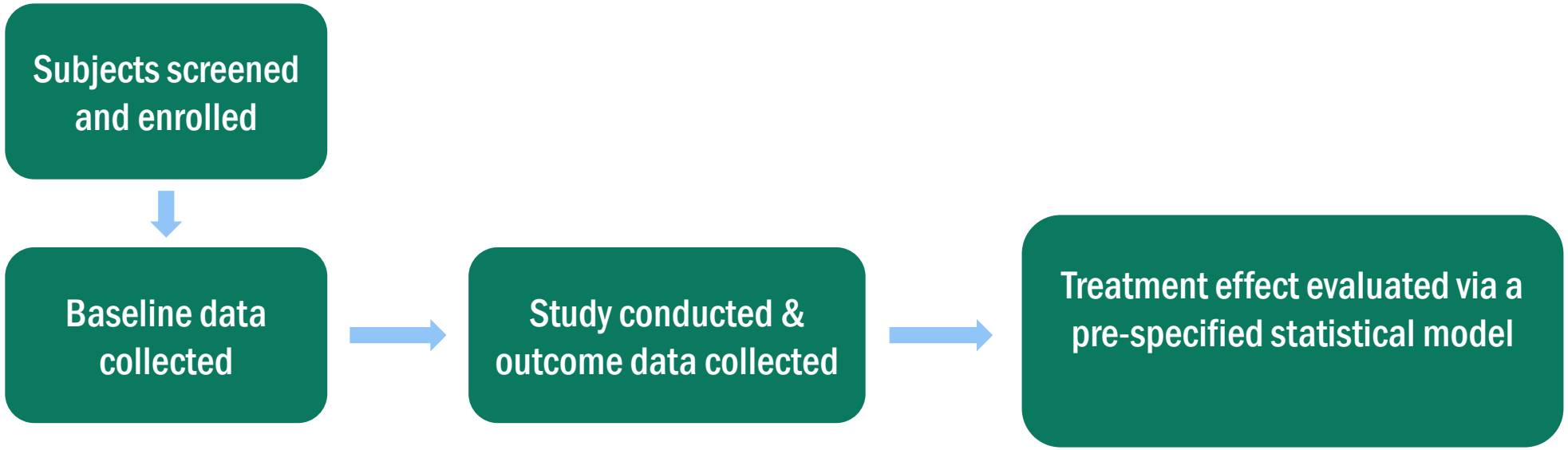
Prognostic Score

- A single quantity that effectively predicts future disease progression (assuming no treatment)
- For RCT use, typically targets a specific disease outcome
- Recently, prognostic scores have also been called digital twins

Biogen

HOW PROGNOSTIC SCORES ENHANCE TRIAL ANALYSES

A typical clinical trial analysis

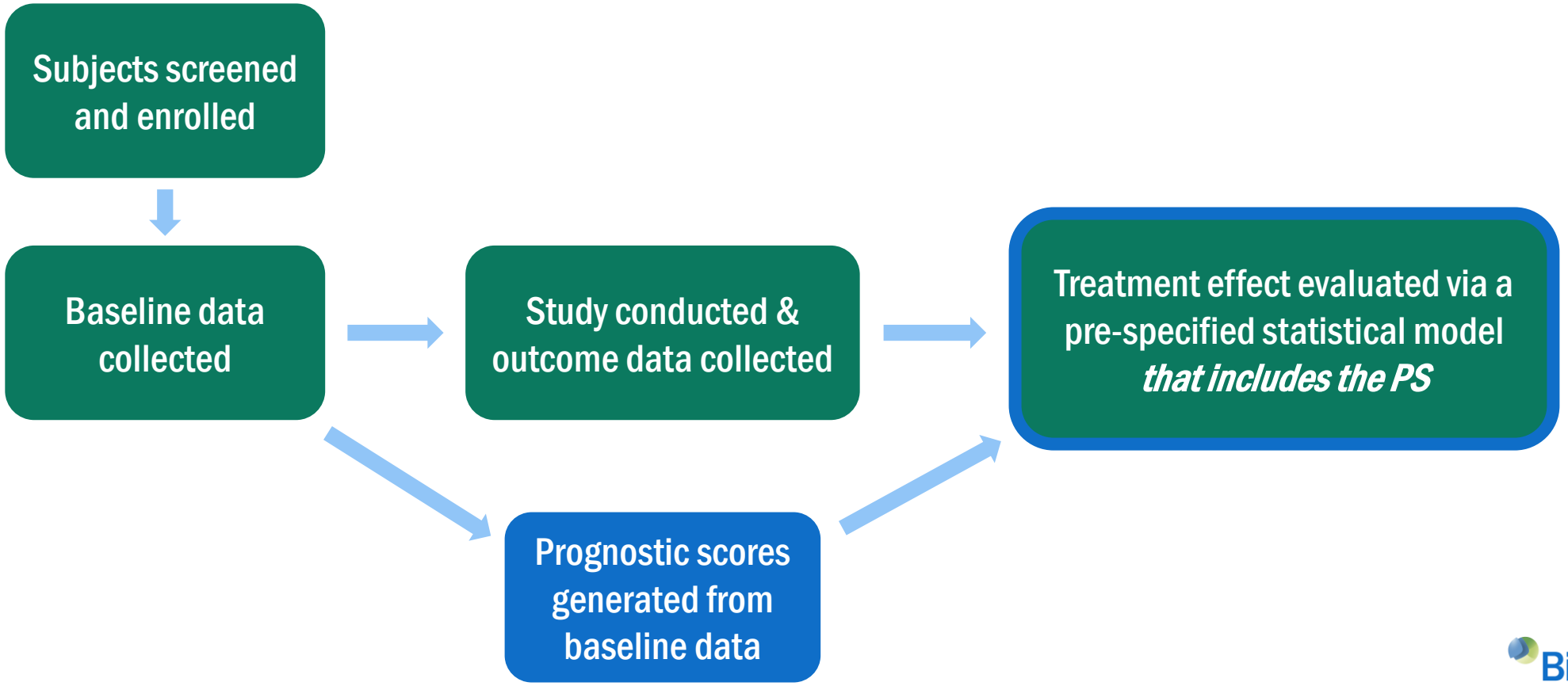


Biogen

HOW PROGNOSTIC SCORES ENHANCE TRIAL ANALYSES

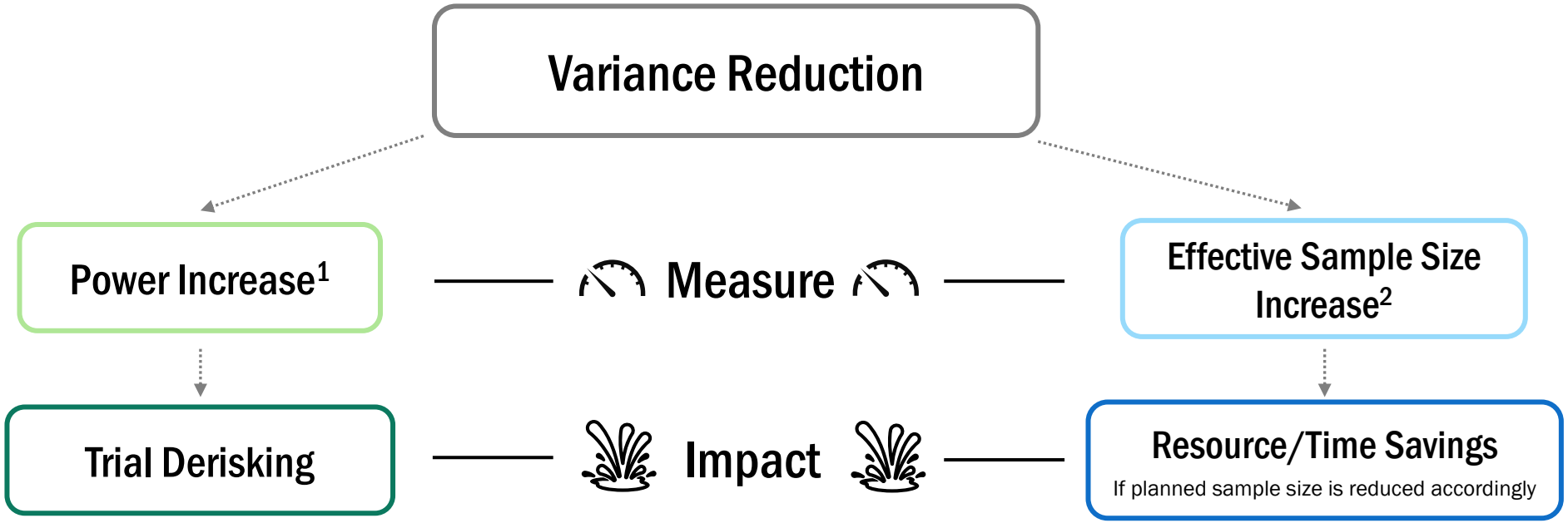
Prognostic scores enable increased study efficiency by reducing unexplained variance via ***covariate adjustment***

A clinical trial analysis with prognostic scores



Biogen

BENEFITS & RISKS OF PS ADJUSTMENT



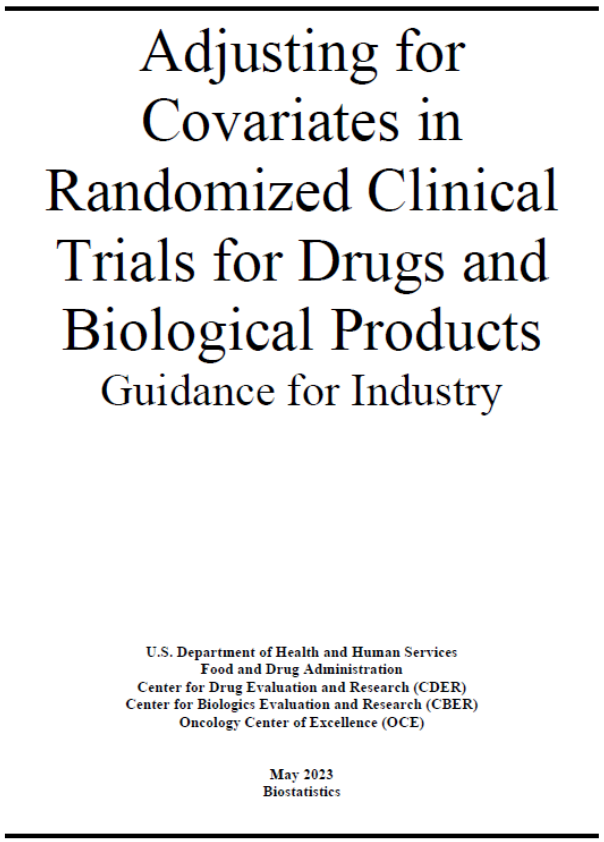
Risks are minimal if PS is prespecified & only incorporates pre-randomization information

- Small power reductions **IF** score has zero prognostic value (due to added degrees of freedom in the model)
- Larger risk **IF** planned sample size is reduced and PS underperforms

[1] Power increase comes from keeping sample size/costs the same combined with variance reduction due to PS adjustment.
[2] There is no actual gain in sample size. Variance reduction makes it as if you had a larger sample.

SUPPORTED BY FDA GUIDANCE

EMA similarly encourages use of prognostic covariate adjustment


Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products

Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Oncology Center of Excellence (OCE)

May 2023
Biostatistics

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. In some circumstances these covariates may be known from the scientific literature. In other cases, it may be useful to use previous studies (e.g., a Phase 2 trial) to select prognostic covariates or form prognostic indices.

Recommend using covariate adjustment to increase trial efficiency

Recommend using historic data to generate "prognostic indices"

Biogen

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

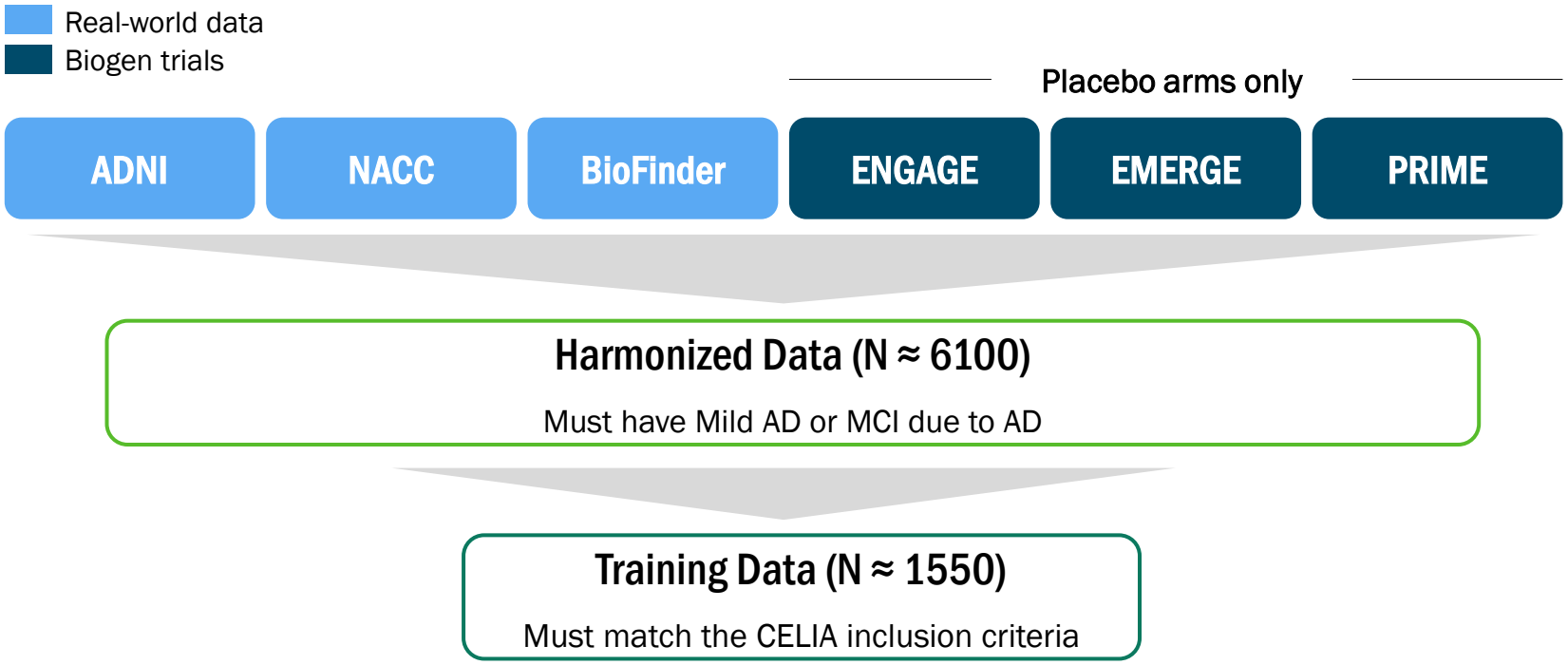
Phase 2: Evaluate

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

Biogen

DATA SOURCES

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



Independent Evaluation

TANGO
N = 650

Biogen

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.

78 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, MoCA

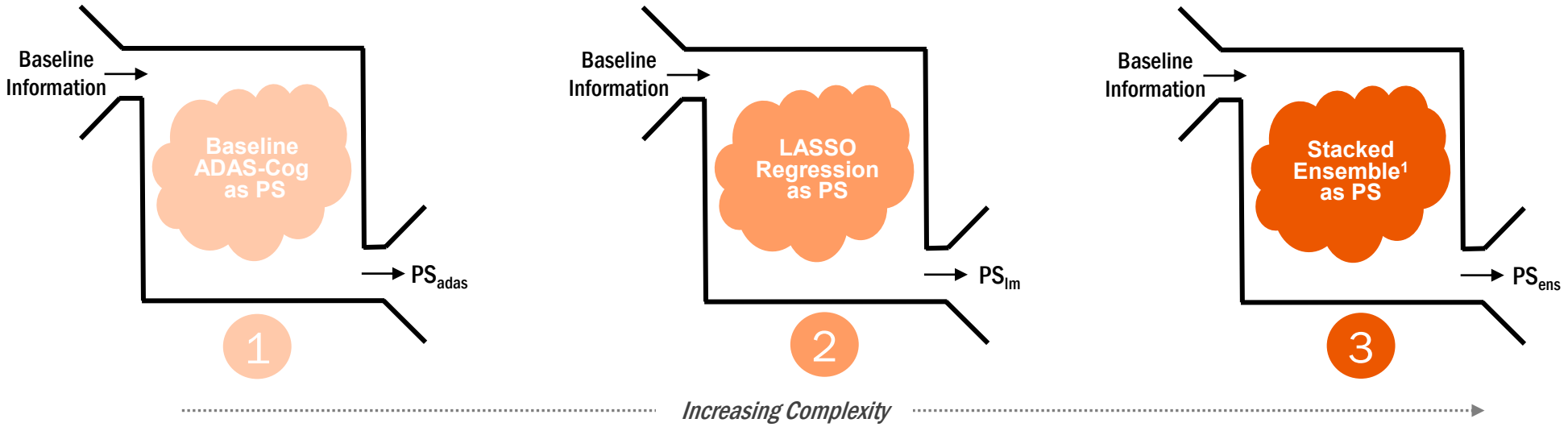
MRI brain region volumes (12)

- E.g., Hippocampus volume

Biogen

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

BENEFIT SUMMARY OF THE PS OPTIONS

The internally-developed ensemble PS is most performant, but there are effective simpler options.

Evaluated in TANGO

Measure[1]	1. ADAS-Cog PS	2. LASSO PS	3. Ensemble PS
Variance Reduction	14.2%	18.6%	19.7%
Effective Sample Size Increase	16.6%	22.9%	24.6%
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.

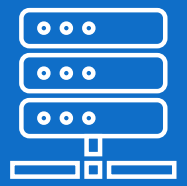
PROS AND CONS OF THE PS OPTIONS

The ensemble PS is most performant, but the LASSO PS benefits from more straightforward interpretation & implementation

	1. ADAS-Cog as PS	2. LASSO PS	3. AI/ML ensemble PS
Pros	• Simplest, easiest to implement operationally • Easiest to interpret	• Good balance between performance and ease of implementation • Easy to interpret: weighted sum of 12 baseline covariates	• Best performing score • Yields the best variance reduction and power increase
Cons	• Leaving value on the table	• Cannot capture non-linear signals in data (if there are any)	• Difficult to interpret; model is a black box • Implementation can be complicated

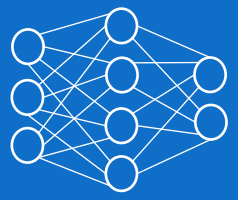
Biogen

TAKEAWAYS & LESSONS LEARNED



High-quality, representative data is critical

- Lack of data standards (structure, variable definitions, etc.) across data sources
- Finding, acquiring, and harmonizing data are often time-consuming and challenging
- Data representativeness usually matters more than volume



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
- Simpler models offer interpretability and ease of implementation

It takes a village!

Cross-group collaboration was essential to the success of this project

- Research statistics
- Medical affairs statistics
- Clinical biostatistics
- Statistical programming
- Internal data sharing/stewardship
- Real world data
- AI and Machine learning

Biogen

PROGNOSTIC SCORES IN RARE DISEASE

Large opportunity for increasing efficiency in resource-constrained environment, but there are unique challenges

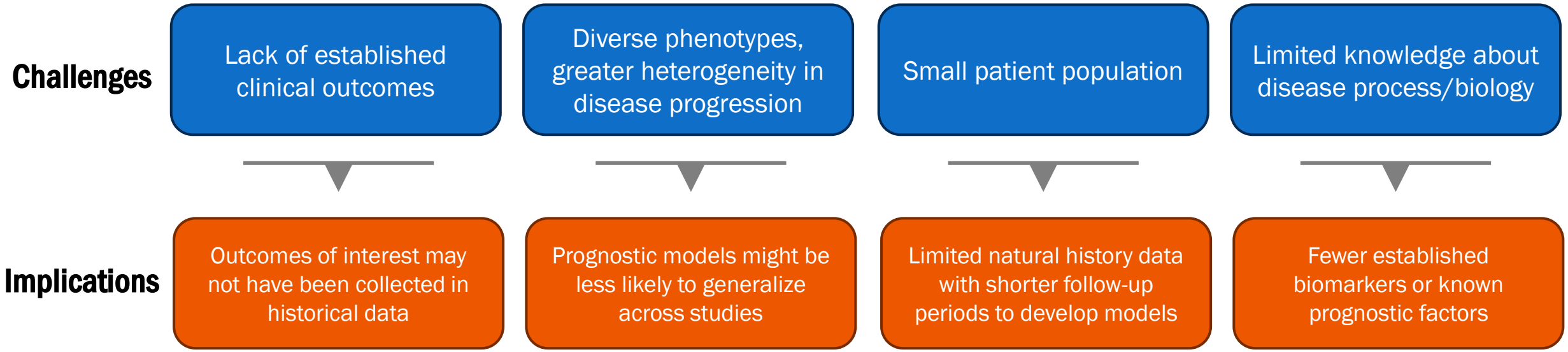
Challenges

- Lack of established clinical outcomes
- Diverse phenotypes, greater heterogeneity in disease progression
- Small patient population
- Limited knowledge about disease process/biology

Biogen

PROGNOSTIC SCORES IN RARE DISEASE

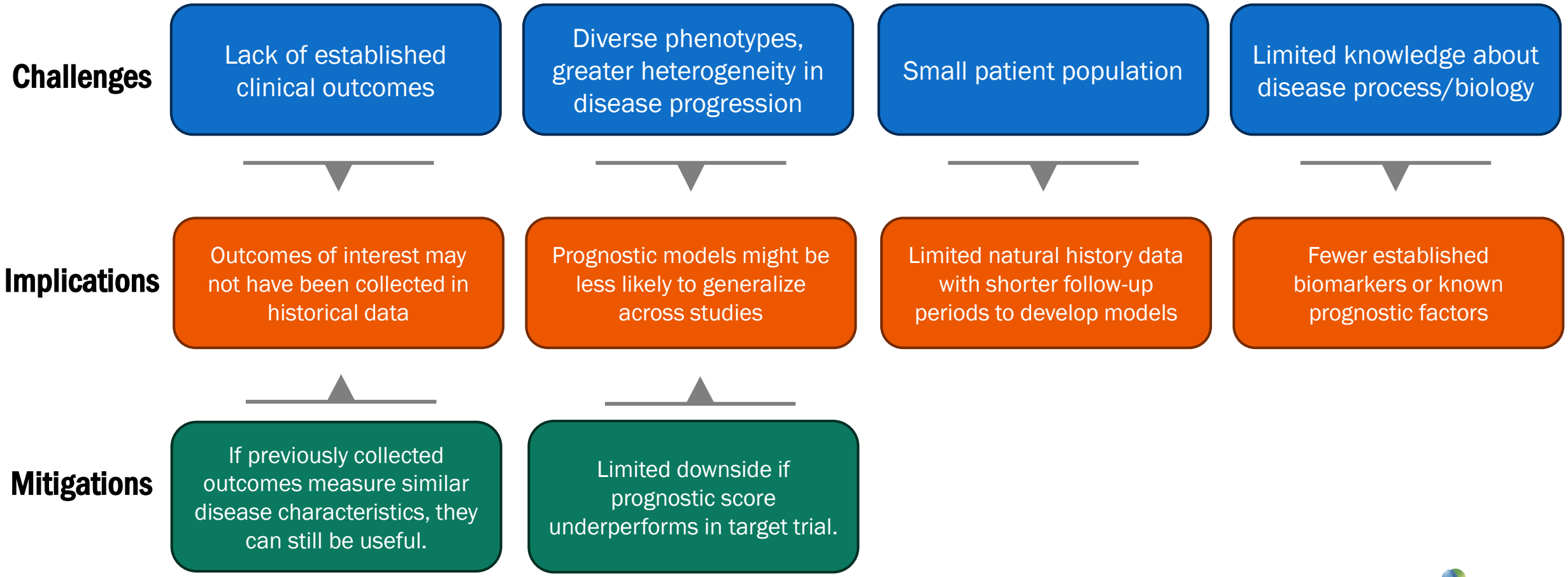
Large opportunity for increasing efficiency in resource-constrained environment, but there are unique challenges



Biogen

PROGNOSTIC SCORES IN RARE DISEASE

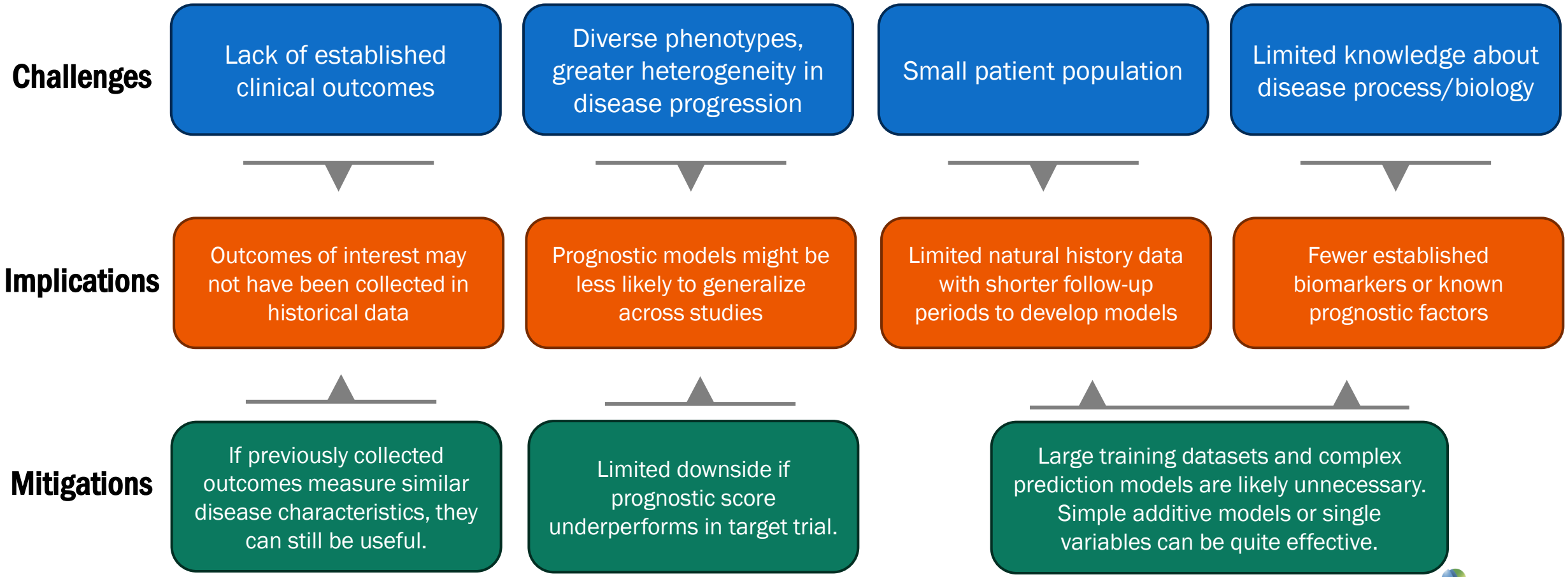
Large opportunity for increasing efficiency in resource-constrained environment, but there are unique challenges



Biogen

PROGNOSTIC SCORES IN RARE DISEASE

Large opportunity for increasing efficiency in resource-constrained environment, but there are unique challenges



Biogen

EFFECTIVENESS OF ADDITIVE MODELS

"Noisy" versus "Non-Noisy" Problems

	Non-Noisy Problems	Noisy Problems
Features	1. Predictors explain the all or most of the variation in the outcome. 2. Outcome is measured/observed without error or with very little error.	1. Predictor set is incomplete/doesn't fully capture the variability in the outcome 2. Outcome is measured with substantial error (e.g., clinical disease outcome measures)
Examples	• Computer vision (e.g., object/facial recognition) • Machine translation & other NLP tasks • Speech processing	• Many (most?) prediction applications in biomedicine • Credit scoring • Predicting power reliability

Biogen

EFFECTIVENESS OF ADDITIVE MODELS

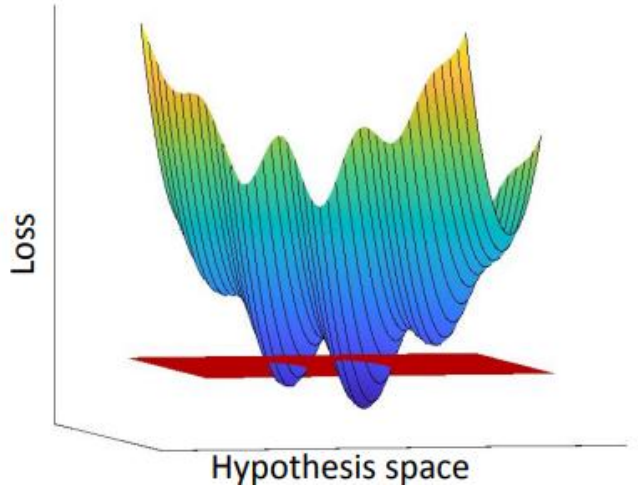
Rashomon effects: When many models with different combinations of features have nearly optimal performance

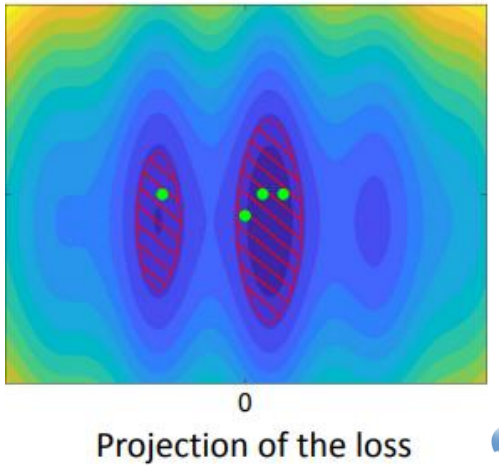
Context

- Rashomon effects are common in “noisy” prediction problems
- The set of models whose loss is below a specific threshold is the **Rashomon Set**
- Sufficiently large Rashomon Sets imply the existence of a simpler model within that set[1]

Consequences for Rare Disease Prognostic Scores

- Predicting progression in rare diseases is a “noisy” problem
- There is unlikely to be an “accuracy-versus-complexity” tradeoff
- Linear PS models will likely provide near-optimal performance (with the added benefit of interpretability and easier implementation)





Biogen

[1]Lesia Semenova, Cynthia Rudin, and Ronald Parr. A study in rashomon curves and volumes: A new perspective on generalization and model simplicity in machine learning. arXiv e-print arXiv:1908.01755, 2019

SUMMARY

Conclusions

- Prognostic scores can increase trial efficiency via trial derisking OR lower required sample sizes
- In AD, ensemble approach is the most performant, enabling
    - 25% effective sample size increase
    - Increase in power of 80% → 88% or 90% → 95%
- Simpler, interpretable models can often achieve near-optimal performance
- Single variables can meaningfully increase precision

Implications for Rare Diseases

- PS adjustment can enable efficiency increases in rare diseases, where limited patient populations make recruitment difficult
- Large historical datasets/complex models are not necessary for developing simple prognostic models or identifying individual prognostic variables
- Low risk of adding a covariate means it can be done without high confidence in degree of precision increase

Biogen

roland.brown@biogen.com