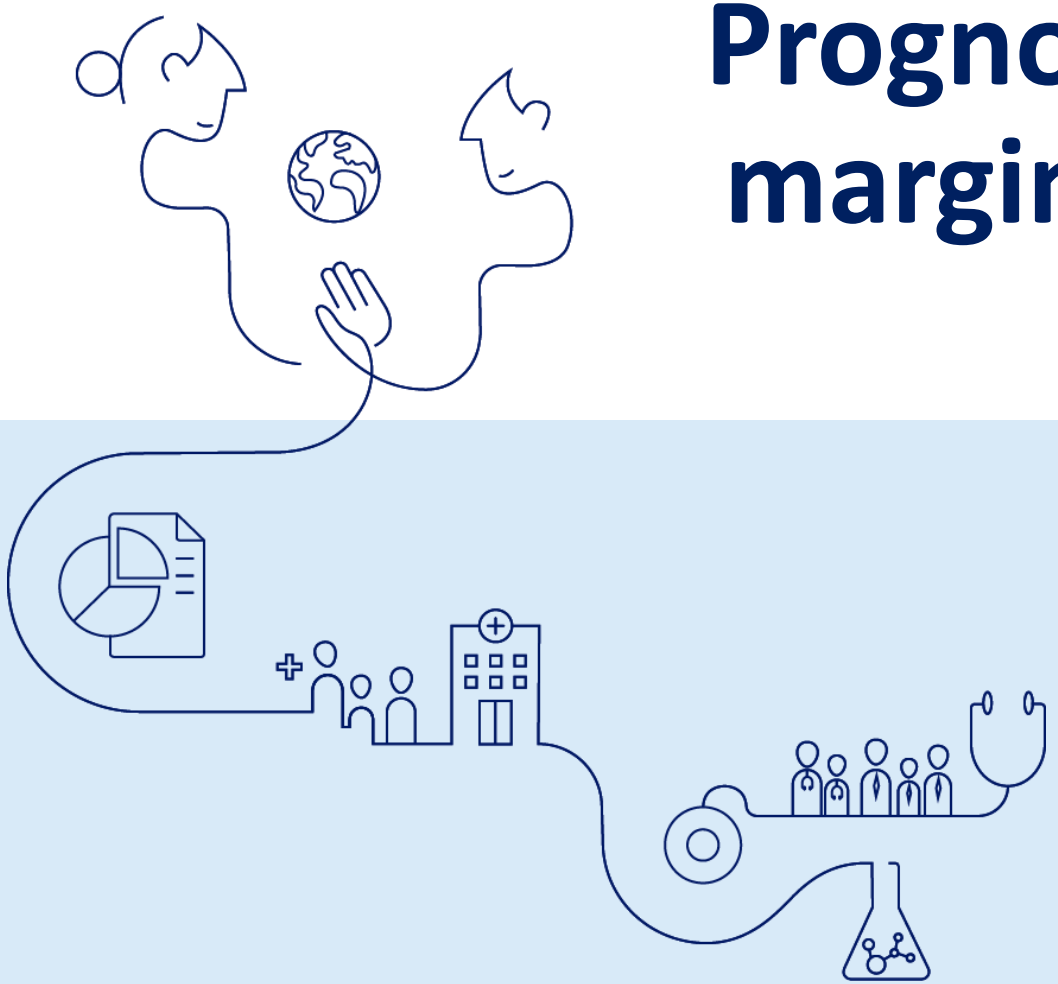


Prognostic score adjustment for marginal effect estimation with GLMs

Enhancing study power through historical data

Emilie Højbjerre-Frandsen, Industrial Ph.D. student at Novo Nordisk and AAU, Mark van der Laan & Alejandro Schuler, UC Berkeley

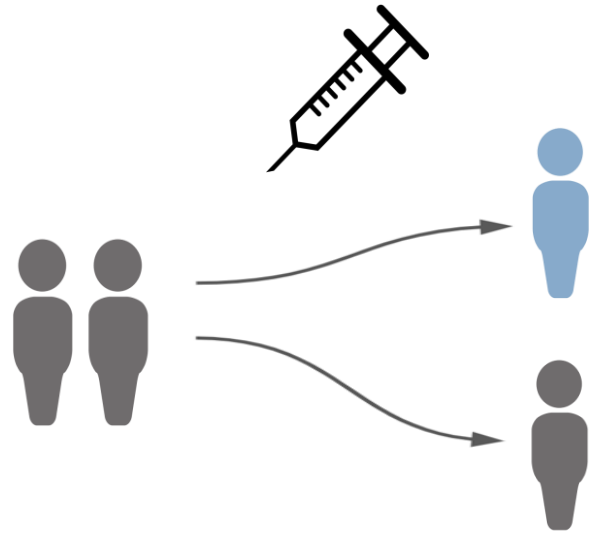
Disclaimer

- Presenter is an employee of Novo Nordisk A/S
- Views and opinions expressed are those of the presenter and not necessarily Novo Nordisk A/S

Average treatment effect

n i.i.d. participants with $O = (W, A, Y) \sim P$

- W is a vector of baseline covariates $(W_1, W_2, W_3, \ldots, W_p)$
- A is indicator for new treatment assignment
- Y represents the primary endpoint continuous variable



New treatment

Existing treatment

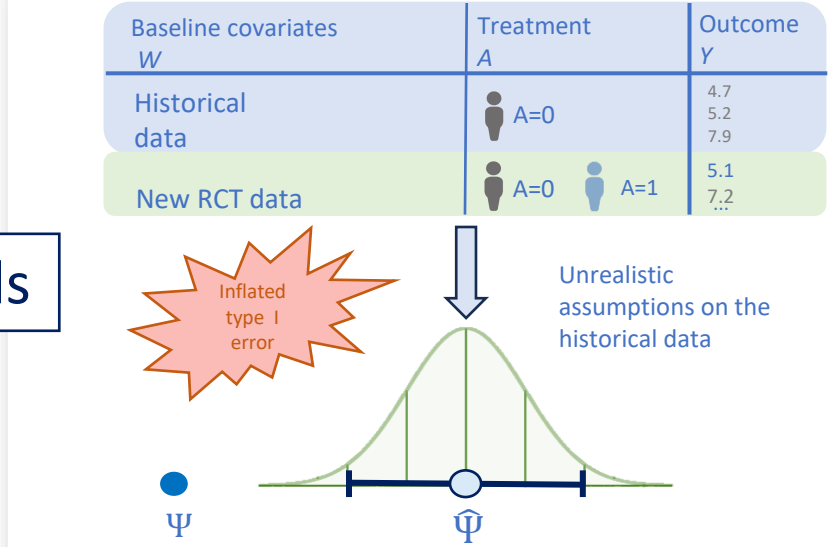
ATE: $\Psi = \Psi_1 - \Psi_0$,

where $\Psi_a = E[Y(a)]$

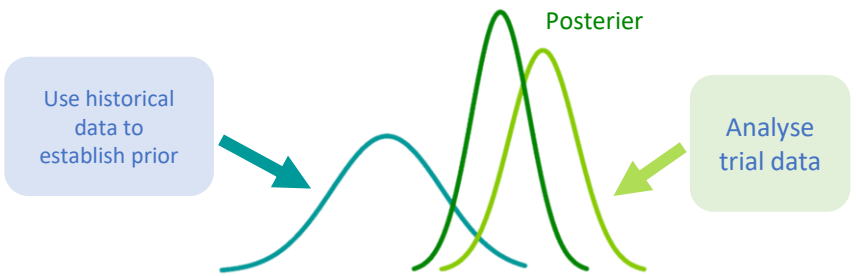
$0 < P(A = a) = \pi_a < 1$

Existing solutions

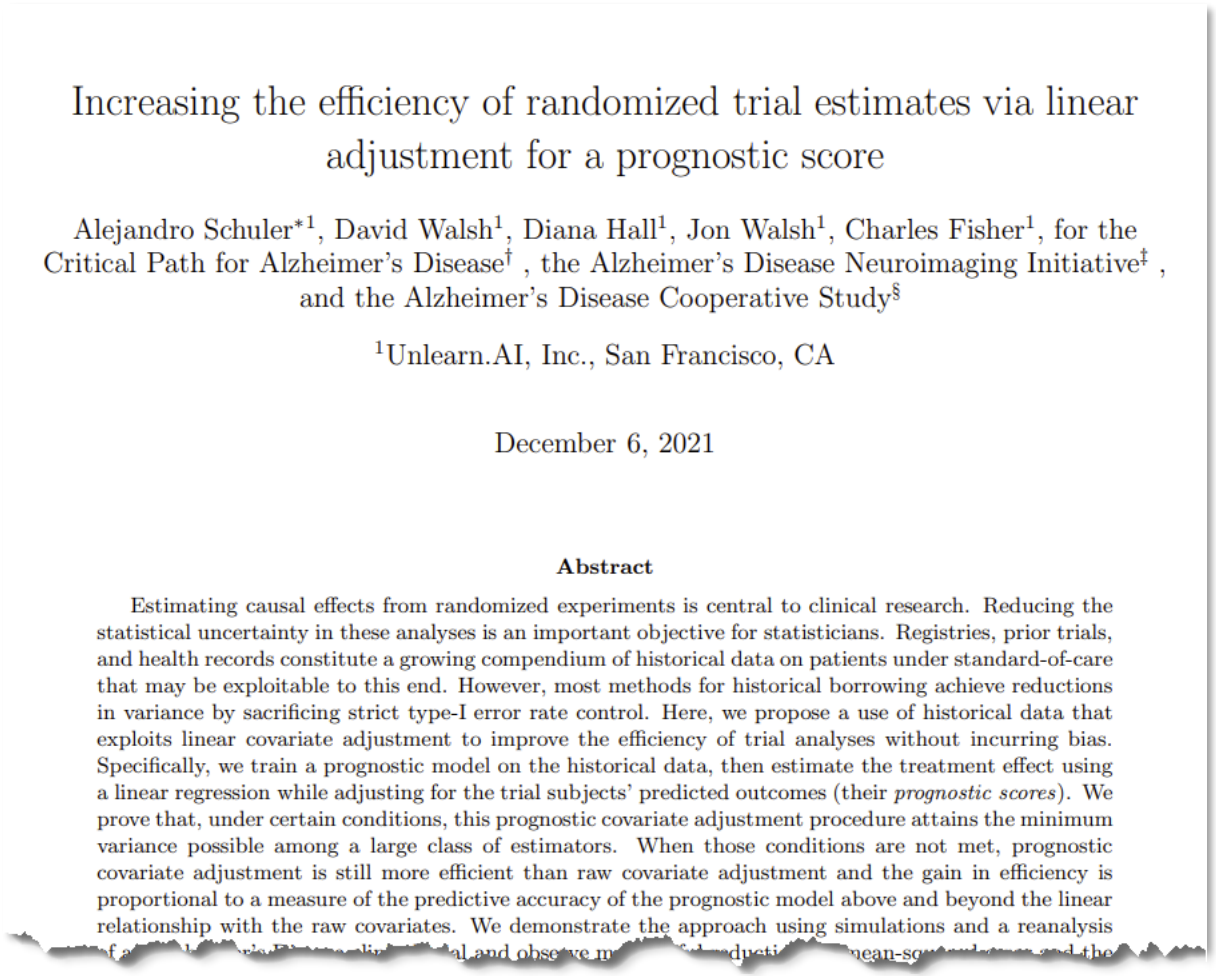
External controls



Bayesian statistics



Prognostic score adjustment for **continuous** endpoint



Increasing the efficiency of randomized trial estimates via linear adjustment for a prognostic score

Alejandro Schuler*[1], David Walsh[1], Diana Hall[1], Jon Walsh[1], Charles Fisher[1], for the Critical Path for Alzheimer's Disease[†] , the Alzheimer's Disease Neuroimaging Initiative[‡] , and the Alzheimer's Disease Cooperative Study[§]

[1]Unlearn.AI, Inc., San Francisco, CA

December 6, 2021

Abstract

Estimating causal effects from randomized experiments is central to clinical research. Reducing the statistical uncertainty in these analyses is an important objective for statisticians. Registries, prior trials, and health records constitute a growing compendium of historical data on patients under standard-of-care that may be exploitable to this end. However, most methods for historical borrowing achieve reductions in variance by sacrificing strict type-I error rate control. Here, we propose a use of historical data that exploits linear covariate adjustment to improve the efficiency of trial analyses without incurring bias. Specifically, we train a prognostic model on the historical data, then estimate the treatment effect using a linear regression while adjusting for the trial subjects' predicted outcomes (their *prognostic scores*). We prove that, under certain conditions, this prognostic covariate adjustment procedure attains the minimum variance possible among a large class of estimators. When those conditions are not met, prognostic covariate adjustment is still more efficient than raw covariate adjustment and the gain in efficiency is proportional to a measure of the predictive accuracy of the prognostic model above and beyond the linear relationship with the raw covariates. We demonstrate the approach using simulations and a reanalysis

Prognostic score adjustment for **continuous** endpoint

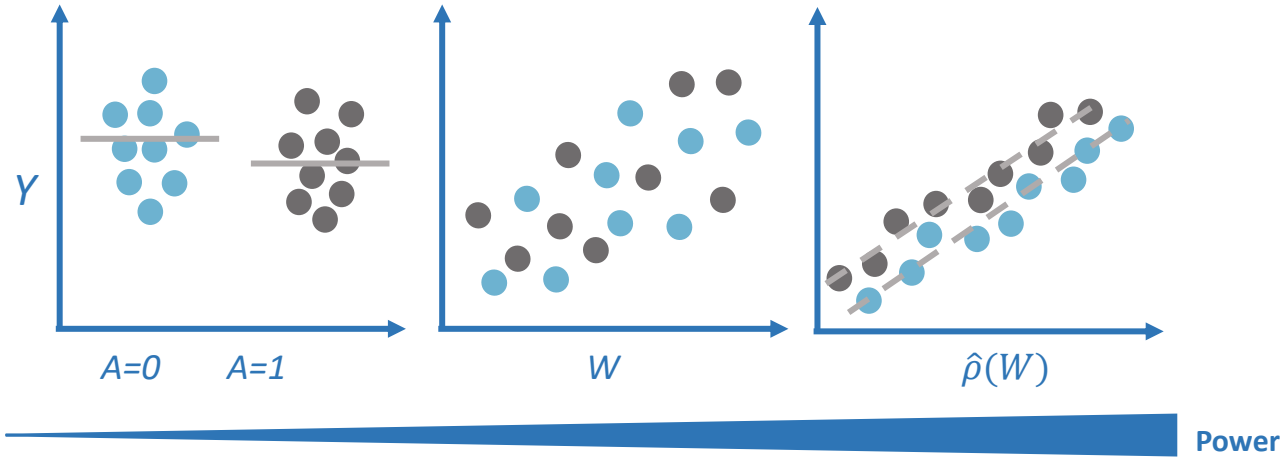
- Determine a prognostic score estimated from historical data

$$\hat{\rho}(W) = \hat{E}[Y \mid W, A = 0, D = 0]$$

- Use ANCOVA model adjusting for $\hat{\rho}(W)$ on the new trial data

$$Y = ATE \cdot A + \beta \cdot W + \alpha \cdot \hat{\rho}(W) + error$$

- The higher correlation with the outcome the higher power increase

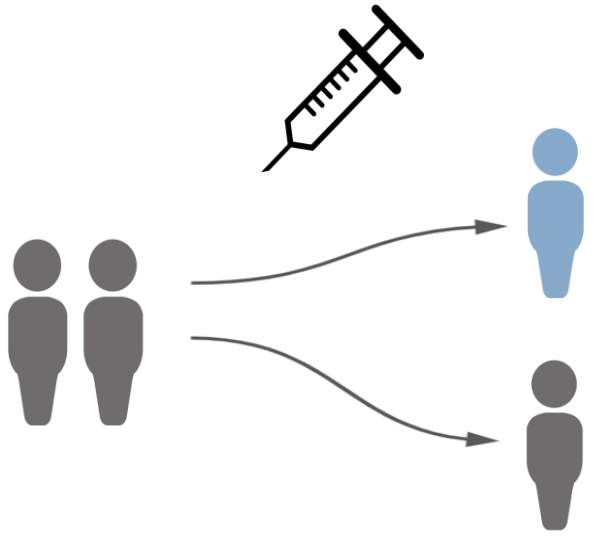


What to do if Y is non-continuous?

Marginal effects

n i.i.d. participants with $O = (W, A, Y) \sim P$

- W is a vector of baseline covariates $(W_1, W_2, W_3, \ldots, W_p)$
- A is indicator for new treatment assignment
- Y represents the primary endpoint variable (continuous, binary, or ordinal)



New treatment

Marginal effects $\Psi = r(\Psi_1, \Psi_0)$,
where $\Psi_a = E[Y(a)]$

Existing treatment

$0 < P(A = a) = \pi_a < 1$

Estimating marginal effects

Rosenblum and van der Laan (2010) proposed a GLM based plug-in estimator:

1. Use MLE to estimate the conditional mean as

$$\hat{\mu}(a,w) = \hat{E}[Y \mid W = w, A = a] = g^{-1}\left(\hat{\beta}_0 + x\hat{\beta}_x\right)$$

   where g is the link function and x is a row in the design matrix

2. Estimate the population mean outcome for each treatment

$$\hat{\Psi}_a = \frac{1}{n}\sum_{i=1\ldots n} \hat{\mu}(a, w_i)$$

3. Plug-in to get the estimator on the right scale

$$\hat{\Psi} = r\left(\hat{\Psi}_1, \hat{\Psi}_0\right)$$

4. Using delta-method to determine the IF and estimate the asymptotic variance as

$$\hat{v}_\infty^2 = \frac{1}{n}\sum_{i=1\ldots n} \left(r_0'\left(\hat{\Psi}_1, \hat{\Psi}_0\right)\cdot \hat{\phi}_0(A_i, W_i, Y_i) + r_1'\left(\hat{\Psi}_1, \hat{\Psi}_0\right)\cdot \hat{\phi}_1(A_i, W_i, Y_i)\right)^2,$$

   where $\hat{\phi}_a(A_i, W_i, Y_i) = \frac{1_a(A_i)}{\pi_a}\left(Y - \hat{\mu}(a, W_i)\right) + \left(\hat{\mu}(a, W_i) - \hat{\Psi}_a\right)$.

*Regular and asymptotically linear (RAL)

Theorem 1 (Rosenblum & van der Laan): The plug-in based estimator of the marginal effect is RAL → remain consistent and asymptotically normal, regardless of the type of misspecification.

The estimator is locally efficient assuming only an RCT setting.

Use GLM plug-in model adjusting for $g(\hat{\rho}(W))$

How the method works

Step 1

- Curate historical data from different sources
- Train a prognostic model $\hat{\rho}$

Step 2

- Evaluate the performance of the prognostic model
- **Estimate population parameters for power estimation** on an independent test data set

Step 3

- Predict the prognostic scores for each of the participants in the new trial

Patient Number	W_1	W_2	W_3	...	W_p	A	Y
1	M	48	175		55	1	34
2	M	34	179		64	0	42
3	K	18	189		87	1	67
4	M	22	165		35	0	21
....							

Patient Number	W_1	W_2	W_3	...	W_p	A	Y	$\hat{\rho}$
1	M	48	175		55	1	34	$\hat{\rho}(w^1)$
2	M	34	179		64	0	42	$\hat{\rho}(w^2)$
3	K	18	189		87	1	67	$\hat{\rho}(w^3)$
4	M	22	165		35	0	21	$\hat{\rho}(w^4)$
....								

Step 4

- Use GLM plug-in model adjusting for $g(\hat{\rho}(W))$.
- Type I error control → only affects precision, not bias

Theorem 1

Assume that the treatment effect is additive on the link-scale, i.e.

$$g(\mu(1,W)) = \zeta + g(\mu(0,W)),$$

and that W and Y have compact support .

Let $f_{\tilde{n}}$ be a uniformly bounded random function learned from historical data obtained from $\tilde{n}$ participants independent from the current trial data. Assume that

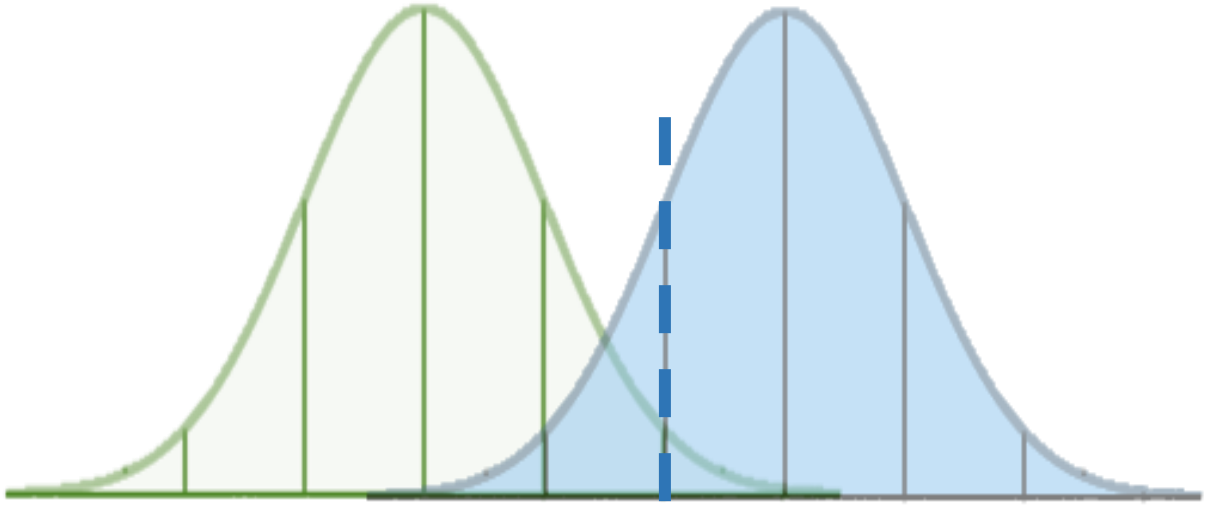
$$|f_{\tilde{n}}(W) - \mu_0(W)| \xrightarrow{L_2} 0,$$

for $\tilde{n} \to \infty$ with $\mu_0(w) = E[Y(0)|w]$ bounded.

Then if $n = \mathcal{O}(\tilde{n})$, the estimator obtained from the plug-in GLM procedure that uses $g(f_{\tilde{n}}(W))$ as an additional covariate is consistent and efficient in the sense that it has the lowest possible asymptotic variance among all RAL estimators with access to W.

**Regular and asymptotically linear (RAL)*

Prospective power estimation



Conservative estimation of asymptotic variance based on few estimable population parameters even when model is misspecified

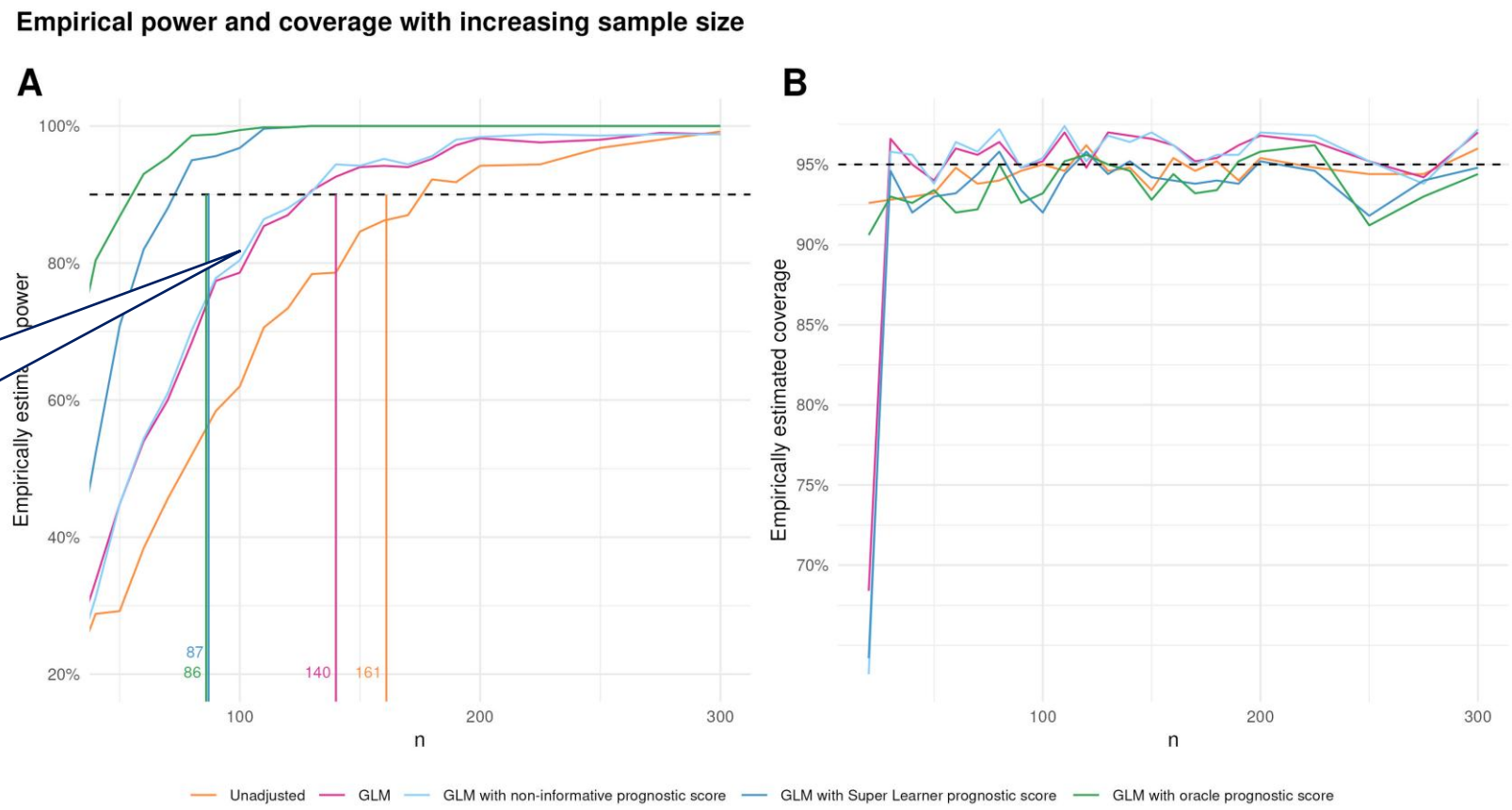
$$v_{\infty}^{\uparrow 2} = r_0{}'^2\sigma_0^2 + r_1{}'^2\sigma_1^2 + \pi_0\pi_1(|r_0'|\frac{\kappa_0}{\pi_0} + |r_1'|\frac{\kappa_1}{\pi_1})^2$$

- The marginal variance of the potential outcome $\sigma_a^2 = Var(Y(a))$
- The expected mean-squared error of best GLM fit $\kappa_a^2 = E\left[(Y(a) - \mu^*(a,W))^2\right]$

Simulation study

Power and coverage in heterogeneous treatment effect scenario

Robustness property



** n is the sample size of the current RCT data, with the historical data amount being ñ=10*n*

Practical experience

- Difficult to combine historical data into a curated data set
- Strong predictors → the gain in precision might not be as big

Compromises

- Lack of power if prognostic score has a lower effect than assumed
  - However, not lower than the power for the analysis without prognostic score with reduced number of participants
- Secondary endpoints
- Subgroup analyses

Questions?