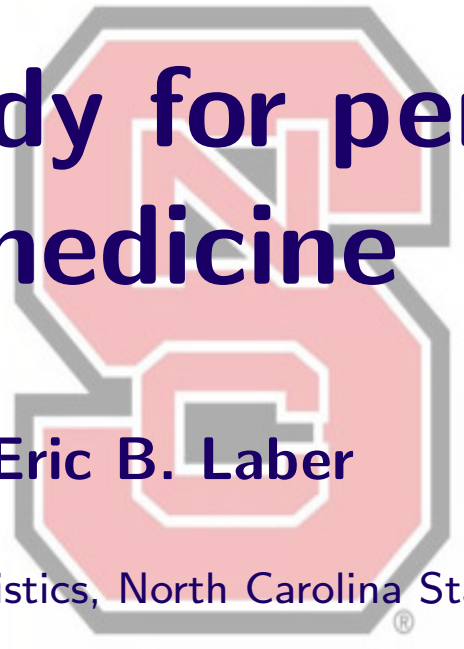
Sizing a study for personalized medicine



Eric B. Laber

Department of Statistics, North Carolina State University

IMPACT, Cary, NC, November 21, 2014

Acknowledgments

- IMPACT organizers
- Duke center for personalized medicine
- NCSU/UNC AI and personalized medicine lab
- NCSU personalized medicine cluster
- Joint work with: Marie Davidian, Michael Kosorok, Todd Regh, Rui Song, Butch Tsiatis, Ming Yuan, Donglin Zeng, Yingqi Zhao

Personalized medicine

- "The right treatment for the right patient at the right time."
  —Mantra of personalized medicine
  - Best clinical care requires treatment decisions based on individual patient characteristics
  - Improve patient outcomes but also reduce cost, patient burden, and resource consumption
- Data-driven personalized medicine
  - Treatment regime formalizes personalized medicine as a function from patient info to a recommended treatment
  - Can we size a clinical trial to ensure estimation of a high-quality treatment regime?

Statistical rigor and personalized medicine

- Identifying subgroups has negative connotations
  - Data snooping
  - Lack of reproducibility/generalizability
  - Perceived lack of statistical rigor*
- Goal of personalized medicine should be reflected in design
  - Identify planned analysis
  - State evaluation criteria
  - Estimate how much data will be needed

Sample size calculation for personalized medicine

- Sample size calculation typically based on primary aim
  - Simple hypotheses, e.g., comparison of two means
  - Clear criteria for power calculations
  - Test stats have regular sampling distns
- Estimation of a treatment regime is typically a secondary, hypothesis generating, analysis
  - Composite hypotheses, high-dim nuisance parameters
  - Many potential criteria for power calculations
  - Test stats have non-regular sampling distns

One approach

- Size the study to ensure:
  - (C1) Power to detect an improvement in the mean outcome under the optimal treatment regime relative to standard care
  - (C2) Near optimality of the estimated optimal regime with high-probability
- Invert projection confidence intervals to obtain sample size
  - Valid despite non-regularity
  - Potentially conservative
- Assume existence of a small pilot study
  - Only elicit clinically meaningful quantities

Setup

- Trial will collect data on n subjects $\mathcal{D}(n) = \{(\mathbf{X}_i, A_i, Y_i)\}_{i=1}^n$
  - $\mathbf{X} \in \mathbb{R}^p$ pre-treatment subj. info.
  - $A \in \{-1, 1\}$ treatment assigned
  - $Y \in \mathbb{R}$ outcome, higher is better
- Treatment regime $\pi : \operatorname{dom} \mathbf{X} \to \operatorname{dom} A$
  - Patient presenting with $\mathbf{X} = \mathbf{x}$ recommended $\pi(\mathbf{x})$
  - Define $V(\pi) = \mathbb{E}^\pi Y$, optimal regime $\pi^{\text{opt}} = \arg\sup_{\pi \in \Pi} V(\pi)$
  - Many methods for estimating π^{opt} (EBL, et al., 2014)
  - Focus on Q-learning (Murphy, 2005)

Analysis method: Q-learning

- Q-function $Q(\mathbf{x}, a) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x}, A = a)$
  - $\pi^{\text{opt}}(\mathbf{x}) = \arg\max_a Q(\mathbf{x}, a)$
  - Estimate π^{opt} via $\widehat{\pi}_n(\mathbf{x}) = \arg\max_a \widehat{Q}_n(\mathbf{x}, a)$, where $\widehat{Q}_n$ is estimator of Q
- Lin. model $Q(\mathbf{x}, a; \beta) = \mathbf{x}_0^\intercal \beta_0 + a\mathbf{x}_1^\intercal \beta_1$, $\mathbf{x}_0, \mathbf{x}_1$ features of $\mathbf{x}$
  - Define $\widehat{\beta}_n = \arg\min_\beta \mathbb{P}_n \{Y - Q(\mathbf{X}, A; \beta)\}^2$
  - $\widehat{Q}_n(\mathbf{x}, a) = Q(\mathbf{x}, a; \widehat{\beta}_n)$
  - Define β^* to be the population analog of $\widehat{\beta}_n$
- Write $V(\beta)$ to denote $V(\pi_\beta)$ where $\pi_\beta(\mathbf{x}) = \arg\max_a Q(\mathbf{x}, a; \beta)$

Analysis method: Q-learning

- Q-function $Q(\mathbf{x}, a) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x}, A = a)$
  - $\pi^{\text{opt}}(\mathbf{x}) = \arg\max_a Q(\mathbf{x}, a)$
  - Estimate π^{opt} via $\widehat{\pi}_n(\mathbf{x}) = \arg\max_a \widehat{Q}_n(\mathbf{x}, a)$, where $\widehat{Q}_n$ is estimator of Q

- Lin. model $Q(\mathbf{x}, a; \beta) = \mathbf{x}_0^\intercal \beta_0 + a\mathbf{x}_1^\intercal \beta_1$, $\mathbf{x}_0, \mathbf{x}_1$ features of $\mathbf{x}$
  - Define $\widehat{\beta}_n = \arg\min_\beta \mathbb{P}_n \{Y - Q(\mathbf{X}, A; \beta)\}^2$
  - $\widehat{Q}_n(\mathbf{x}, a) = Q(\mathbf{x}, a; \widehat{\beta}_n)$
  - Define β^* to be the population analog of $\widehat{\beta}_n$

- Write $V(\beta)$ to denote $V(\pi_\beta)$ where $\pi_\beta(\mathbf{x}) = \arg\max_a Q(\mathbf{x}, a; \beta)$

Estimator of $V(\beta^*)$

- For any β, let $\widehat{V}_{\mathcal{D}(n)}(\beta)$ denote an augmented IPWE for $V(\beta^*)$ (Zhang et al., 2012)
  - $\widehat{V}_{\mathcal{D}(n)}(\beta^*)$ is unbiased for $V(\beta^*)$
  - $\widehat{V}_{\mathcal{D}(n)}(\widehat{\beta}_n)$ is an augmented IPWE of $V(\beta^*)$ (Zhang et al., 2012)
- $\mathbb{G}_n = n^{1/2}\left\{\widehat{V}_{\mathcal{D}(n)}(\widehat{\beta}_n) - V(\beta^*)\right\}$ is non-regular and need not even be $O_P(1)$
  - Standard asymptotic theory does not apply
  - Sample size calculations cannot be based directly on $\mathbb{G}_n$

Formalizing (C1) and (C2)

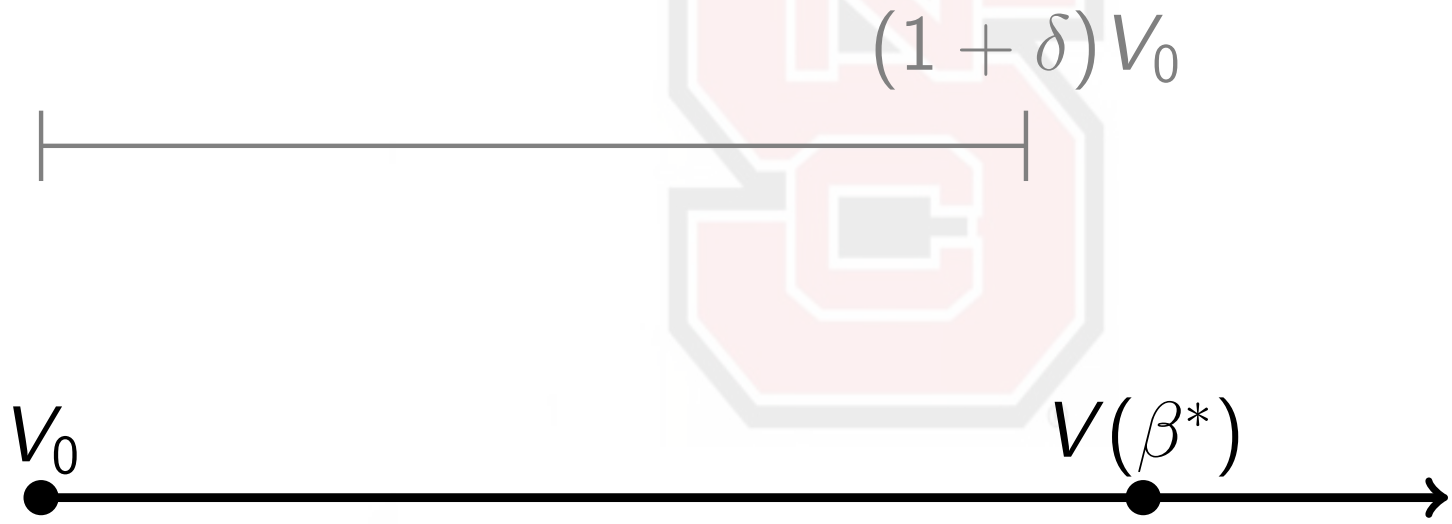
- Let V_0 be an elicited expected outcome under standard care
- Pilot data $\mathcal{D}_p = \mathcal{D}(n_p)$, want $n^* = n^*(\mathcal{D}_p)$ so that:

  (C1) The power to reject the hypothesis $V(\beta^*) = V_0$ in favor of $V(\beta^*) > V_0$ is at least $1 - \rho$ when $V(\beta^*) \geq (1 + \delta)V_0$

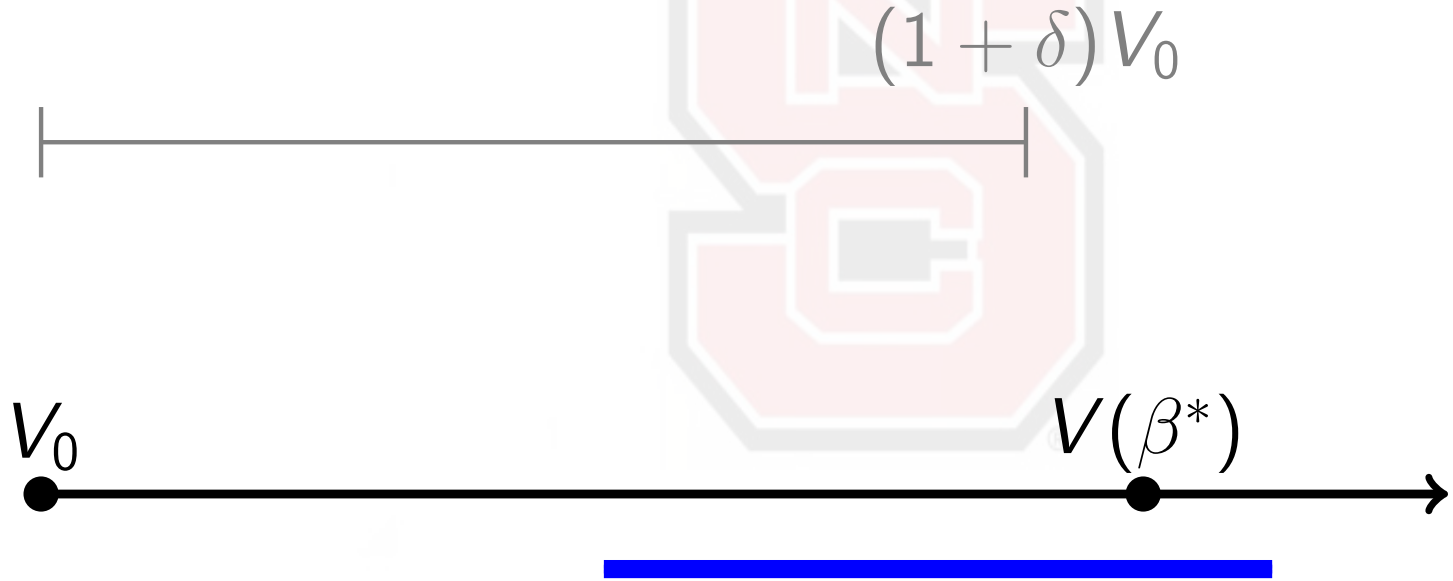
  (C2) $P(|\widehat{V}_{\mathcal{D}(n^*)}(\widehat{\beta}_{n^*}) - V(\beta^*)| \leq \epsilon) \geq 1 - \rho$

  where $\rho, \epsilon, \delta > 0$ are fixed constants

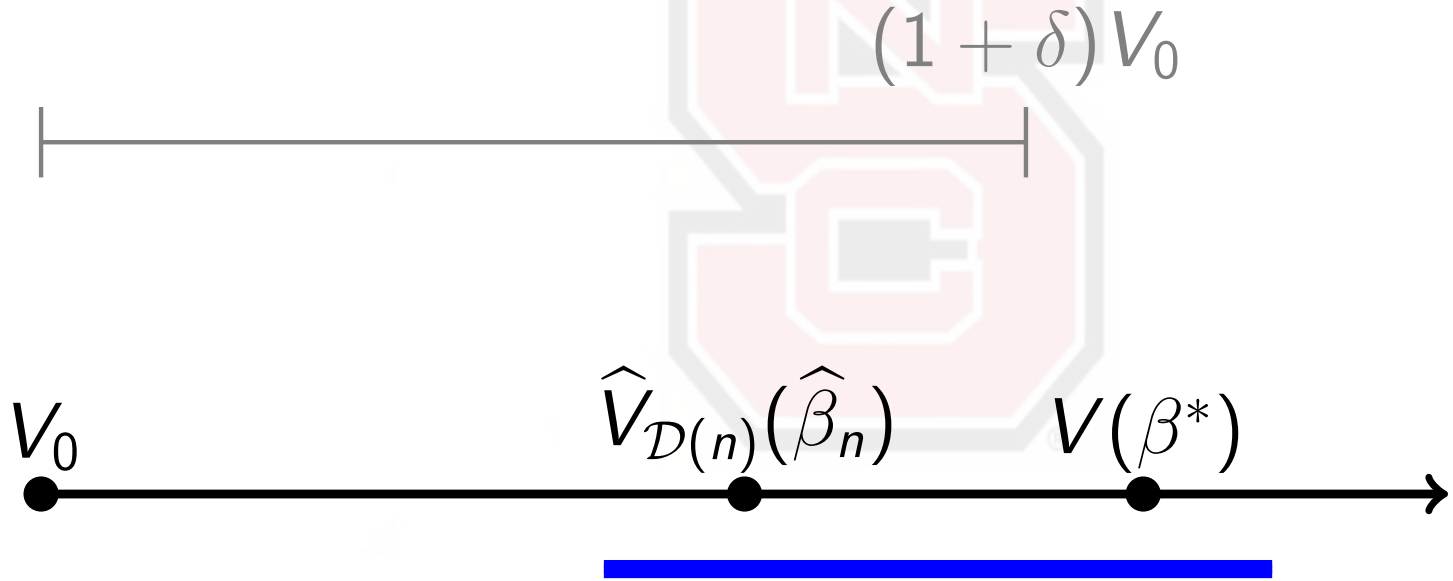
Inverting a confidence interval



Inverting a confidence interval



Inverting a confidence interval



Inverting a confidence interval

- Suppose $\zeta_{n,1-\rho}$ is a $(1-\rho)\times 100\%$ confidence interval for $V(\beta^*)$ with diameter $d(n)$
  - If $d(n) \leq \delta V_0$ then $V(\beta^*)$ and V_0 cannot simultaneously belong to $\zeta_{n,1-\rho}$ if $(1+\delta)V_0 \leq V(\beta^*)$
  - If $d(n) \leq \epsilon$ and $\widehat{V}_n(\widehat{\beta}_n) \in \zeta_{n,1-\rho}$ then $P\left\{|\widehat{V}_n(\widehat{\beta}_n) - V(\beta^*)| \leq \epsilon\right\} \geq 1-\rho$
- Let $\widehat{d}_{\mathcal{D}_p}(n)$ be an estimator of $d(n)$ then

$$\widehat{n}^* = \inf\left\{n : \widehat{d}_{\mathcal{D}_p}(n) \leq \min(\delta V_0, \epsilon)\right\}$$

is our estimator of n^*

Constructing a confidence interval for $V(\beta^*)$

- Idea: treat β^* as nuisance parameter and construct projection interval (Berger and Boos, 1994; Robins, 2004)
- For fixed β, $n^{1/2}\left\{\widehat{V}_n(\beta) - V(\beta)\right\}$ is asymptotically normal with mean zero and variance $\sigma^2(\beta)$

$$\mathbb{I}_\gamma(\mathcal{D}_p, n_p) = \left(\widehat{V}_{n_p}(\gamma) - \frac{z_{1-\mu/2}\widehat{\sigma}_{n_p}(\gamma)}{\sqrt{n_p}}, \widehat{V}_{n_p}(\gamma) + \frac{z_{1-\mu/2}\widehat{\sigma}_{n_p}(\gamma)}{\sqrt{n_p}}\right)$$

  is a valid $(1-\mu)\times 100\%$ CI for $V(\beta)$

- Let $\mathbb{T}(\mathcal{D}_p, n_p)$ be a $(1-\xi)\times 100\%$ Wald-type CI for β^*

Constructing a confidence interval for $V(\beta^*)$ cont'd

- A $(1-\mu-\xi)\times 100\%$ projection CI for $V(\beta^*)$ is

$$\bigcup_{\beta\in\mathbb{T}(\mathcal{D}_p,n_p)} \mathbb{I}_\beta(\mathcal{D}_p,n_p),$$

which has diameter

$$d(n_p) = \sup_{\beta\in\mathbb{T}(\mathcal{D}_p,n_p)} \mathbb{I}_\beta(\mathcal{D}_p,n_p) - \inf_{\beta\in\mathbb{T}(\mathcal{D}_p,n_p)} \mathbb{I}_\beta(\mathcal{D}_p,n_p)$$

- Estimated diameter is

$$\widehat{d}_{\mathcal{D}_p}(n) = \sup_{\beta\in\mathbb{T}(\mathcal{D}_p,n)} \mathbb{I}_\beta(\mathcal{D}_p,n) - \inf_{\beta\in\mathbb{T}(\mathcal{D}_p,n)} \mathbb{I}_\beta(\mathcal{D}_p,n),$$

which determines $\widehat{n}^* = \inf\left\{n : \widehat{d}_{\mathcal{D}_p}(n) \le \min(\delta V_0, \epsilon)\right\}$

Simulated experiment

- Use generative model

$$
\begin{aligned}
U_i &\sim \text{Bernoulli}(\nu), & X_i &\sim \text{N}_p\left(0, I_p\right), \\
Z_i &= (I - U_i P_{\beta^*}) X_i, & A_i &\sim \text{Uniform}\left\{-1, 1\right\}, \\
e_i &\sim \text{N}\left(0, 1\right), & Y_i &= X_i^\top \alpha^* + A_i Z_i^\top \beta^* + e_i,
\end{aligned}
$$

  where P_{β^*} is projection matrix onto $\text{span}\, \beta^*$

- Addl details
  - Dimension of $\mathbf{X}$ is $p = 5$
  - Pilot study size $n_p = 20$
  - Set $\delta = \epsilon = V_0 = 1$, $V(\beta^*) = 2$, 2.25, and $\rho = 0.80$
  - Compare with standard normal-based confidence interval

Results

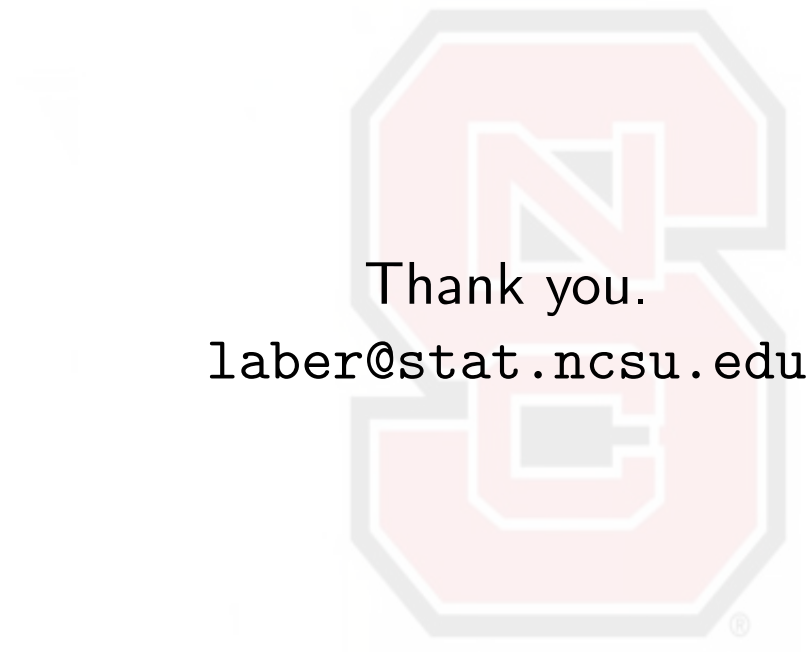
		Projection sample size				Normal sample size			
V^*	ν	$V(\beta^*)$	V_0	$\mathbb{E}\widehat{n}^*$	SD	$V(\beta^*)$	V_0	$\mathbb{E}\widehat{n}^*$	SD
2	0	**0.77**	**0.08**	156	151	**0.77**	0.14	111	62
2	0.05	**0.77**	**0.06**	144	132	**0.77**	0.13	107	62
2	0.1	**0.80**	**0.07**	144	127	0.76	0.14	100	60
2	0.25	**0.80**	**0.08**	149	123	0.76	0.16	98	65
2	0.5	**0.79**	**0.09**	191	204	0.72	0.19	85	65
2	0.75	**0.79**	**0.09**	191	204	0.70	0.23	101	104
2.25	0	0.74	0.02	145	138	**0.78**	**0.073**	120	68
2.25	0.05	**0.77**	**0.02**	147	143	**0.78**	0.07	112	65
2.25	0.1	**0.77**	**0.03**	142	118	**0.77**	0.067	114	71
2.25	0.25	**0.78**	**0.04**	166	157	0.75	0.09	112	75
2.25	0.5	**0.79**	**0.03**	228	216	0.74	0.11	105	84
2.25	0.75	**0.77**	**0.06**	347	390	0.69	0.17	129	135

Simulation conclusions

- Reliable sample size estimates in simulated examples
- Works with small sample size and moderate number of covariates
- Large standard deviation
- Mid-stream update (not shown) reduces variability and $\mathbb{E}\widehat{n}^*$ while maintaining operating characteristics

Discussion

- Sample size calculation for personalized medicine is possible
- Many possible criteria for sizing the study, more thought is needed
- Extending this to the multistage setting without becoming excessively conservative is an open problem



Thank you.

laber@stat.ncsu.edu