

**Jason Abaluck**  
optimal Representation  
Ephraim Sutherland

**Setup**

1. Suppose a physician can only see ATE and some measure of representativeness. They have prior  $\bar{\beta}$  and  $\beta_{ATE} = (1/N) \sum \beta_j$ .
2. need model for betas related to each other based on  $x'$ s. WLOG, suppose

$$\beta(x_j) = x_j \gamma$$

If you knew the gamma, then you would know beta for any given patient.

3. You don't observe gamma, you instead observe:  $\beta_{ATE} = \bar{x} \gamma$  where  $\bar{x} = (\frac{1}{N}) \sum x_i$
4. We know  $\beta_i$  for patients with characteristics  $\bar{x}$  (it is  $\beta_{ATE}$ ).
5. For other patients, need to solve

$$E(x_i \gamma | \bar{x} \gamma = \beta_{ATE})$$

6. to solve

(a)

$$\begin{aligned} \beta_{i,post} &= E(x_i \gamma | \bar{x} \gamma = \beta_{ATE}) \\ &= E((x_i - c\bar{x})\gamma | \bar{x} \gamma = \beta_{ATE}) + cE(\bar{x}\gamma | \bar{x} \gamma = \beta_{ATE}) \\ &= E((x_i - c\bar{x})\gamma | \bar{x} \gamma = \beta_{ATE}) + c\beta_{ATE} \end{aligned}$$

For any constant  $c$ .

Choose  $c$  so that

$$\text{Cov}((x_i - c\bar{x})\gamma, \bar{x}\gamma) = 0$$

maybe assume normality so that this guarantees independence. Then,

$$E((x_i - c_i \bar{x})\gamma | \bar{x} \gamma = \beta_{ATE}) = (x_i - c_i \bar{x})E(\gamma)$$

So then

$$(x_i - c\bar{x})E(\gamma) + c_i \beta_{ATE} = x_i E(\gamma) + c_i (\beta_{ATE} - \bar{x} E(\gamma))$$

(c depends on  $x_i$ )

In other words, your belief is your prior, adjusted based on the difference between the observed ATE and your prior about the ATE. The key question is how much adjustment you do which depends on "c". We choose c to solve:

$$\begin{aligned} \text{Cov}((x_i - \bar{x})\gamma, \bar{x}\gamma) &= 0 \\ \iff \text{Cov}(x_i\gamma, \bar{x}\gamma) - c\text{Cov}(\bar{x}\gamma, \bar{x}\gamma) &= 0 \\ \iff \text{Cov}(x_i\gamma, \bar{x}\gamma) &= c_i\text{Var}(\bar{x}\gamma) \\ \iff c_i &= \frac{\text{Cov}(\beta_i, \beta_{ATE})}{\text{Var}(\beta_{ATE})} \end{aligned}$$

The random variable in this context is  $\gamma$  (the coefficients on the  $x$ 's) in this case  $\text{Var}(\beta_{ATE})$  is a measure of how uncertain one was about what  $\beta_{ATE}$  would be before doing the trial.

$c$  is the equation for a regression of  $\beta_i$  on  $\beta_{ATE}$ . In other words, we take a bunch of patients with characteristics  $x_i$  and we keep redrawing the gammas from our prior distribution the we ask how correlated  $\beta_i$  and  $\beta_{ATE}$  are. If they are more correlated (as they would be for patients where the  $x_i$  are closer to  $\bar{x}$  we update more.

To compute  $c$ , we just need to know  $x_i$ ,  $\bar{x}$ , and the distribution of  $\gamma$ .

Suppose we want to design the trial to minimize:

$$\min E[(\beta_i - \beta_{i,post})^2]$$

## Simple Cases

1. There is just one  $x$  and it is binary (old v young). Can it be solved analytically?
2. Can you solve a 2-dimensional case?

## Thoughts

$$\min E[(\beta_i - \beta_{i,post})^2]$$

ought to also be minimized when

$$\min(\beta_i - \beta_{i,post})^2$$

is minimized