Model-based matching for causal inference in observational studies

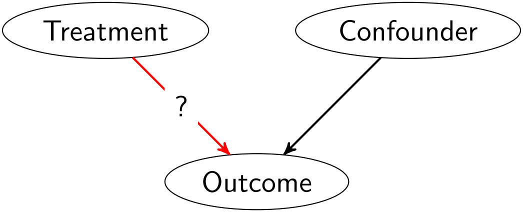
Kellie Ottoboni
with Philip B. Stark

Department of Statistics, UC Berkeley

March 14, 2016

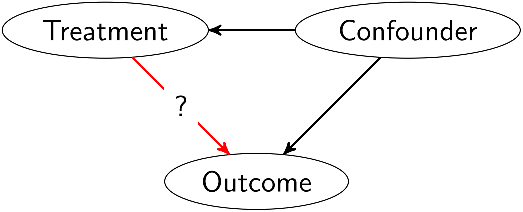
Observational Studies vs Experiments

- **Problem:** Estimate the causal effect of a treatment on outcome of interest
- In randomized experiments, treatment is assigned to individuals at random.
- In observational studies, the way individuals select into treatment groups is unknown.



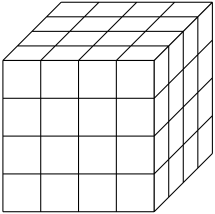
Observational Studies vs Experiments

- **Problem:** Estimate the causal effect of a treatment on outcome of interest
- In randomized experiments, treatment is assigned to individuals at random.
- In observational studies, the way individuals select into treatment groups is unknown.



Matching

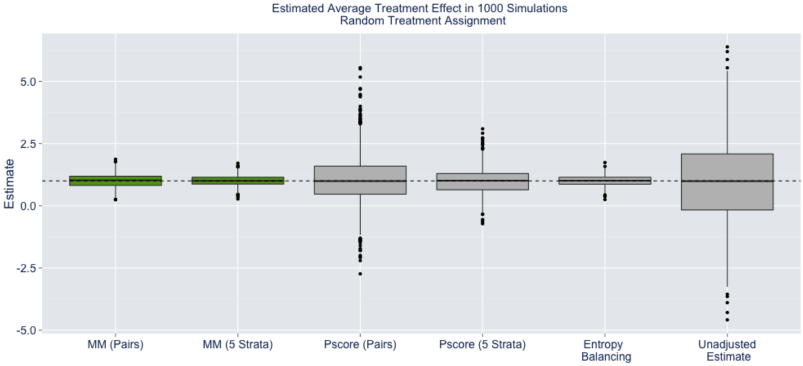
- **Ideal:** group individuals by important confounders to estimate subgroup treatment effects and then average over subgroups
- **Reality:** many covariates, perhaps continuous, make it difficult to stratify



- **Solution:** use a one-dimensional score to match or group individuals
- Stratify on $\hat{Y}$, the "best" prediction of the response based on all covariates except for the treatment

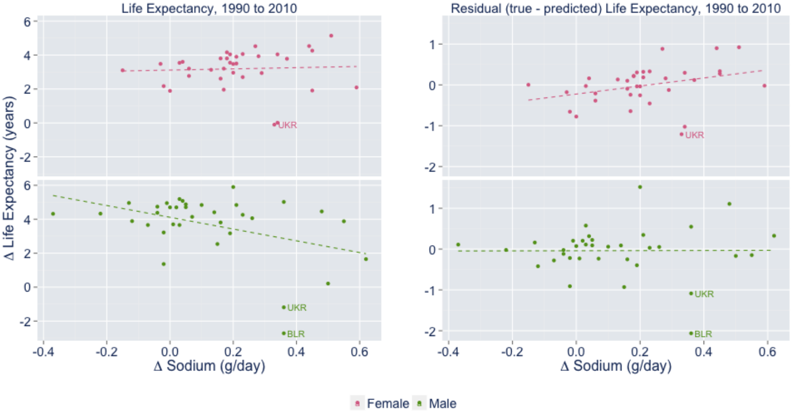
Estimation

- Under standard assumptions, we can estimate the **average treatment effect** nonparametrically using the difference in average residuals, $Y - \hat{Y}$, between treated and controls



Hypothesis testing: example

- Test the **strong null hypothesis** of no relationship between salt consumption and country-level mortality rates
- Control for other health predictors: look at residuals $Y - \hat{Y}$
- Stratified permutation test requires no distributional assumptions



Future Directions

- Do different test statistics give greater power when the treatment effect is nonlinear?
- What is the optimal way to stratify?
- How to quantify uncertainty – standard errors and confidence intervals?