Lecture 2

Quantitative methods for addressing selection bias due to confounding

Content

- Causal inference, purpose, motivation
- Propensity score matching
- Genetic matching
- Sensitivity analyses
- Latest developments

Statistical Methods for addressing confounding

- Causal Framework and estimands
- Assume no unobserved confounding
  - Regression adjustment
  - Matching methods
    - Propensity score matching
    - Genetic Matching
- Allow for observed and unobserved confounding:
  - Instrumental variable estimation
  - Regression discontinuity design
  - Sensitivity analysis for unobserved confounding

Problem of causal inference (Rubin 1977, Holland 1986)

- T_i is treatment indicator: 1 treatment group, 0 control
- Interested in causal relationship between T_i and Y_i
- Each individual, *i* faces potential outcomes Y_{i0} and Y_{i1} under control and treated states
- Ideally observe treatment effect for each individual $\tau_i = Y_{i1} - Y_{i0}$
- BUT cannot observe both outcomes
- **Objective of methods: impute missing potential outcome**

Which estimand?

Which population are we interested in?

- Average treatment effect (ATE):
  - Characteristics of treated and controls
- **Average treatment effect for treated (ATT)**

i	T	Y_0	Y_1	$Y_1 - Y_0$
1	1		8	
2	1		4	
3	1		8	
4	0	8		
5	0	10		
6	0	7		

Which estimand?

Which population are we interested in?

- Average treatment effect (ATE):
  - Characteristics of treated and controls
- **Average treatment effect for treated (ATT)**

i	T	Y_0	Y_1	Y_1-Y_0
1	1	5	8	3
2	1	3	4	1
3	1	6	8	2
4	0	8	9	1
5	0	10	10	0
6	0	7	6	-1

Which estimand?

Which population are we interested in?

- Average treatment effect (ATE):
  - Characteristics of treated and controls
- **Average treatment effect for treated (ATT)**

i	T	Y_0	Y_1	Y_1-Y_0
1	1	5	8	3
2	1	3	4	1
3	1	6	8	2
4	0	8	9	1
5	0	10	10	0
6	0	7	6	-1

ATT= 2

ATE= 1

Regression for average treatment effects

Want to estimate incremental cost-effectiveness

INTEREST: effect of treatment on mean costs, QALYs

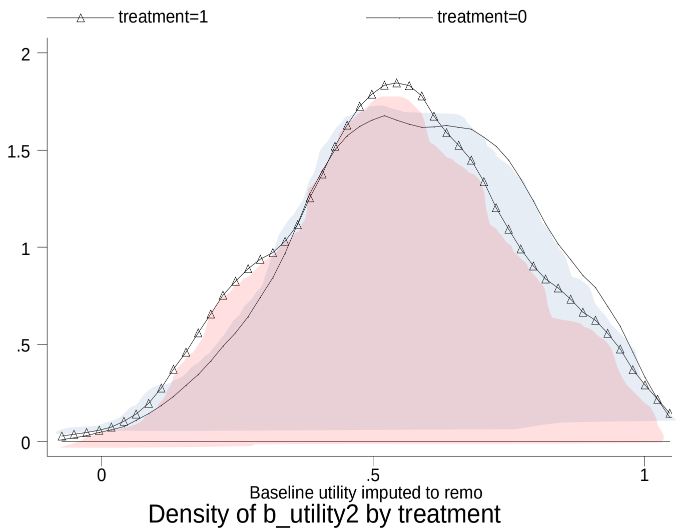
- Regression controls for observed covariates through modelling the outcome

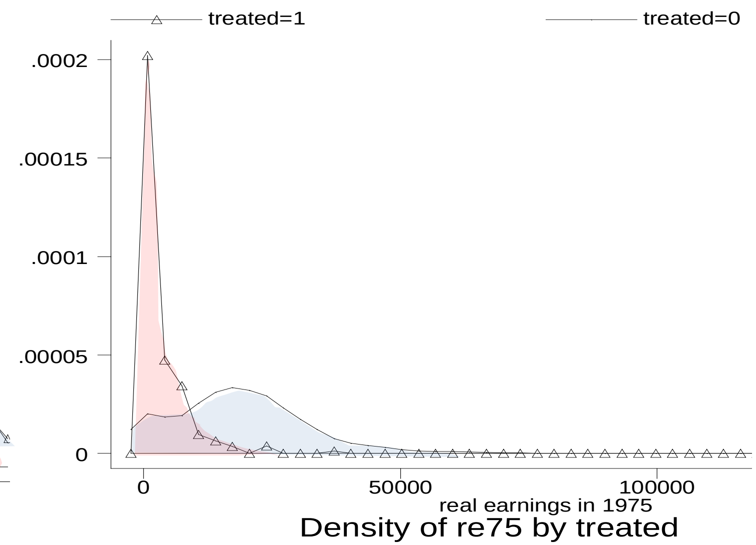
- Estimates regression model for the mean outcome $E[Y \mid T, X]$
  - E.g. $E[Y \mid T, X] = \beta_1 T + X_1 \beta_2 + X_1^2 \beta_3$

- Predicts both potential outcomes for each individual
  - $\hat{Y}_{i0}$ as $E[Y_i \mid T = 0, X_i]$
  - $\hat{Y}_{i1}$ as $E[Y_i \mid T = 1, X_i]$

- Estimates ATT (ATE) average prediction differences among treated (everyone)

Regression challenge: overlap

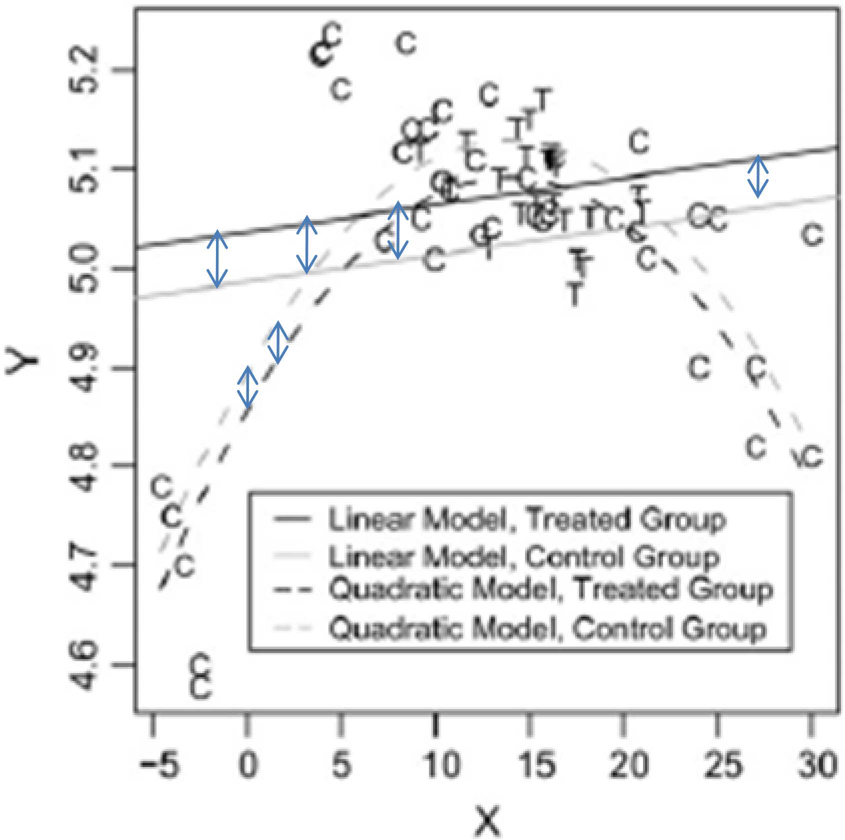




good overlap:
Baseline utility

weak overlap
Baseline earnings:

Example: weak overlap, sensitivity to functional form



linear model:
treatment effect 0.05

quadratic model:
treatment effect of -0.04

Source: Ho et al. 2007

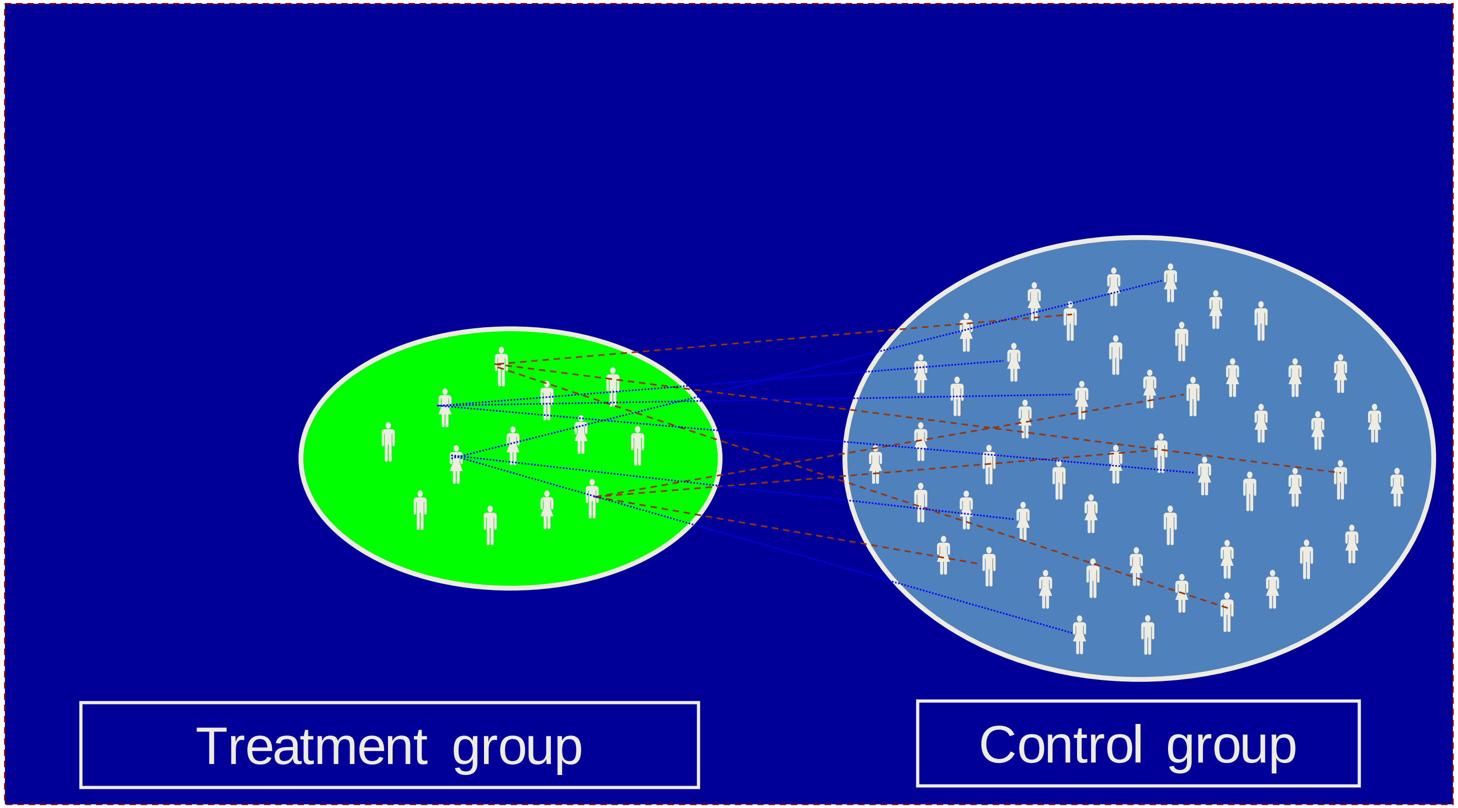
Matching: motivation

- Regression correct functional form never known
- Incorrect relationship: parameter to endpoint
- Or treatment effect multiplicative not additive
- Biased and inconsistent estimates
- Especially severe when weak overlap (Ho et al. 2007)
- Regression involves extrapolation
- Endpoint variable always in sight

Matching (Stuart 2010)

- AIM: ensure groups are balanced
- Covariates similar between treatment and control groups
  - Means, but also variances et
  - RCT similar baseline covariate distributions
- Imputes missing potential outcome by finding a “similar” individual from control group according to *observed* characteristics
- Key assumptions
  - 1. No unobserved confounders
  - 2. Covariates overlap between groups $0<Pr(T_i=1|x_i)<1$

Intuition behind matching, e.g. for ATT



Require matching method that achieves **best balance** in observed characteristics x_i between treatment and control groups

Pscore: background

- Most non-parametric way match exactly on x
- Only feasible if very few, discrete confounders
- Reduce dimensionality with Pscore methods
- Rosenbaum and Rubin, *Biometrika* 1983
- Google Scholar citations: n=20,157 as of May, 15th, 2018
- Key result: Pscore is a balancing score
  - Sufficient to ‘control’ for true Pscore only
  - Matching, subclassification, adjustment, weighting

- Matching performs relatively well (Austin 2009)

Pscore: estimation

$$e(X_i) = \Pr(T_i = 1 \mid X_i)$$

- Model of the probability of treatment, given observed covariates
- Choice of treatment depends on patient, clinician choice
- Matching Pscore can unbiased estimate ATT (Rosenbaum and Rubin 1983)
- If Pscore is correctly specified
  - Pscore generally unknown, must be estimated
  - How do we get correct functional form?
  - Balance can be directly assessed, shows if Pscore is specified correctly
  - Assess balance post matching, modify accordingly
  - Achieving balance on many terms is challenging..

Pscore matching: key stages

- Define target population, estimand of interest (ATT, ATC, ATE)
- Define 'treatment' and 'control' groups
- Assess overlap and if required redefine target population
- Estimate the Pscore
- Check balance, re-estimate the Pscore
- Extract matched data, and estimate treatment effects
- Sensitivity analyses (e.g. regression on matched data)

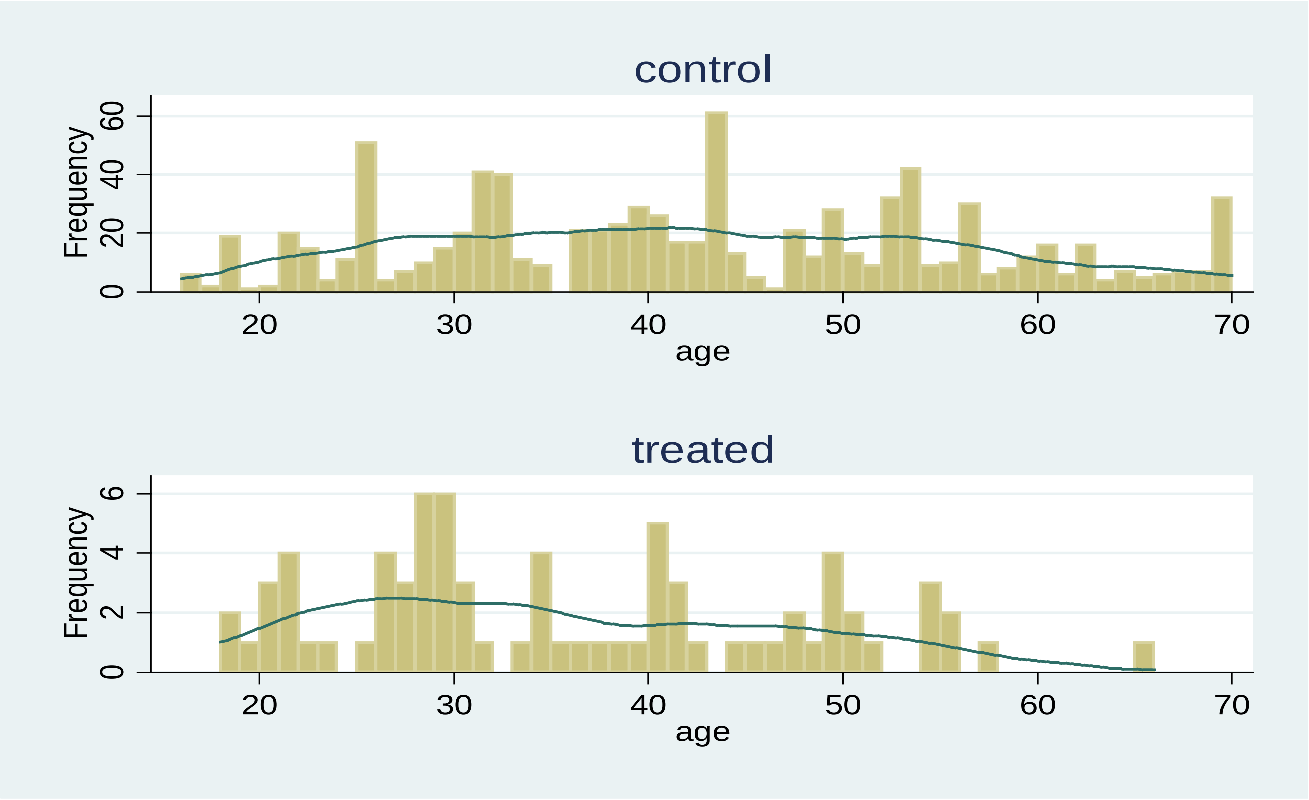
Assessing overlap

- Describe each covariate, treatment versus control
  - e.g. Histograms for continuous variables
- Remedy, apply explicit exclusion criteria
- Excluded from pop. of interest for decision problem
- Can look at distribution of Pscore
- Could drop observations don’t overlap on Pscore
- Unclear then what is being estimated
- Instead consider individual covariates
- Make exclusions explicit, helps interpretation

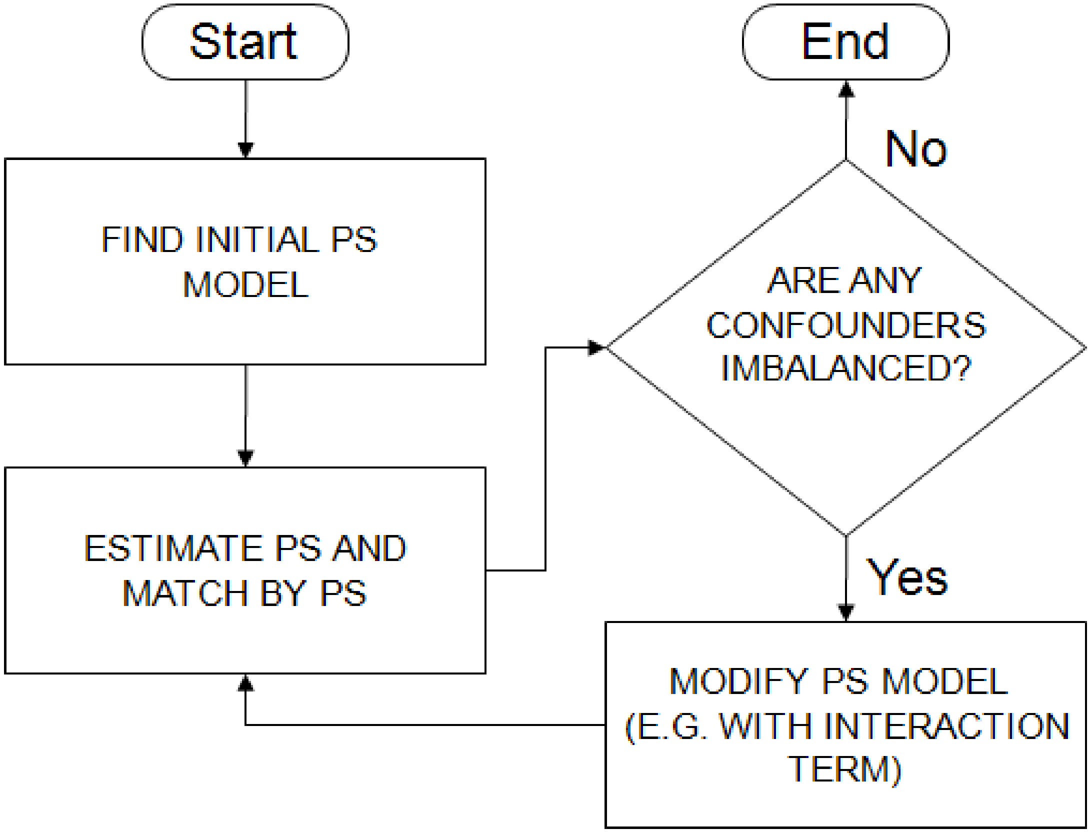
Examples assessing overlap

H1N1: ECMO treatment versus control

Noah et al, JAMA 2012



Iterative process for specifying the Pscore



Assessment of balance

See Austin (2009)

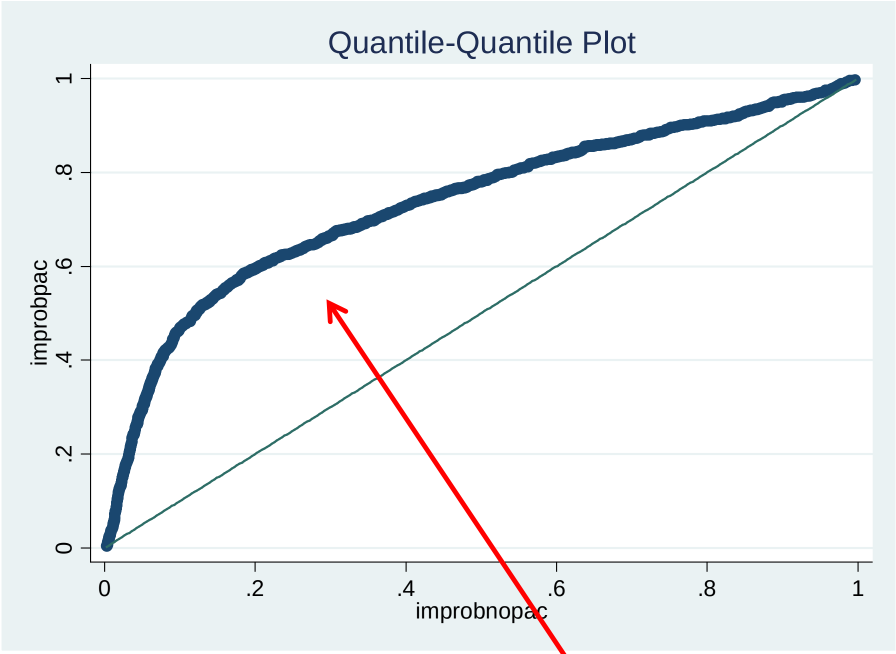
- Should not use standard t-tests
- Considering means necessary but insufficient
- Appropriate balance measures:
  - sample size invariant
  - consider moments of the distribution beyond mean
- Standardised differences- means divided by pooled SD
- Quantile-Quantile plots (continuous variables)
- P values from non-parametric tests

The importance of checking balance

- Pulmonary artery catheterization (PAC)
- Invasive monitoring device used in ICU
- Observational study using Pscore
- PAC higher mortality & cost vs. No PAC (Connors 1996)
- PAC use declined subsequently
- Further observational study undertaken by Harvey et al, 2005,
- Critical care data from ICNARC (1052 PACs, 32,000 no PACs)
- 65 baseline covariates
- Later re-visited by Sekhon and Grieve 2012

Empirical Quantile-Quantile Plot (eQQ)
PAC versus no PAC
Baseline probability death (IMProb)

before Pscore matching

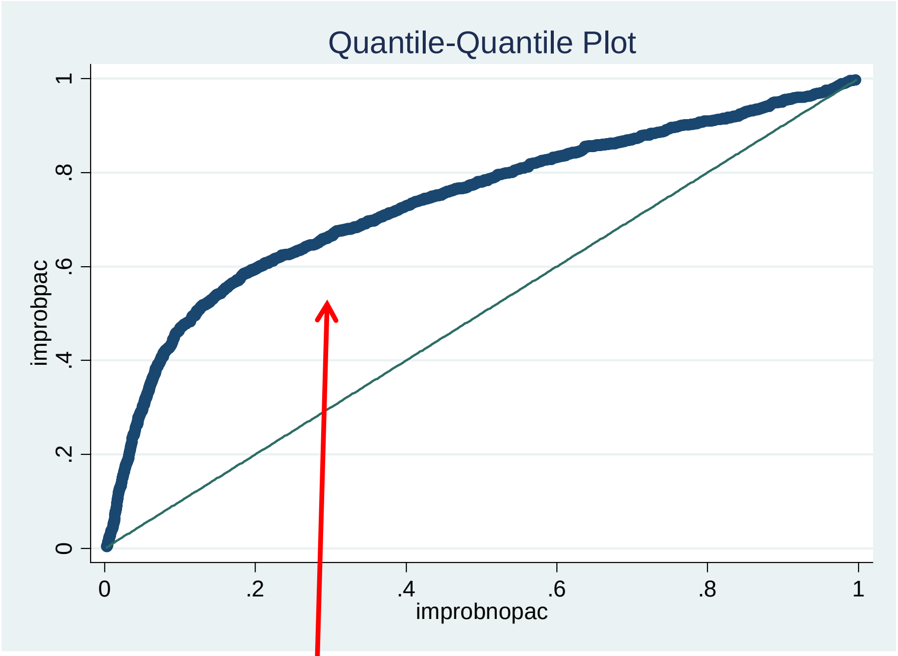


Want the gap to be small

i.e linked p value to be large

Empirical Quantile-Quantile Plot (eQQ)
PAC versus no PAC
Baseline probability death (IMProb)

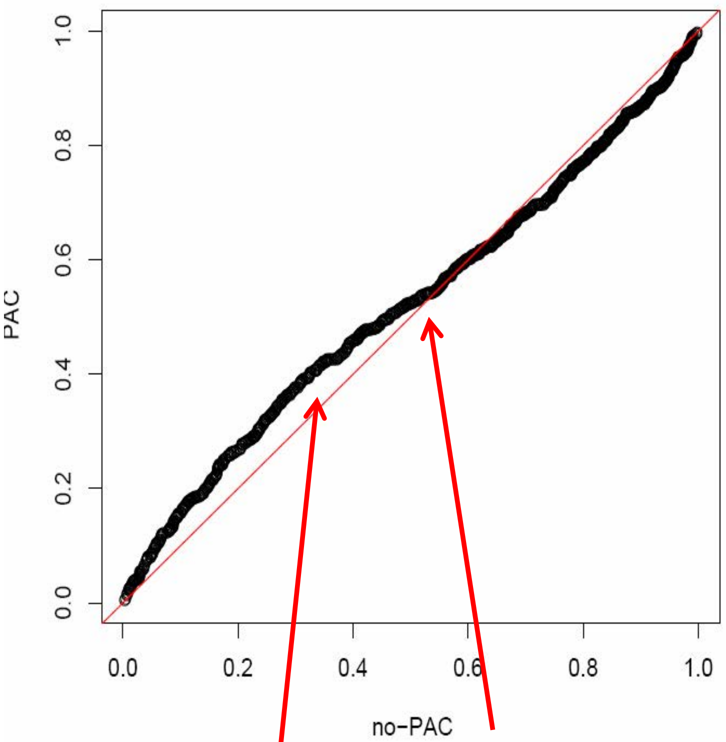
before Pscore matching



Want the gap to be small

i.e linked p value to be large

after Pscore matching



Means balanced

Still a gap, albeit smaller

IPW estimator

- Propensity score: $p(X) = \Pr(T=1 \mid X)$
- Inverse probability of treatment weighting (IPW) for the ATE:

  reweighting treated with $\dfrac{T_i}{\hat{p}(X_i)}$

  and control sample with $\dfrac{1-T_i}{1-\hat{p}(X_i)}$

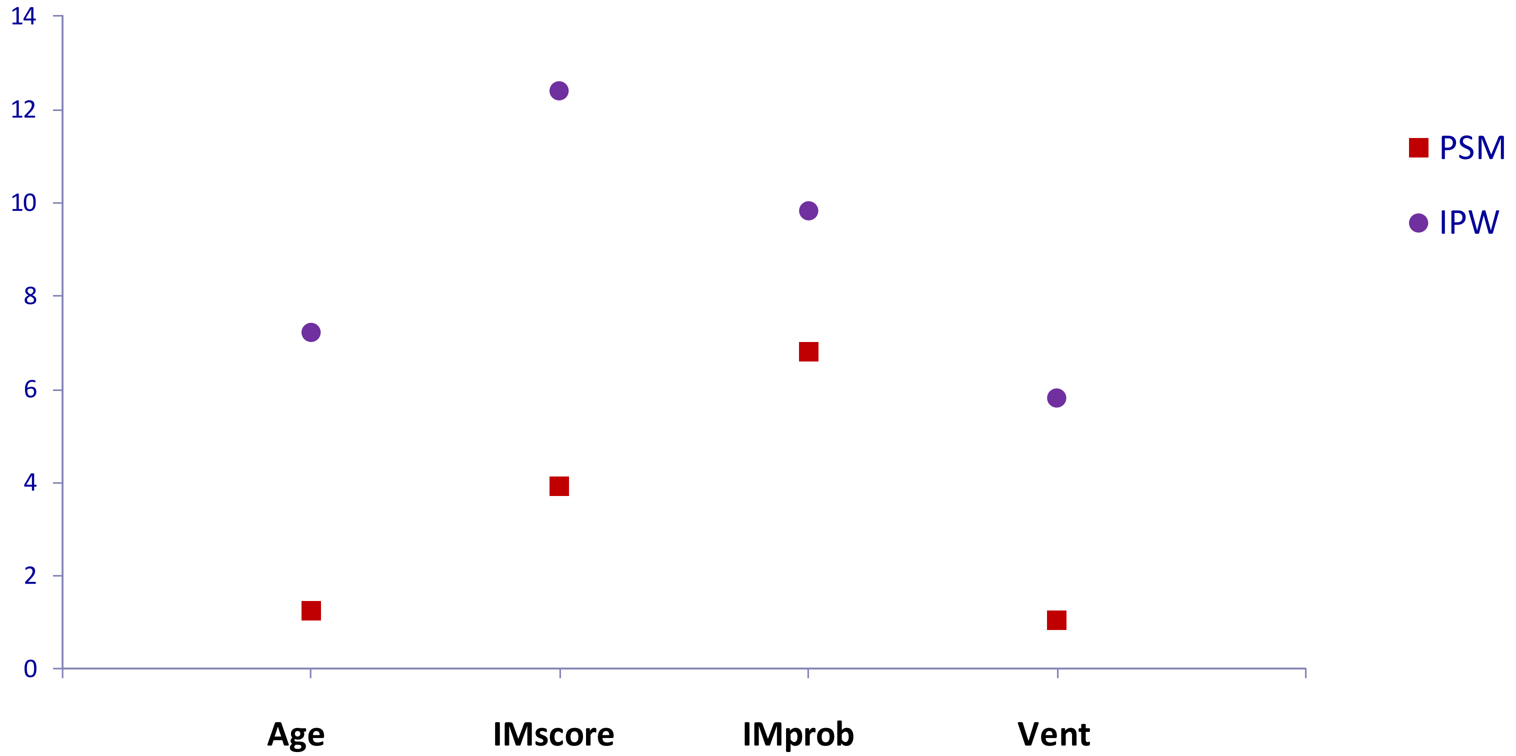
- Theory: if Pscore correct, unbiased + most efficient way to use PS
- Poor overlap -> close to 0 or 1 -> extreme weights –> bias, inefficiency

• Can be combined with regression, in double-robust models (e.g. Bang and Robins, 2005 Biometrics)

• Can allow for time varying treatments (e.g. Marginal structural models, Hernán et al., 2000 Epidemiology)

Xigris for severe sepsis: subgroup with 3-5 organ failures
Covariate balance PSM vs IPW

Standardized differences



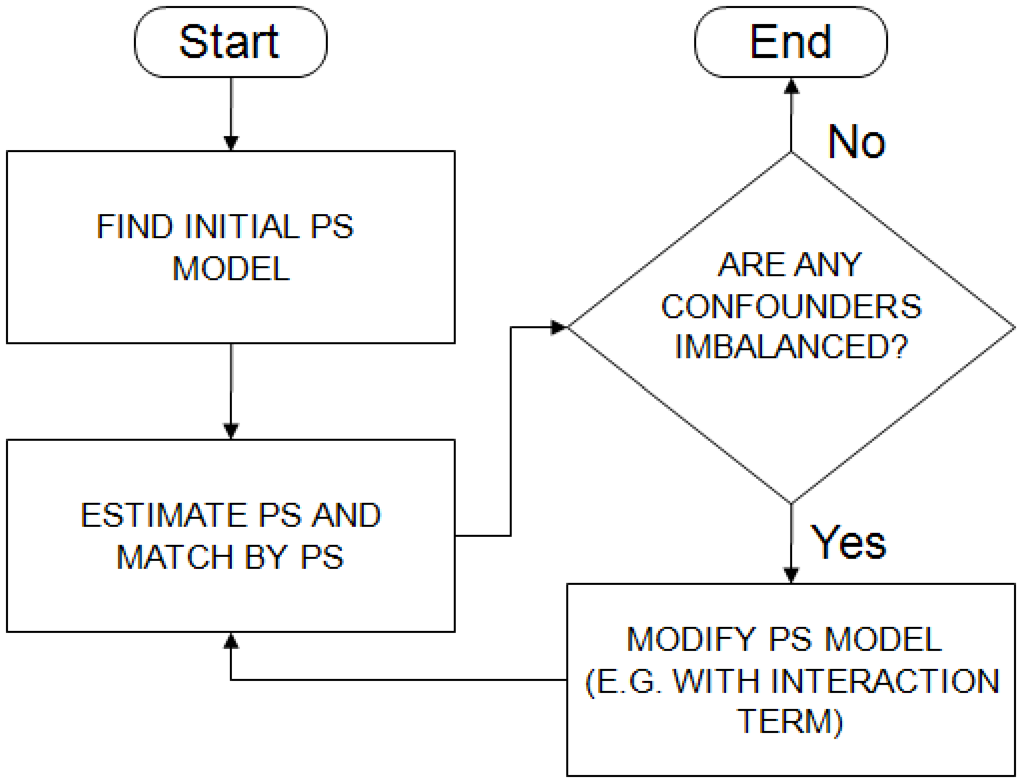
Summary: pscore methods

- Fundamental to define the target population
- Pscore less reliant correct specification outcome regression model.
- Challenge correct Pscore model
- Vital to report full range of balance statistics
- Poor balance pscore matching, consider other pscore approaches
- Inverse probability weighting (IPW) and double-robust estimation appealing alternatives especially with dynamic treatment regimes (see for example Vander Laan and Robins, 2007)
- For now consider other matching alternatives..

Genetic Matching (GenMatch)

- Methods so far assume correct model specification
- Difficult to specify correct Pscore i.e. balance covariates
- In many evaluation settings covariates often non-normal
- **Genetic Matching**: automated search algorithm maximises balance
- Follows principle recommended by Rosenbaum and Rubin (1985)
- Recommendations for Pscore ignored (Austin 2008)
  - Follow iterative process of balance checking
  - In addition, match on underlying covariates
- Can give less bias (Diamond and Sekhon 2010; Sekhon and Grieve 2011, Radice et al., 2011, Kreif et al. 2012)

Iterative process for pscore specification



GenMatch

MOTIVATION 1: Automates cumbersome iteration process

MOTIVATION 2: Focuses on balancing covariates

What is GenMatch?
see Sekhon (2011)

- **Aim:** max balance between treatment and controls
- Automated search algorithm maximises balance
- Algorithm searches data for ‘best’ matches
- Repeatedly checks balance, then improves balance
- Automated not manual balance checking
- Can match with Pscore *and* covariates
- Maximise *balance* on most important confounders
- As recommended by original developers of Pscore

  (Rosenbaum and Rubin 1985)

Multivariate distance matching

See Glance et al. (2007)

- GenMatch extends other multivariate matching
- Common matching metric Mahalanobis distance (MD):

$$md(X_i, X_j)=\{ (X_i - X_j)\ '\ S^{-1} (X_i - X_j)\}^{1/2}$$

- X_i and X_j vector of covariates for 2 different observations;
- S is sample covariance matrix of X
- Minimise multivariate distance metric for each matched pair -> may not result in optimal balance in matched sample
- Weight according to sample covariance
- Performs badly when covariates are non-normal

GenMatch: Multivariate matching

(see Sekhon 2011, Sekhon and Grieve, 2011, Noah et al, 2011, Pennington et al, 2013, Sadique et al, 2011, Kreif et al, 2012; Radice et al, 2012; Ramsahai et al, 2011)

- GenMatch generalises Mahalanobis distance measure
- $GMD(X_i,X_j) = \{ (X_i - X_j)' (S^{-1/2})' \boldsymbol{W} S^{-1/2}(X_i - X_j) \}^{1/2}$
  - X_i and X_j vector of covariates for 2 different observations;
  - S is sample covariance matrix of X
  - ***W*** is a weight matrix
- Considers many alternative sets of weights
- A genetic algorithm searches data to pick the weights W
- Picks those weights that maximise overall covariate balance
- **Creates matched dataset using optimal weights**

GenMatch: Key Stages

- Specify variables want to match on (*X* matrix)
- Specify variables vital to balance (balance matrix)

THIS DECISION IS KEY. MUST INLUDE ALL CONFOUNDERS VITAL TO BALANCE. THE CHOICE IS NOT AUTOMATED BUT IS A JUDGEMENT BY THE ANALYST. MUST CONSIDER A PRIORI REASONING, PREVIOUS LITERATURE. THE CHOICE OF VARIABLES TO MATCH MUST BE ACCORDING TO THOSE **JUDGED** VITAL TO BALANCE.

- Choose balance statistics (e.g. t-tests, KS statistics)
- Specify matching options (e.g. 1 to 1, replacement)
- Ask Genetic Matching to optimise balance

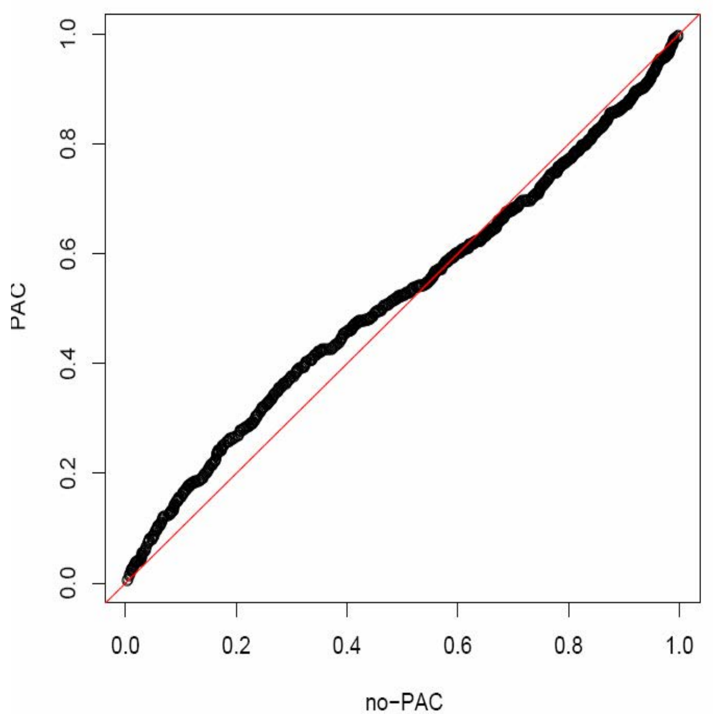
Choosing matching options

General choices (all matching methods)

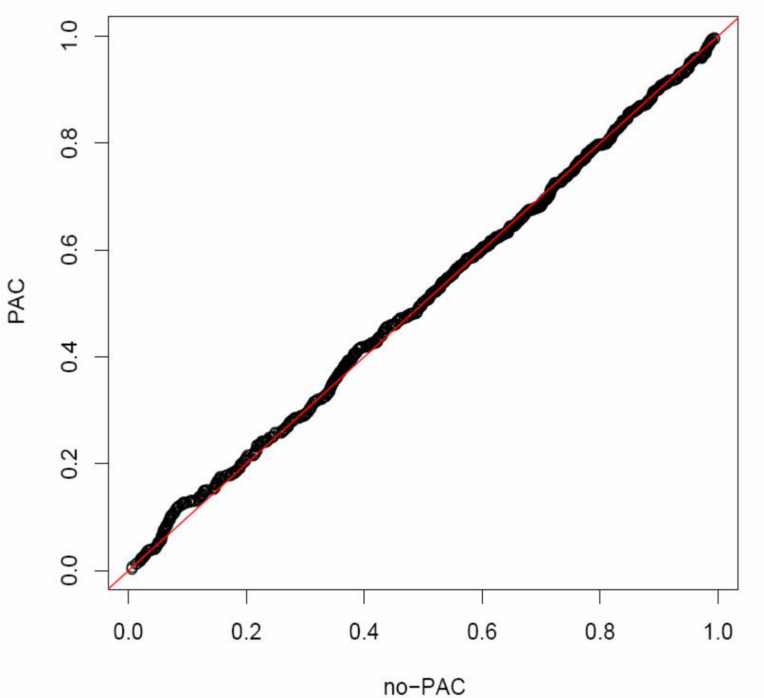
- Matching with versus without replacement
- Matching 1:1 versus 1: n (Austin, 2010)
- Least bias option is 1:1 with replacement (Stuart, 2010)
- “Abadie & Imbens standard errors” allow for dependencies within the matched data (Abadie and Imbens, 2006)
- Inference is conditional on the matched data (Ho et al 2007)

PAC study, see Sekhon and Grieve 2011
Covariate Balance: eQQ-plot
Baseline Probability Death (IMProb) PAC vs. No PAC

Pscore matching



Genetic Matching



Incremental net benefit (INB)
PAC vs. No PAC

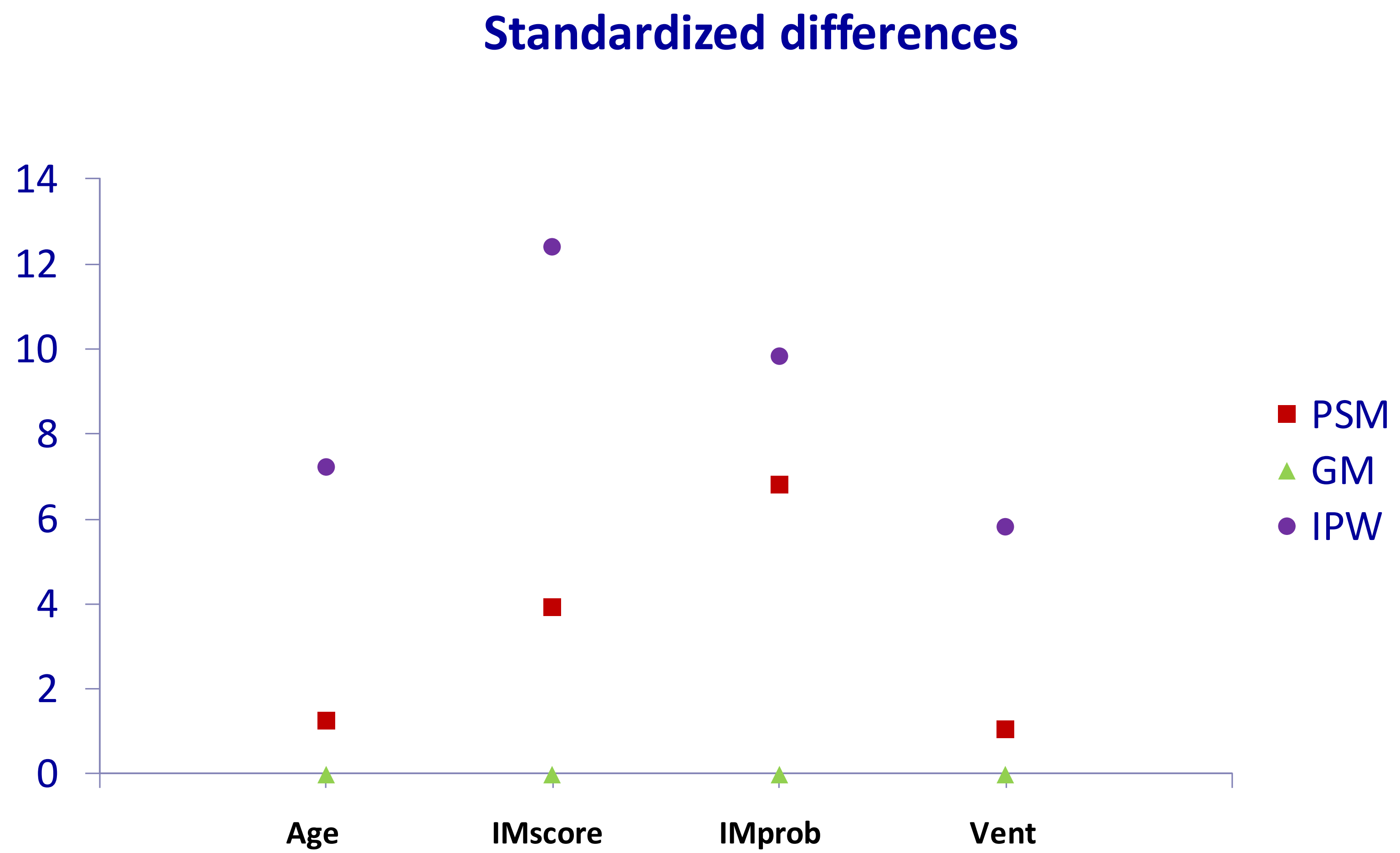
	INB (95% CI)
Pscore matching	-£27,215 (-£38,864 to -£14,154)
GenMatch	-£11,830 (-£24,960 to £834)
RCT	-£3,089 (-£19,234 to £13,265)

λ=£30,000 per QALY

CIs calculated with non-parametric bootstrap

Xigris for severe sepsis: subgroup with 3-5 organ failures

Covariate balance PSM vs IPW vs GM



Xigris for severe sepsis, subgroup with 3-5 organ failures

Cost-effectiveness results

	Using subgroup specific PS mean (95% CI)*		
	Inc cost £	Inc QALY	INB**
Genetic Matching	19,948 (17,610 to 22,286)	1.28 (0.86 to 1.70)	5,690 (-2,543 to 13,924)
IPW	19,023 (15,636 to 22,102)	0.542 (-0.66 to 1.55)	-8,175 (-31,787 to 11,845)
Pscore matching	19,384 (17,696 to 21,071)	0.98 (0.65 to 1.33)	391 (-6,350 to 7,133)

*Non-parametric bootstrap CI

**INB at £20,000 per QALY

GenMatch steps

see Sekhon (2011)

1. Specify the covariates to match on

```
X <- cbind(age,sex,Improb,bloodpr)
```

- can include the Pscore

2. Specify the terms to balance

```
BalanceMatrix<-cbind(age,sex,Improb,bloodpr)
```

- can be identical to X

3. Set GenMatch options

4. Call GenMatch (computational time)

```
gen1<-GenMatch(Tr=PAC,X=X,
    BalanceMatrix=BalanceMatrix,popsize=1000)
```

GenMatch options

see Help for more options and details

- The population size: number of ‘trials’ i.e. possible sets of weights within each ‘run’ or generation
- Larger can be better for balance, 1000 is reasonable: `pop.size=1000`
- The number of generations: the number of ‘runs’ again larger can be better, controlled with `wait.generations` and `max.generations`

Obtaining balance from GenMatch

- Have to first call Match() to extract the Genmatch matched dataset

```
mgen1 <- Match(Tr = pac, X = X, weight.matrix=gen1)
```

- Then use these matched datasets to get balance statistics

```
mb_GM <-MatchBalance(pac ~ IMprob match.out = mgen1, data=
   pacdata, nboots=500)
```

Estimating treatment effects

- Not until satisfied with balance achieved
- Report estimand of interest e.g. ATT
- mean differences in say costs for treated,

```
m_gm1_cost<-Match(Y= totalcost,Tr=treated, X=X,   Weight.matrix =
   gen1, estimand = "ATT")

summary(m_gm1_cost)
```

- Inference allow for joint distribution costs and outcomes
- use non-parametric bootstrap to report uncertainty
- Report inference *conditional* on matched data

What if, I can't get good balance?

- GenMatch maximise balance according to the loss function
- Will improve worst balance of variables in balance matrix
- Can customise loss function according to problem
- For example, prioritise variables according to previous literature, expert opinion, or insights from DAGs
- Ramsahai et al. 2011, drew on expert opinion to define 'high priority'; 'medium priority' and 'low priority' variables.
- Wrote customised loss function, to maximise balance

Matching alone, and in combination

- Balance is key
- Advantages combining matching with regression (Adabie and Imbens 2011)
- Performs at least as well as double robust estimation (Kreif et al 2014)
- Machine learning methods for treatment effect estimation (Kreif et al SMMR 2014)
- Throughout, overarching design cross sectional data
- Assumed no unobserved confounding..

Conclusions

- Causal inference framework requires analyst to define estimand and assumptions
- Matching methods, can be flexible according to causal question, and reduce reliance on parametric assumptions
- Essential define causal assumptions, sensitivity analyses.
- Don't rely on a single method
- Matching methods offer advantage of simplicity, transparency
- Recent extensions broaden range of settings, and offer useful ways of combining matching with regression + other approaches.

References

- Rosenbaum , P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1): 41-55.
- Austin PC (2009). The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. *Med Decis Making*. 29(6):661-77. doi: 10.1177/0272989X09341755.
- Stuart, E. A. (2010). Matching Methods for Causal Inference: A review and a look forward. *Statistical Science* 25(1).
- Kreif, N., et al (2012). Methods for estimating subgroup effects in cost-effectiveness analyses that use observational data *Medical Decision Making*,32(6):750-63.
- Diamond, A. & Sekhon, J. S. 2013. Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. Review of Economics and Statistics. 95(3): 932-945
- Sekhon, J. S. 2011. Matching: multivariate and propensity score matching with automated balance search. Journal of Statistical Software. 42(7)
- Sekhon, J. S. & Grieve, R. 2011. A Matching Method for Improving Covariate Balance in Cost-Effectiveness Analyses. Health Economics 21(6):695-714
- Kreif N et al. (2014). Regression-adjusted matching and double-robust methods for estimating average treatment effects in health economic evaluation. Health Services and Outcomes Research Methodology 13 (2-4), 174-202, 2013.