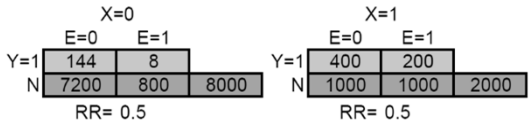
EPID 765
Pharmacoepidemiology

Propensity Scores

1

Are Stratified Tables Collapsible?



Example:
- E = inhaled beta agonists (yes/no)
- Y = Asthma mortality (yes/no)
- X = Asthma severity (high vs. low)

2

2

Collapsibility Rules

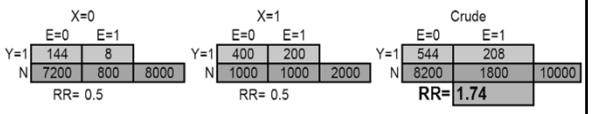
- Notation:
  - Independent of = ⫫
  - Conditional on = |
- Tables are collapsible for RR or RD if either
  - X ⫫ E
    Covariate is independent of exposure = Exposed and unexposed are exchangeable

  or
  - X ⫫ Y | E
    Covariate is not a risk factor for the outcome in the unexposed

3

3

Check Collapsibility Rules



	E=0	E=1
N	8,200	1,800
X (%)	1,000 (12%)	1,000 (56%)

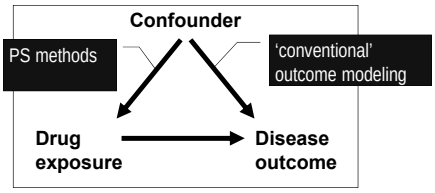
$RR_{XY|E=0}$=400/1000/144/7200=20

Note: Collapsibility requires only one condition to be true
Confounding requires both conditions to be false

4

4

Either/or of Collapsibility Rule: 2 Ways to Address Confounding



5

5

Exchangeability in Cohort Studies

- Randomization
- Matching
- Weighting (e.g., Inverse Probability of Treatment Weighting [IPTW])
- Restriction

Lead to unconfounded "crude" RR and RD in defined populations

Assumption of no unmeasured confounding (more plausible with randomization!)

6

6

Summary Scores

- Reduce dimensionality of confounders
- Peters 1941, Belson 1965, Cornfield 1971
  - Optimize matching (dimensionality)
  - Assess treatment effect heterogeneity
  - Construct validity (ability to predict outcome)
- Miettinen 1976
  - Disease and exposure risk score
- Rosenbaum & Rubin 1983
  - Exposure risk score = propensity score
    - Balancing properties
    - Causal implications
    - Estimation
    - Implementation

7

7

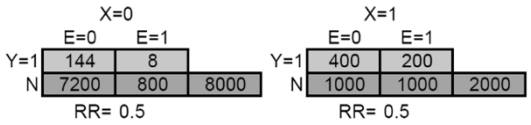
Propensity Score

- Quantifies probability that a person is exposed given his/her observed covariates
- Prob (E = 1 | **X**)
- Estimated from data
  - Multivariable logistic regression
  - CART, neural networks, etc.
- Given same PS exposed and unexposed
  - Tend to have same distribution of covariates used to estimate PS (in expectation, large N!)
  - Are exchangeable
- Unconfounded risk comparisons (RR, RD)

8

8

Example Propensity Score



- Estimate probability of exposure based on covariates:
  - pE(X=0) = 800 / 8,000 = 0.1
  - pE(X=1) = 1,000 / 2,000 = 0.5
- Predicted probability for each observation = PS

9

9

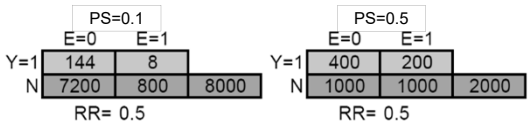
Propensity Score Implementation

- Sub-classification
  - Stratification into e.g., quintiles, deciles, 50-100
- Individual matching
  - Exchangeable pairs of exposed and unexposed
- Weighting
  - Inverse probability of treatment (IPTW)
  - Others (e.g. SMR, matching, overlap)
- Modeling
  - Control for PS in outcome model
  - Least appealing (doubly un-robust!)

10

10

Stratification on Propensity Score



- Estimate RR_{EY} within PS strata (e.g., quintiles)
- Combine stratum-specific RR using Mantel-Haenszel (dummy variables, stratified model)
- Assumption: uniform effects!
- Consider fine stratification (Desai Epidemiol17)

11

11

PS Advantages: 1. Many Covariates & Rare Outcome

- Rule of thumb: 10 outcomes per covariate
- Recent studies: closer to 6
- If less: small sample bias
  - Bias away from 0 with every additional covariate
  - Easily misinterpreted as confounding
- No need for outcome model with PS
- Assuming enough exposed (or unexposed, whichever is the smaller group) we can
  - Control for many covariates (fit a rich PS model)
  - Without running into small sample bias

12

12

PS Advantages: 2. Detect Treatment Barriers

	Propensity quintile				
% with in last 12 mo	1	2	3	4	5
Nursing home stay	82	46	24	12	5
Cardiac arrhythmias	27	22	19	16	14
Congestive heart failure	38	29	24	22	21
Dementia	31	10	3	1	.5
COPD	26	23	21	18	14

Initiation of lipid lowering therapy in enrollees in New Jersey benefits programs age 65+ (Glynn et al, Basic Clin Pharmacol Toxicol 2006) 13

13

PS Advantages: 3. Timing of Covariates and Exposure

- Prediction of exposure can only be made on things prior to exposure (Yogi Berra: it is difficult to make predictions, especially about the future)
- In pharmacoepidemiology often "new user"
- Predict initiating exposure as function of covariates prior to initiation
- Do not include anything after initiation
  - Not relevant for prediction
  - May be affected by exposure (bias!)

14

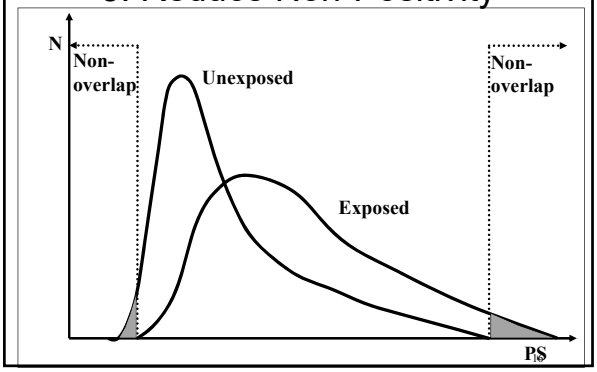
14

PS Advantages: 4. Report Covariate Balance

	Entire cohort				Matched cohort			
	Estrogen		Comparator drugs		Estrogen		Comparator drugs	
N	7,824		37,425		6,957*		6,957	
Age (years), mean (SD)	70.9	(10.4)	79.0	(8.3)	72.7	(9.3)	72.8	(9.3)
White race	5,822	(74.4)	37,912	(83.8)	5,413	(77.8)	5,415	(77.8)
Nursing home	388	(5.0)	4,540	(12.1)	374	(5.4)	355	(5.1)
Diagnoses prior year								
Myocardial infarction	178	(2.3)	1,248	(3.3)	173	(2.5)	157	(2.3)
Ischemic stroke	207	(2.7)	2,186	(5.8)	206	(3.0)	204	(2.9)
PTCA	141	(1.8)	691	(1.9)	128	(1.8)	126	(1.8)
CABG	31	(0.4)	275	(0.7)	31	(0.5)	26	(0.4)
Combined prior CVD	467	(6.0)	3,704	(9.9)	450	(6.5)	449	(6.5)
Angina	1,009	(12.9)	5,870	(15.7)	931	(13.4)	909	(13.1)
Congestive heart failure	1,097	(14.0)	9,970	(26.6)	1,056	(15.2)	1,044	(15.0)
Hypertension	5,385	(68.8)	28,731	(76.8)	4,940	(71.0)	4,932	(70.9)
Diabetes	1,979	(25.3)	12,435	(33.2)	1,821	(26.2)	1,794	(25.8)
Cancer	600	(7.7)	4,623	(12.4)	564	(8.1)	549	(7.9)

15

PS Advantages: 5. Reduce Non-Positivity



16

16

PS Advantages: 6. Causal Contrasts

- Most multivariable models assume uniform treatment effects
- In presence of treatment effect heterogeneity (non-uniform effects, effect-measure modification)
  - Models assuming uniform effects invalid
  - Results do not pertain to any definable population
- Stratum-specific effect estimates valid
- But what about overall effect?
  - Overall estimates based on standardization and weighting valid in presence of heterogeneous effects
  - Because they apply to defined populations

17

Causal Contrasts

- PS matching and weighting allow us to estimate causal contrasts based on counterfactuals:
  - PS matching/SMR weighting:
    - What would have happened to treated if they had been untreated?
    - ATT: average treatment effect in the treated
  - Inverse probability of treatment weighting (IPTW)
    - What would have happened if everyone had been treated vs. no one had been treated
    - ATE: average treatment effect in the entire population
    - Same as RCT

18

Populations Matter!

TABLE 2. Proportion of deaths among 6,269 ischemic stroke patients registered in a German stroke registry between 2000 and 2001 who were treated or not treated with tissue plasminogen activator, according to percentiles of the propensity score for the entire study population

Percentile	Treated (n = 212) Score†	No.	Deaths No.	Deaths %	Not treated (n = 6,057) Score†	No.	Deaths No.	Deaths %	Empirical OR*
99 to 100	0.5809	36	3	8.3	0.5474	26	7	26.9	0.25
95 to <99	0.3143	73	13	17.8	0.2912	178	27	15.2	1.21
90 to <95	0.1393	55	8	14.6	0.1363	258	19	7.4	2.14
75 to <90	0.0585	31	3	9.7	0.0459	910	82	9.0	1.06
50 to <75	0.0115	10	4	40.0	0.0084	1,558	87	5.6	11.27
25 to <50	0.0017	5	2	40.0	0.0014	1,561	54	3.5	18.60
10 to <25	0.0004	2	1	50.0	0.000267	940	36	3.8	25.11
5 to <10		0	0	0	0.000066	313	6	1.9	
1 to <5		0	0	0	0.000027	251	8	3.2	
0 to <1		0	0	0	0.000007	62	1	1.6	
Overall	0.2521	212	34	16.0	0.0262	6057	327	5.4	3.35

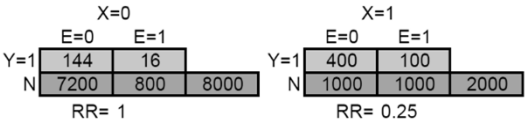
* Propensity-stratum-specific-treatment–mortality odds ratio.
† Mean propensity score in percentile.

Kurth et al., AJE 2006

Average treatment effect in the treated? OR = 1.2
Average treatment effect in the population? OR = 11.0

19

Treatment Effect Heterogeneity

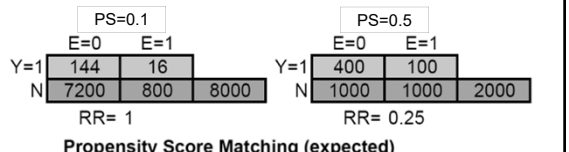


- Assume beta agonists only prevent asthma mortality in patients with severe asthma

20

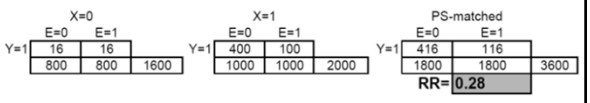
20

Propensity Score Matching



Propensity Score Matching (expected)

pE(X=0)= 0.1 pE(X=1)= 0.5



	E=0	E=1
X (%)	1000 (55%)	1000 (55%)

RR = 0.28 = average treatment effect in the treated (ATT)

21

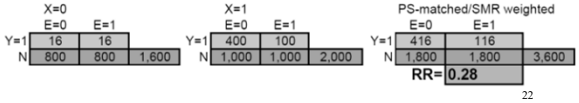
21

ATT (SMR) Weighted Estimate

- All treated w = 1
- All untreated w = PS/(1-PS)
  - sw_{00} = (PS)/(1-PS) = 0.1/(1-0.1) = 0.1111
  - sw_{10} = 1 = 1 = 1.0
  - sw_{01} = (PS)/(1-PS) = 0.5/(1-0.5) = 1.0
  - sw_{11} = 1 = 1 = 1.0
- Multiply all cells (observations!) by these weights
- Collapse tables, calculate RR from collapsed table

Propensity Score Matching / SMR Weighting

w(E=0,X=0)= 0.1111 w(E=1,X=0)= 1 w(E=0,X=1)= 1 w(E=1,X=1)= 1



22

22

ATT (SMR) Weighted

```
data smrw;
 set tables2;
 id=_n_;
 smrw=e + (1-e)*ps/(1-ps);

proc means data=smrw;
 var smrw;
proc means data=smrw;
 var smrw;
 where e=0;

proc freq data=smrw order=data;
 tables e*y/cmh;
 weight smrw;

proc phreg data=smrw
covs(aggregate);
** robust variance **;
 id id;
 model time*y(0) = e/rl;
 freq smrw/notruncate;
run;
```

23

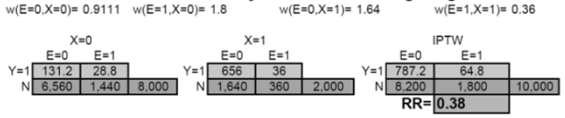
23

IPTW Estimate

- Weight = inverse probability of **actual** exposure
- Stabilized: times marginal prevalence of **actual** exposure
  - sw_{00} = $(1- P_E)/(1-PS)$ = (1-0.18)/(1-0.1) = 0.9111
  - sw_{10} = $(P_E)/(PS)$ = 0.18/0.1 = 1.8
  - sw_{01} = $(1- P_E)/(1-PS)$ = (1-0.18)/(1-0.5) = 1.64
  - sw_{11} = $(P_E)/(PS)$ = 0.18/0.5 = 0.36
- Multiply all cells (observations!) by these weights
- Collapse tables, calculate RR from collapsed table

Inverse Probability of Treatment Weighting

w(E=0,X=0)= 0.9111 w(E=1,X=0)= 1.8 w(E=0,X=1)= 1.64 w(E=1,X=1)= 0.36



	E=0	E=1
X (%)	1,640 (20%)	360 (20%)

RR = 0.38 = average treatment effect in the population (ATE)

24

24

IPTW

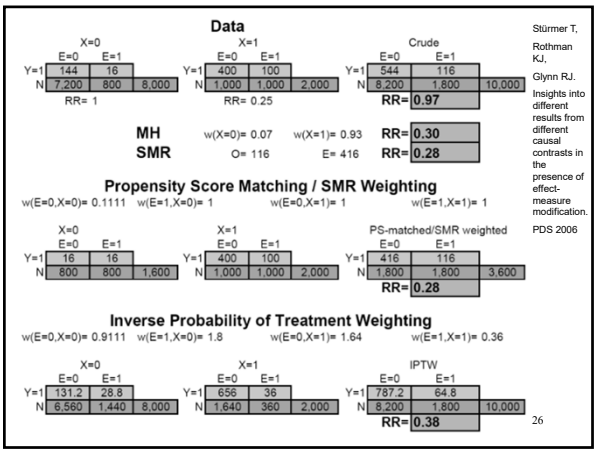
```
data iptw;
 set tables2;
 id=_n_;
 Pe=1800/10000;
 sw=e*pe/ps + (1-e)*(1-pe)/(1-ps);

proc freq data=iptw order=data;
 tables e*y/cmh;
 weight sw;

proc phreg data=iptw covs(aggregate);
** robust variance **;
 id id;
 model time*y(0) = e/rl;
 freq sw/notruncate;
run;
```

25

25



Stürmer T, Rothman KJ, Glynn RJ. Insights into different results from different causal contrasts in the presence of effect-measure modification. PDS 2006

26

26

Practical Issues: PS Variable Selection

- Simulation study with 3 covariates:
  - Confounder, risk factor, instrumental variable

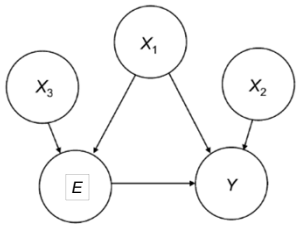


FIGURE 1. Causal diagram for simulation experiment 1.

Brookhart MA et al. AJE 2006

27

27

Practical Issues: Variable Selection

- Simulation study with 1 confounder & 2 different non-confounders
  - 1 predicting exposure but not outcome (i.e., an instrumental variable)
  - 1 predicting outcome but not exposure
- Confounder (X_1) needs to be in PS (dah)
- Predictor of exposure (X_3)
  - Reduced efficiency
  - Increased bias if confounder is not in model (unmeasured confounder)
- Predictor of outcome (X_2)
  - Increased efficiency

Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. AJE 2006

28

28

Which Matching Algorithm?

Biometrical Journal 51 (2009) 1, 171–184 DOI: 10.1002/bimj.200810488

Some Methods of Propensity-Score Matching had Superior Performance to Others: Results of an Empirical Investigation and Monte Carlo simulations

Peter C. Austin[1,2]

treated subjects. Seven of the 8 propensity-score matched samples resulted in qualitatively similar estimates of the reduction in mortality due to statin exposure. 5 → 1 digit matching resulted in a qualitatively different estimate of relative risk reduction compared to the other 7 methods. Using Monte Carlo

The relative risk for 3-year mortality for statin users compared to statin non-users in the entire sample was 0.56 (95% CI: 0.51–0.62). The relative risks estimates using the different propensity-score matched samples were 0.85 (95% CI: 0.75–0.97), 0.87 (95% CI: 0.77–0.98), 0.91 (95% CI: 0.81–1.03), 0.87 (95% CI: 0.77–0.98), 0.87 (95% CI: 0.77–0.98), 0.85 (95% CI: 0.75–0.96), 0.86 (95% CI: 0.76–0.97), and 0.87 (95% CI: 0.77–0.98) when calipers of 0.2 standard deviations of the logit of the propensity score, calipers of 0.6 standard deviations of the logit of the propensity score, 5→1 digit matching, calipers of width 0.005, 0.01, 0.02, 0.03, and 0.1, respectively, were employed. The largest relative risk (0.91) was observed when 5 → 1 digit matching was employed (statin use reduced 3-year mortality by 9%). Furthermore, when 5 → 1 digit matching was used, the 95% confidence interval included 1, indicating that the relative risk was not significantly different from 1. The remaining 7 relative risks ranged from 0.85 to 0.87, and all 7 had confidence intervals that excluded 1 (statin use significantly reduced 3-year mortality by 13% to 15%). Thus, qualitatively similar results were obtained using 7 of the 8 different propensity-score matching methods. A qualitatively different conclusion was obtained when 5 → 1 digit matching was employed.

29

29

Which Matching Algorithm?

Method	Mean % of cases matched	Mean treatment effect	Relative bias	MSE	Mean % of cases matched	Mean treatment effect	Relative bias	MSE
Unmatched		9.90	890%	80.052		9.89	889%	79.543
0.2 SD logit PS	91.7	1.03	3%	0.279	81.8	1.05	5%	0.143
0.6 SD logit PS	95.5	1.24	24%	0.326	87.6	1.47	47%	0.371
5 → 1 digit	90.1	1.07	7%	0.311	80.4	1.00	0%	0.163
0.005 caliper	78.3	1.01	1%	0.382	71.8	0.99	−1%	0.199
0.01 caliper	83.9	1.00	0%	0.339	76.6	0.98	−2%	0.170
0.02 caliper	87.7	1.00	0%	0.310	79.2	0.99	−1%	0.155
0.03 caliper	89.3	1.00	0%	0.296	80.1	1.00	0%	0.150
0.1 caliper	92.5	1.06	6%	0.278	82.7	1.12	12%	0.155
		P_E=0.10				P_E=0.20		

Peter Austin, Biometrical Journal 2009; 1,000 simulation runs

Note: tradeoff %matched vs. bias

30

30

Conclusions PSs

1) No theoretical/practical evidence for intrinsically better control for confounding compared with outcome models
2) Great for rare outcomes and prevalent exposures
3) Help us to think about treatment barriers, timing of confounding, and populations (causal contrasts)
4) Importance of variable selection
   - Avoid entering variables not associated with outcome
   - Report % of exposed for whom unexposed matches were found
5) Look for non-uniform effects over range of PS
   - Use matching, weighting
   - Discuss residual confounding vs. treatment heterogeneity
   - Consider range restrictions, trimming
6) Implementation of PS (modeling, stratification, matching, weighting) minor issue given uniform effects

31

31