

Propensity Scores for Control of Confounding

Dave Glidden
Biostatistics 210

I

Potential Outcomes

- Pursues rigorous “causal” targets
- So-called “marginal” approach
- Comparing factors in standardized population
- Answer depends on confounder mix
- Doesn’t assume homogeneous effect
- Model $f(Y|C,E)$: familiar in that sense
- No clear generalization of hazard ratios

Counterfactuals

- $Y_i(1)$: survival if living kidney
- $Y_i(0)$: survival if cadaveric kidney
- $Y_i(1) - Y_i(0)$: causal effect in i th person
would be rigorous to observe this!
- Only get to see one realization
even in a randomized trial
possible exception: crossover trials

3

Average Causal Effect

- Effect of exposure in population
- A.C.E. $\text{mean}[Y(1)] - \text{mean}[Y(0)]$
 $= 1/n \sum Y_i(1) - Y_i(0)$
- average causal OR or RR can also be defined
- Population is important here
- Causal inference: let estimate A.C.E.
or a related quantity

4

Conditional Independence

- Core assumption of causal inference
- Synonymous with “no unmeasured confounders” beyond C
- Estimate effect of exposure with C
and combine in some sense
- Basis of several methods
regression, inverse weighting, propensity scores, potential outcomes estimation

5

Causal Effects from Regression

- Requires no unmeasured confounders
- Directly model exposure and confounders on outcome
- Not only tool for obtaining causal effects
 - propensity scores
 - stratification
 - inverse weight based methods

6

Propensity Scores

- A different approach to balance confounders
- Useful for **binary exposures**
- Modeling effort on the level of exposure
- Little modeling of effect on the outcome
- Can give rise to marginal and conditional effects

7

Example

- UNOS data
- Survival of pediatric kidney rx
- Effect of living v. cadaveric organ on survival
- 351 deaths
- Model on level of outcome (survival) or
- Model on level of propensity (txtype)

8

Flavor of Propensity Scores

- Two stage approach
- Very intuitive approach
- Has a feeling like randomized trial
doesn't involve complex outcome model
- Can deal with confounding by matching
- Seems so direct -- more causal

Differences with regression more minor than major

9

What is the approach?

- Identify set of possible confounders
- Model exposure in terms of the confounders (e.g., logistic regression)
- Estimate predicted probability for each person (estimated propensity)
- Compared outcome by exposure within strata of the estimated propensity
- Alternative approach to controlling for confounders

10

Causal Inference

- Compare “like with like”
- Exposure holding confounder the same
- Propensity distills confounders into single continuous confounder
 - shows overlap
 - flexibility in outcome model
 - can be more careful about causal target

11

Basic Theorem

- $f(C|\text{propensity}, Z) = f(C|\text{propensity})$
- Two people: one exposed, one not exposed have equal confounders on average
- Reduces confounding to single dimension
- Thing of propensity as our single confounder

12

Variables for Propensity

considerations identical to regression model

- No mediators or colliders
- Confounders must be included
- Don't include predictors of exposure which is not associated with outcome
- Strong predictor of outcome is OK
can reduce residual variation – no help for confounding

13

Adjusting for Instrument

- Predictor of exposure but not outcome
hence, not a confounder
- Provide no protection against confounding
- Propensity will have less overlap
and hence less efficiency
- Worst of both world:
lose efficiency without helping confounding

14

Building Model for Propensity

- Outcome: exposure
- Ideal common exposure/rarer outcome
- Bias/Variance trade off tilts more toward to minimizing bias, less parsimonious
- Include splines and interactions
- Any binary model is OK
extract the predicted probabilities

15

The steps

- Predictor Selection
- Build model for exposure
calculate the prob. of exposure for each person
- Assess overlap/positivity
- Compare exposure by level of propensity
quintiles, splines, matching

16

Propensity Fit: Example

- Splines for age, age of donor year
- Interactions of these with presence of previous transplant
- `logistic txtype prevtx prevtx###c.years*
prevtx###c.agesp* prevtx###c.agedonsp* i.hla`
- `predict prop`
- `xtile group=prop, nq(5)`

17

Quintile Overlap

5 quantiles of prop	txtype		Total
	Living	Cadaveric	
1	1,375	95	1,470
2	1,252	217	1,469
3	666	804	1,470
4	197	1,272	1,469
5	14	1,455	1,469
Total	3,504	3,843	7,347

18

Calculate Effect by Quintile

cc death txttype, by(group)

5 quantiles of p	OR	[95% Conf. Interval]		M-H Weight
1	2.139647	.7233504	5.240355	2.542857 (exact)
2	1.350844	.6446765	2.618979	7.256637 (exact)
3	1.436236	.8626344	2.431353	13.92245 (exact)
4	1.390791	.6236271	3.654351	5.765827 (exact)
5	.9183223	.1352256	39.42104	.9251191 (exact)
Crude	1.604015	1.278398	2.018582	(exact)
M-H combined	1.450304	1.048069	2.006913	
Test of homogeneity (M-H) chi2(4) = 1.00 Pr>chi2 = 0.9102				

19

1st Quintile

death	txttype		Total
	Living	Cadaveric	
No	1,333 96.95	89 93.68	1,422 96.73
Yes	42 3.05	6 6.32	48 3.27
Total	1,375 100.00	95 100.00	1,470 100.00

OR = 2.14, 95% Conf Int (0.7 to 5.2)

Survival if living recipients got a cadaveric organ

20

5th Quintile

death	txtype		Total
	Living	Cadaveric	
No	13 92.86	1,359 93.40	1,372 93.40
Yes	1 7.14	96 6.60	97 6.60
Total	14 100.00	1,455 100.00	1,469 100.00

OR = 0.92, 95% Conf Int (0.14 to 39)

Survival if cadaveric recipients got a living organ

21

Stata Code

- `logistic txtype predictors`
- `predict prop, p`
- `xtile group=prop, nq(5)`
- `cc death txtype, by(group)`
- `logistic death txtype i.group`

22

Testing Interaction

- LR Test
- `logistic death txtype###i.group`
- `est store A`
- `logistic death txtype i.group`
- `lrtest A`

Likelihood-ratio test
(Assumption: . nested in A)

LR chi2(4) = 0.91
Prob > chi2 = 0.9229

23

Why Quintiles

- Overall
 - easily assessed
 - absent/variable effect in quintile
- Interactions
 - seamlessly checks for them
 - easy to calculate average causal effects

make it all transparent

24

Adjusting for Propensity

- Quintile: easy but residual confounding
- Linear: just. don't.
- Spline: flexible but hard to assess interaction
- Inverse Weight: causal rigorous, extreme weight, opaque method
- Matching: default, hard to construct

25

Matching

- Align on individual level based on propensity scores
- Needn't be 1-1
- Needn't be K-1
- Any matching can be wasteful
has arbitrary element
- Extreme version of dialing up quintile to be a finer percentile
- Not a my preferred method

26

Spline

```
. mkspline prop_sp = prop, cubic nknots(7)

. logistic death txttype prop_sp*
```

Logistic regression	Number of obs	=	7347
	LR chi2(7)	=	23.23
	Prob > chi2	=	0.0016
Log likelihood = -1398.3454	Pseudo R2	=	0.0082

death	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
txttype	1.399637	.2433085	1.93	0.053	.9955092	1.967821
prop_sp1	144.3345	1024.966	0.70	0.484	.0001302	1.60e+08
prop_sp2	9.0e-203	5.8e-200	-0.73	0.467	0	.
prop_sp3	.	.	0.72	0.470	0	.
prop_sp4	2.6e-213	1.8e-210	-0.71	0.478	0	.
prop_sp5	4.36994	197.448	0.03	0.974	1.52e-38	1.26e+39
prop_sp6	1.03e+67	1.47e+69	1.08	0.281	1.79e-55	5.9e+188
_cons	.0241336	.0161688	-5.56	0.000	.0064914	.0897228

27

Quintile

```
. logistic death i.txttype i.group
```

Logistic regression	Number of obs	=	7347
	LR chi2(5)	=	23.85
	Prob > chi2	=	0.0002
Log likelihood = -1398.0372	Pseudo R2	=	0.0085

death	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
1.txttype	1.460813	.2433996	2.27	0.023	1.053823	2.024982
group						
2	1.301893	.2544977	1.35	0.177	.8875235	1.909723
3	1.274882	.2676887	1.16	0.247	.8447766	1.923968
4	1.07496	.2522625	0.31	0.758	.6786397	1.702727
5	1.480443	.3480316	1.67	0.095	.9338694	2.346915
_cons	.0327919	.0048379	-23.16	0.000	.0245576	.0437872

28

Causal Effects

- p_0 = Probability of death if everyone got living
- p_1 = Probability of death “ “ got cadaveric
- Can be estimated by “margins” based on a particular model
- $p_1 - p_0 = ACE$
- $p_1/p_0 = \text{Causal RR}$
- $p_1(1-p_0)/(1-p_1)p_0 = \text{Causal OR}$

29

Marginal Effects

Margined Over Quintile

```
. margins txttype
```

	Margin	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
txttype						
0	.0386335	.0041469	9.32	0.000	.0305057	.0467613
1	.0554319	.0046479	11.93	0.000	.0463223	.0645416

```
. margins, dydx(txttype)
```

	dy/dx	Delta-method Std. Err.	z	P> z	[95% Conf. Interval]	
1.txttype	.0167985	.0072948	2.30	0.021	.0025008	.0310961

30

Code for ACE Risk Difference

- `logistic death i.txttype i.group`
fits propensity model
- `margins txttype`
calculate predicted
- `margins dx dy(txttype)`
calculate different

31

Code for Causal RR

- `logistic death i.txttype i.group`
could be any model for the data
- `margins txttype, post`
post saves some key results
- `nlcom (RR: _b[1.txttype] / _b[0.txttype])`
SE can a little off
- `nlcom (logRR: log(_b[1.txttype]) -
log(_b[0.txttype]))`
then exponentiate -- always better SE/CI

32

Causal RR -- Quintile Fit

```
logistic death i.txttype i.group
margins txttype, post
```

```
nlcom (risk_ratio: _b[1.txttype]/_b[0.txttype] )
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
risk_ratio	1.434816	.228147	6.29	0.000	.9876558 1.881976

```
logistic death i.txttype i.group
margins txttype, post
```

```
. nlcom (logRR: log(_b[1.txttype]) - log(_b[0.txttype]))
```

```
logRR: log(_b[1.txttype]) - log(_b[0.txttype])
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
logRR	.3610364	.1590079	2.27	0.023	.0493867 .6726861

marginal RR = 1.43 (1.05, 1.96)

33

Causal OR -- Quintile Fit

```
logistic death i.txttype i.group
margins txttype, post
```

```
. nlcom (OR: (_b[1.txttype]*(1-_b[0.txttype])) / (_b[0.txttype]*(1-_b[1.txttype])) )
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
OR	1.460333	.2433062	6.00	0.000	.9834615 1.937204

```
logistic death i.txttype i.group
margins txttype, post
```

```
nlcom (OR: log(_b[1.txttype])+log(1-_b[0.txttype])-log(_b[0.txttype])-log(1-_b[1.txttype])) )
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
OR	.3786644	.1666101	2.27	0.023	.0521146 .7052142

marginal OR = 1.46 (1.05, 2.02)

34

Usual Regression

Basic Model

```
. logistic death txttype prevtx year age age_don i.hla
```

```
Logistic regression                                Number of obs   =          7347
                                                    LR chi2(11)    =          109.57
                                                    Prob > chi2    =           0.0000
Log likelihood = -1355.1776                      Pseudo R2      =           0.0389
```

death	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
txttype	1.352607	.2104378	1.94	0.052	.9971076	1.834852
prevtx	1.025545	.148438	0.17	0.862	.7722383	1.361941
year	.8667789	.0149232	-8.30	0.000	.8380179	.8965269
age	.9693932	.0102048	-2.95	0.003	.9495972	.989602
age_don	.9976164	.0042928	-0.55	0.579	.9892381	1.006066
hlamat						
1	1.118378	.2534091	0.49	0.621	.7173287	1.74365
2	.9684949	.2219639	-0.14	0.889	.6180381	1.517677
3	.8122357	.1911163	-0.88	0.377	.51215	1.288151
4	.8067894	.2186669	-0.79	0.428	.4743032	1.372348
5	.9442997	.3425764	-0.16	0.874	.4637757	1.922701
6	.89599	.3324897	-0.30	0.767	.4329461	1.854268

35

Usual Regression

Spine/Interaction Model

```
logistic death txttype prevtx prevtx##c.yearsp* prevtx##c.agesp* prevtx##c.agedonsp* i.hla
```

```
Logistic regression                                Number of obs   =          7347
                                                    LR chi2(32)    =          162.11
                                                    Prob > chi2    =           0.0000
Log likelihood = -1328.9077                      Pseudo R2      =           0.0575
```

death	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
txttype	1.472337	.2414501	2.36	0.018	1.067623	2.030471
prevtx	6.5e+189	3.3e+192	0.86	0.389	1.8e-242	.
yearsp1	.975281	.1156388	-0.21	0.833	.7730419	1.230429
yearsp2	1.058858	.8618707	0.07	0.944	.2147788	5.220161
yearsp3	.2970671	1.00785	-0.36	0.721	.0003846	229.4555
yearsp4	3.538337	15.02097	0.30	0.766	.0008615	14531.97
prevtx#c.yearsp1						
1	.8033063	.2045974	-0.86	0.390	.4876241	1.323358
prevtx#c.yearsp2						
1	3.321273	6.292198	0.63	0.526	.0810385	136.1187

and more!

36

20 Strata

```
xtile group3=prop, nq(20)
logit death txttype, i.group3
```

```
logistic death txttype i.group3
```

Logistic regression

```
Number of obs   =      7347
LR chi2(20)     =      50.29
Prob > chi2     =      0.0002
Pseudo R2      =      0.0178
```

Log likelihood = -1384.8183

death	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
txttype	1.427705	.2478669	2.05	0.040	1.01592	2.0064
group3						
2	.326775	.1705254	-2.14	0.032	.1175046	.9087467
3	.7250355	.292959	-0.80	0.426	.3284139	1.600653
4	1.137948	.4122547	0.36	0.721	.5594356	2.314701
5	1.500861	.5153622	1.18	0.237	.765697	2.941874
6	.8503287	.3286302	-0.42	0.675	.3986753	1.813654

ETC

37

Models

Model	Risk Diff	Marginal OR	Cond. OR
Prop: Quint	0.017	1.46	1.46
Prop: Spline	0.015	1.4	1.4
Reg: Simple	0.014	1.35	1.35
Reg: Spline+Int	0.017	1.47	1.46

38

Regression	Propensity Score
assess confounding	assess confounding
draw DAG	draw DAG
assess predictor size	assess predictor size
select confounders	select confounders
model for outcome	model exposure
check overlap	check overlap
	check for interaction
	compare outcome exposure adjusted for propensity

39

Major Differences

- Propensity scores (PS) only useful for binary exposure
- Model size can be very different
PS great for rare outcome/common exposure
- PS involves less modeling of the outcome
- More choices for controlling for propensity
- PS involves many ad-hoc choices

40

Final Recommendations

- Useful for binary variable or possibly ordinal variable
- Switches modeling to exposure *rather than outcome*
 - Advantage if exposure is common, outcome is rare
- Makes overlap transparent
- Analysis for like “randomized trial” less modeling of the outcome
- Often similar results₄₁