Propensity Scores for Control of Confounding

Dave Glidden
Biostatistics 210

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

Propensity Fit: Example

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

Spline

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

. logistic death txtype 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]
-------------+----------------------------------------------------------------
      txtype |   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
------------------------------------------------------------------------------
```

Quintile

```
. logistic death i.txtype 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.txtype |   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
------------------------------------------------------------------------------
```


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 = *Causal RR*
- $p_1 (1-p_0)/(1-p_1)\, p_0$ *Causal OR*

Marginal Effects

Margined Over Quintile

```
. margins txtype

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

. margins, dydx(txtype)

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


Code for ACE Risk Difference

- logistic death i.txtype i.group
  fits propensity model
- margins txtype
  calculate predicted
- margins dxdy(txtype)
  calculate different

Code for Causal RR

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

Causal RR -- Quintile Fit

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

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

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


logistic death i.txtype i.group
margins txtype, post
. nlcom (logRR: log(_b[1.txtype]) - log(_b[0.txtype]))

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

------------------------------------------------------------------------------
             |      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)

Usual Regression

Basic Model

```
. logistic death txtype 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]
txtype	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

Usual Regression

Spine/Interaction Model

```
logistic death txtype 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]
txtype	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!

20 Strata

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

 logistic death txtype i.group3

Logistic regression                               Number of obs   =       7347
                                                  LR chi2(20)     =      50.29
                                                  Prob > chi2     =     0.0002
Log likelihood = -1384.8183                       Pseudo R2       =     0.0178

------------------------------------------------------------------------------
       death | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      txtype |   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
```


Models

Model	Risk Diff	Marginal OR	Cond. OR
Prop: Quint	0.017	1.46	1.46
Prop: Spline	0.015	1.40	1.40
Reg: Simple	0.014	1.35	1.35
Reg: Spline+Int	0.017	1.47	1.46

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

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

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 results

Models for Prediction

Dave Glidden

Distinct Objectives

for a regression model

1. Developing a Model for Prediction/ Stratification
2. Adjusted Effect of Single Variable
3. Identifying Multiple Predictors of Outcome
4. Adjustment to Improve Efficiency in RCT

Example: Prediction

- Cohort: 11,701 community-dwelling elders
- Predictors: age, co-morbidities, functional status
- 42 candidate predictors
- Follow-up: mortality over 4 years
  1,361 deaths
- Who is at the highest risk of death?

Lee, SJ et al. JAMA. 2006;295:801-808.

Study Context

- Objective: inform patients/risk stratification
- Predictors obtained from patient report
- Wanted index "as simple as possible"
- Combine functional measure + co-morbidities
- Doesn't state if high or low risk more important: important for grouping

Statistical Objective

- Develop a regression model
- Discriminates between high and low risk
  how to quantify discriminatory ability?
- Pure prediction is not enough
  want some parsimony/interpretation
  external validity
- Implications for predictor selection
  all 42 significantly associated with death

Getting Best Model

- Need a measure of model predictiveness
- Sift through large # models
- Handle optimism/overfitting
  "how many predictors can the data handle?"
- Consider what will validate externally

Finding Our Model

Look at every possible model and choose the best one in terms of prediction

1. Infeasible in many cases
2. Susceptible to "overfitting"
3. Places no value on simplicity or plausibility

Measures of Prediction

- Log-likelihood
  measure how close model is to the data
  how max. likelihood methods to select model
  analogous to a sum of squares in linear model
- Binary data: C-statistic
  area under ROC curve
  "proportion of person with event has higher risk than someone without event"

A good model optimizes these

Best Subsets

- Consider all possible models
- Each variable may be in or out
  2 possibilities
- Repeated for each predictor
  total number of predictors p
- 2 x .... x 2 (p times) = 2^p
- Increases very fast 2^{42} = 4.4 trillion models
  trillion seconds > 31,000 years
  cut predictors in half: 2.1 million models

Backward Selection

1. Start all variables in the model
2. Fit model
3. Drop term with highest p-value
4. Repeat 2 and 3 until all p-value less than some threshold (e.g., $p < 0.05$)

what is the right threshold? Avoid overfitting

Stepwise Model Selection

- Reduces the model selection immensely
- 42 predictors => 42 possible models
  43 predictors => 43 possible models
- Gives a logical sequence for fitting
  probably don't need to fit all 42
  fit until some criterion is reached
- But will not select the best model

Overfitting

Overfitting is a broad term that refers to entering too many terms into the model -- for the amount of available data.

This produces unstable, variable estimates. The extreme ones *tend* to be both spurious and also *tend* to get the most focus.

Hence, results are frequently unreplicable.

Testimation Bias

The combination of testing a hypothesis and then estimating a quantity, resulting in estimation of an effect which is stronger than is true. One example the process of estimating regression coefficients only among significant predictors.

Example

- Take 5% sample of Lee data
- Fit all variables to data subset
- 0.05 criterion with backward selection
- Would results be reliable?

Model Replication

mortality model

	Odds Ratio	
Variable	test data	full data
hrsdm	2.9	1.8
bmicat	3.3	2.2
hrsmemory	45.7	2.7
hrsarth	7	1.3
hrscad	4.2	1.4
hrslung	4.8	2.3
meals	5.4	2.7
walkjog	9.5	3.8

yellow: test result lies outside CI for full data

Poor Validation

- Odds Ratios consistently overestimated
  consistent overestimation -> bias
- Many estimates outside CI
- More extreme OR -> more overestimated
- Artifact of model selection procedure
  retained only if $p < 0.05$

Cross-Validation

- Correcting for model-based optimism
- Split data into a series of random subsets
  usually 10 of them
- one-at-time, leave subset out
- Build model on 90%
  use other 10% for validation

Stata Code

- gen u = uniform()
  makes random # and puts in "u"
- xtile cat=u, nq(10)
  creates "cat" 10 groups based on "u"
- logistic dead *predictors* if cat!=k
  xtile cat1=u if dead==1, nq(10)
- predict *prediction* if cat==k
  save prediction for kth group

How it works

- Uses data to build and validate model
- 90% builds model, 10% validates it
- Every observation gets used in 1 test set
- Provides an internal validation

Model Fit and Complexity

- Log-likelihood (LL) measures closeness of data to model
- -2*Log-likelihood: like a sum of squares
  smaller means more variation explained
- Always goes down with more predictors
- LR test: must go down by 3.84 for 1 predictor to be significant

Aikake Info. Criterion

- Turns out -2*log-likelihood is biased
- Assessed on same data use to minimize
- Unbiased version:
- -2*log-likelihood - 2*K
- K is # parameters in model
- Call the Aikake Info. Criterion (AIC)
- p-value criterion $p < 0.16$

Bayesian Info. Criteria

- A more stringent version
- -2*LL - K*log(n)
- n is number of predictors
- bigger penalty for adding predictor *increases with sample size*

BIC Table

N	Chi2 >	p-value
20	3	0.083
100	4.6	0.032
200	5.3	0.021
500	6.2	0.013
2000	7.6	0.006

Information Criteria

- Choose model with lowest AIC/BIC
- Very big difference
- AIC will allow some non-sig predictors
- BIC will exclude some non-sig predictors
  BIC can lead to underfitting

Building Prediction Models

- Balance utility, prior knowledge
- Predictors powerful, portable, interpretable, accessible
- Mitigate overfitting: BIC, cross-validation, Construct simple score
- Context dependent cutpoints

Prediction w/ Logistic Regression

Predictor $x_1, \ldots\ldots, x_p$

Linear Pred. $x_1\beta_1 + \ldots\ldots + x_p\beta_p \longrightarrow$ Score

Predicted Probability

$$\frac{\exp(\beta_0 + x_1\beta_1 + \ldots\ldots + x_p\beta_p)}{1+\exp(\beta_0 + x_1\beta_1 + \ldots\ldots + x_p\beta_p)}$$

17 df predictor model

coefficients

```
. xi: logistic dead i.agecat male hrsdm hrsca hrslung hrschf bmi smoke bath money walkjog push, coef
i.agecat          _Iagecat_0-6        (naturally coded; _Iagecat_0 omitted)

Logistic regression                               Number of obs   =      11652
                                                  LR chi2(17)     =    2080.12
                                                  Prob > chi2     =     0.0000
Log likelihood = -3132.1774                       Pseudo R2       =     0.2493

------------------------------------------------------------------------------
        dead |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  _Iagecat_1 |   .6345948   .1454731     4.36   0.000     .3494727    .9197169
  _Iagecat_2 |   1.039148   .1413782     7.35   0.000     .7620513    1.316244
  _Iagecat_3 |   1.314136   .1370846     9.59   0.000     1.045455    1.582817
  _Iagecat_4 |   1.689447   .1371779    12.32   0.000     1.420583    1.958311
  _Iagecat_5 |    2.11972   .1429917    14.82   0.000     1.839461    2.399979
  _Iagecat_6 |   2.789075   .1451113    19.22   0.000     2.504662    3.073487
        male |    .712165   .0704824    10.10   0.000     .5740221    .8503079
       hrsdm |   .5781861   .0848568     6.81   0.000     .4118699    .7445023
       hrsca |   .7208027   .0851805     8.46   0.000      .553852    .8877534
     hrslung |   .8225759   .1244037     6.61   0.000     .5787491    1.066403
      hrschf |    .851144    .145293     5.86   0.000     .5663751    1.135913
      bmicat |   .5017212   .0698006     7.19   0.000     .3649147    .6385278
       smoke |   .7204122   .0927079     7.77   0.000     .5387081    .9021163
        bath |   .6721684   .1055242     6.37   0.000     .4653448    .8789919
       money |    .643781   .0961402     6.70   0.000     .4553498    .8322122
     walkjog |   .7358055   .0786537     9.36   0.000     .5816471    .8899639
        push |   .4231833   .0791656     5.35   0.000     .2680215     .578345
       _cons |  -4.903092   .1321857   -37.09   0.000    -5.162171   -4.644013
------------------------------------------------------------------------------
```

Creating a Score

- Mimics the linear predictor (lp)
  if model is right, lp summarizes risk
- Simple score requires categorical predictor
- Recode them so 1=high risk, 0 = low
- Score is sum of points for each predictor
- Score for jth predictor: round(β_j/min(β))

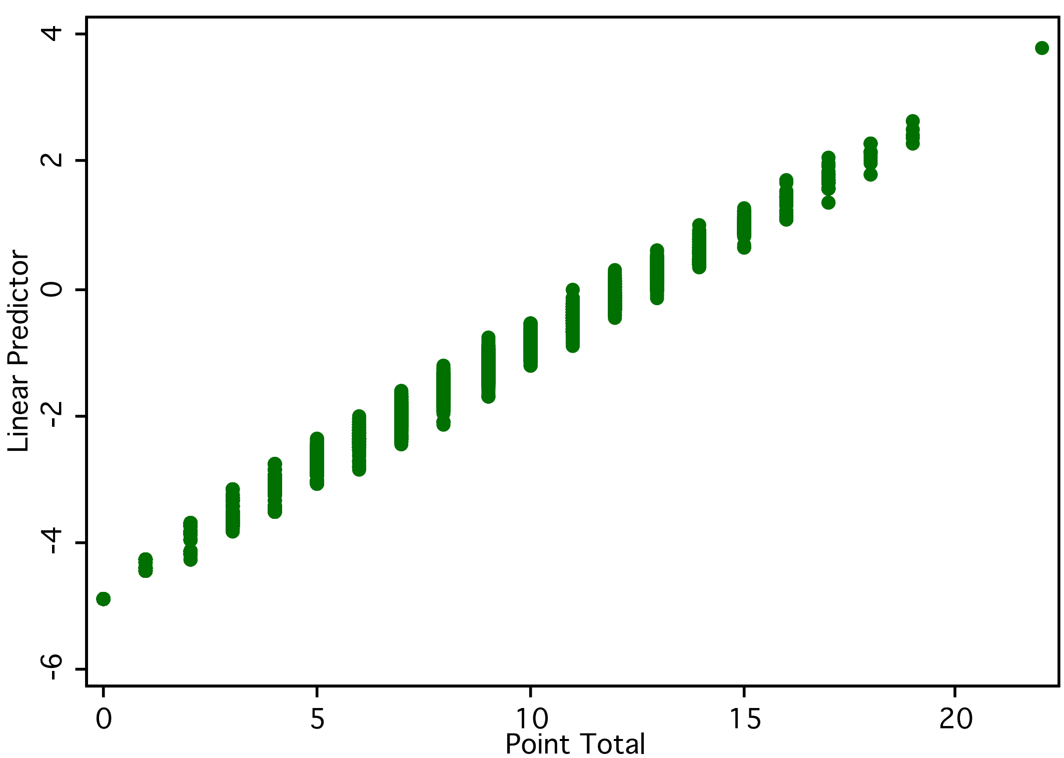
Calculating Points

Push: HR = 1.53
coef = 0.42 (smallest coef)
points = +1 0.42/0.42

Walk/Jog HR = 2.1
coef = 0.736
points = +2 $0.736/0.42 \approx 2$

Variable	Lee Points
Age 60-64	+1
Age 65-69	+2
Age 70-74	+3
Age 75-79	+4
Age 80-84	+5
Age > 85	+7
Male	+2
Diabetes	+1
Cancer	+2
Lung Disease	+2
Heart Failure	+2
BMI < 25	+1
Smoking	+2
Bathing	+2
Money	+2
Walking	+2
Pushing	+1

Points v. Linear Predictor



Probability of Death