Propensity Scores

Cover:

Propensity score and its properties

Methods for using propensity score to control confounding:

Methods for using propensity score:

in analysis

- Simple control: subclassification/modeling
- Weighting
- Tests of ignorable treatment assignment

in design

- Matching

Propensity score:

- conditional probability of receiving treatment/exposure given covariates X
- $e(X) = pr(A = 1 | X)$

for a binary treatment (as here), this is a scalar quantity properties:

balancing score $b(X)$:

defined as function of covariates for which $X \perp\!\!\!\perp A | b(X)$

- balances covariates between treated and untreated
- treatment and covariates are conditionally independent given balancing score

propensity score is a balancing score

$$f\{X|e(X),A\} = f\{X|e(X)\}$$

$$X \coprod A|e(X)$$

that is, the distribution of covariates at different treatment/exposure levels is the same once one conditions on the propensity score

proof:

$$pr\{A=1|X,e(X)\} = pr\{A=1|X\} = e(X)$$

$$pr\{A=1|e(X)\} = E\{E(A|X)|e(X)\} = E\{e(X)|e(X)\} = e(X)$$

propensity score is coarsest balancing score

Theorem 2 of Rosenbaum and Rubin (1983)

Let $b(X)$ be a function of X. $b(X)$ is a balancing score iff $b(X)$ is finer than $e(X)$; i.e., $e(X)=f\{b(X)\}$ for some function f

proof: see Rosenbaum and Rubin

provide examples of balancing score $b(X)$

X (trivially)

$\{e(X), q(X)\}$ for any arbitrary function q (i.e., propensity score + any covariate)

$g\{e(X)\}$: any 1-1 function of the propensity score

let $expit(y) \equiv \exp(y)/\{1+\exp(y)\}$ (inverse of logit function)

e.g., if $e(X) = expit(X\beta)$, then $X\beta$ is a balancing score

If $b(X)$ is balancing score and treatment assignment strongly ignorable given X, then treatment assignment strongly ignorable given $b(X)$

If X sufficient to control for confounding, then $b(X)$ also sufficient

$$pr\{A=1|b(X),\underline{Y}\}=pr\{A=1|b(X)\}$$

proof (see Rosenbaum and Rubin for more complete version)

$$pr\{A=1|b(X),\underline{Y}\}=E\{pr(A=1|X,\underline{Y})|b(X),\underline{Y}\}$$

$$=E\{pr(A=1|X)|b(X),\underline{Y}\} \qquad \text{(ignorability)}$$

$$=E\{e(X)|b(X),\underline{Y}\}=e(X)=pr\{A=1|b(X)\}$$

How could one use propensity score (or other balancing score) to control for confounding, estimate effect of treatment in estimation?

Methods for using propensity score in analysis:

- Control for propensity score as regular variable
- Use in weighting
- Use in estimating equations in tests of strongly ignorable treatment assignment

consequence of properties of propensity score, strong ignorability:

$$E\{Y^a | b(X)\} = E\{Y^a | b(X), A = a\} = E\{Y | b(X), A = a\}$$

at given level of $b(X)$, subjects who are treated are representative of what would have happened to all subjects had they been treated

true for propensity score $e(X)$ as balancing score

why can one not in general use propensity score $e(X)$ directly in data analysis?

Propensity score $e(X)$ unknown

How can this be remedied?

Estimate propensity score

estimated score $\hat{e}(X)$

e.g., use logistic regression: $logit\{pr(A=1|X)\}=X\beta$

estimate β as $\hat{\beta}$ (e.g., by ML)

estimate $e(X)$ by $\hat{e}(X)=expit(X\hat{\beta})$

Use estimated propensity score

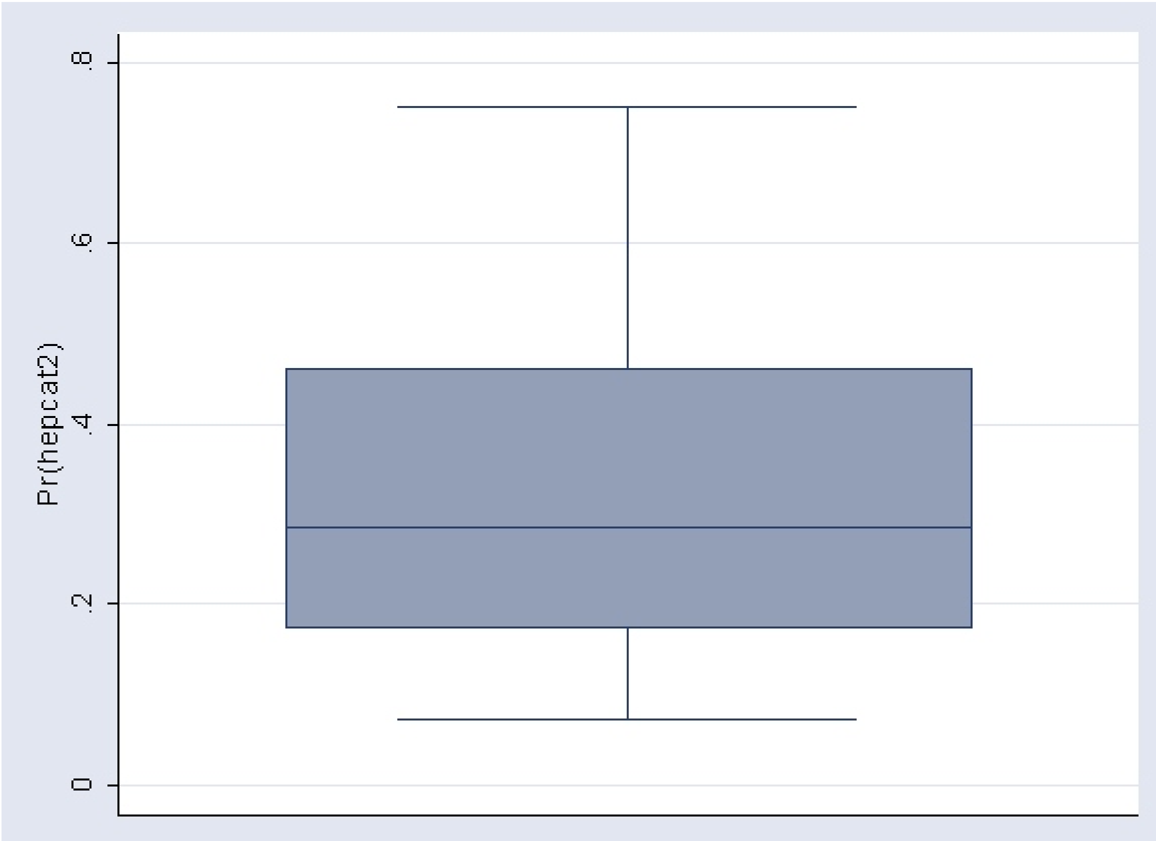
Illustrate using AVF data, simulated data

Discuss additional features/properties

AVF data:

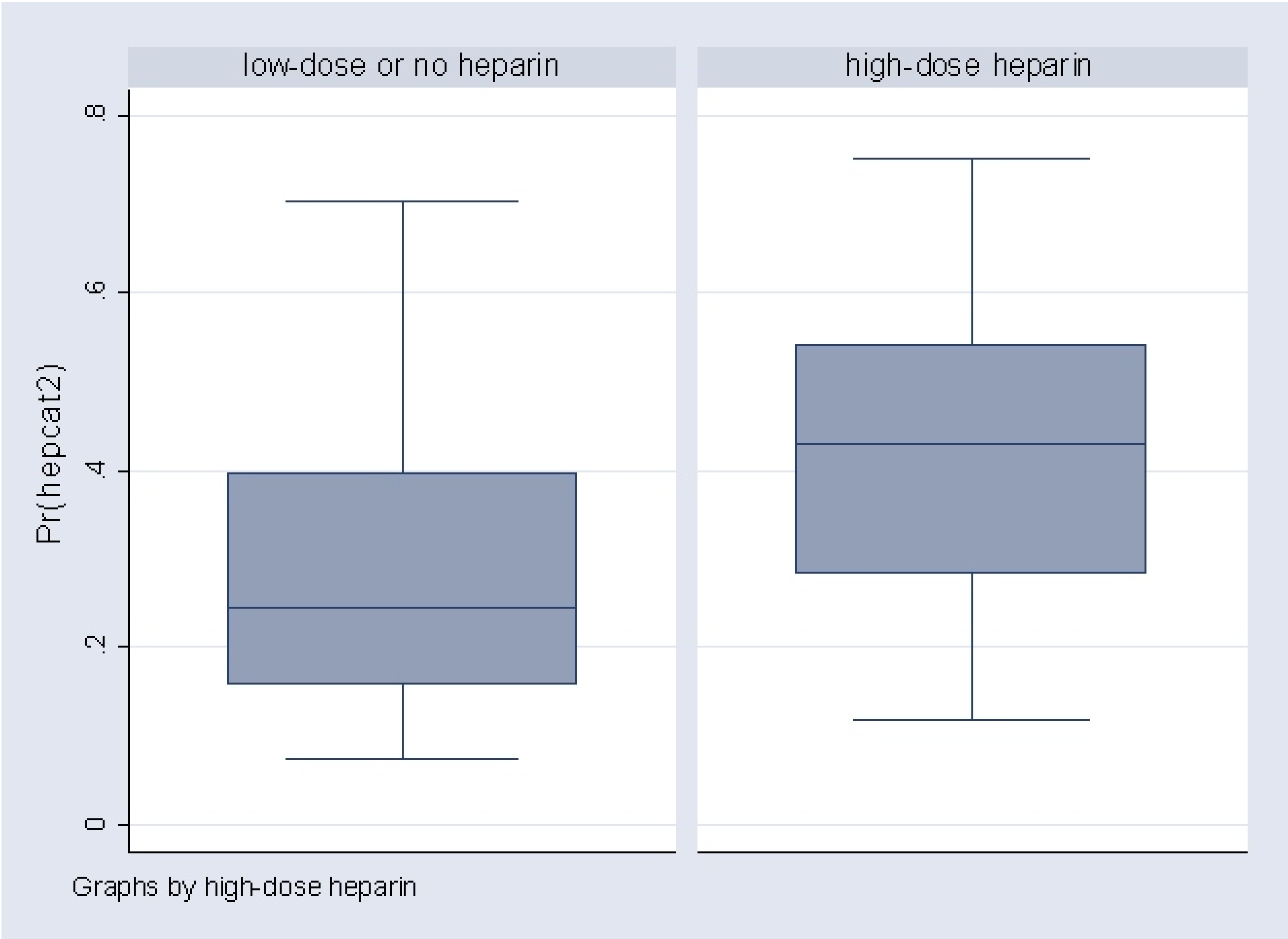
considered previously model for high-dose heparin use

Distribution of propensity score:



large spread of scores, clustering in middle of range

By heparin use category:



why should these distributions differ?

If model used to develop propensity score good, discriminates between heparin users and nonusers

distribution of exposure by propensity score:

. tab2 p_hep_pct hepcat2
-> tabulation of p_hep_pct by hepcat2

p_hep_pct	high-dose heparin low-dose	high-dose	Total
0	49	11	60
1	56	8	64
2	33	14	47
3	41	26	67
4	27	38	65
Total	206	97	303

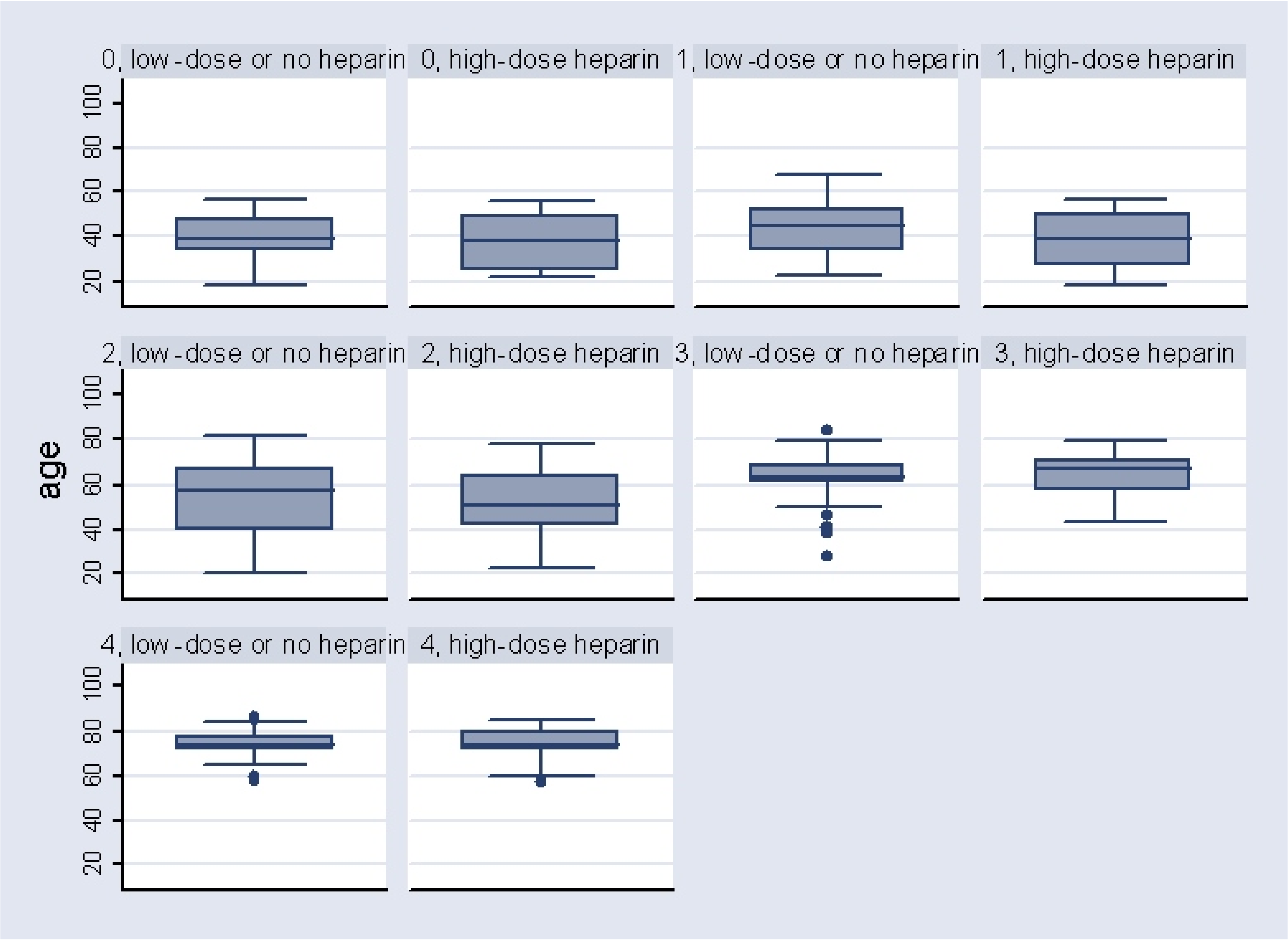
evidence that propensity score model may be misspecified?

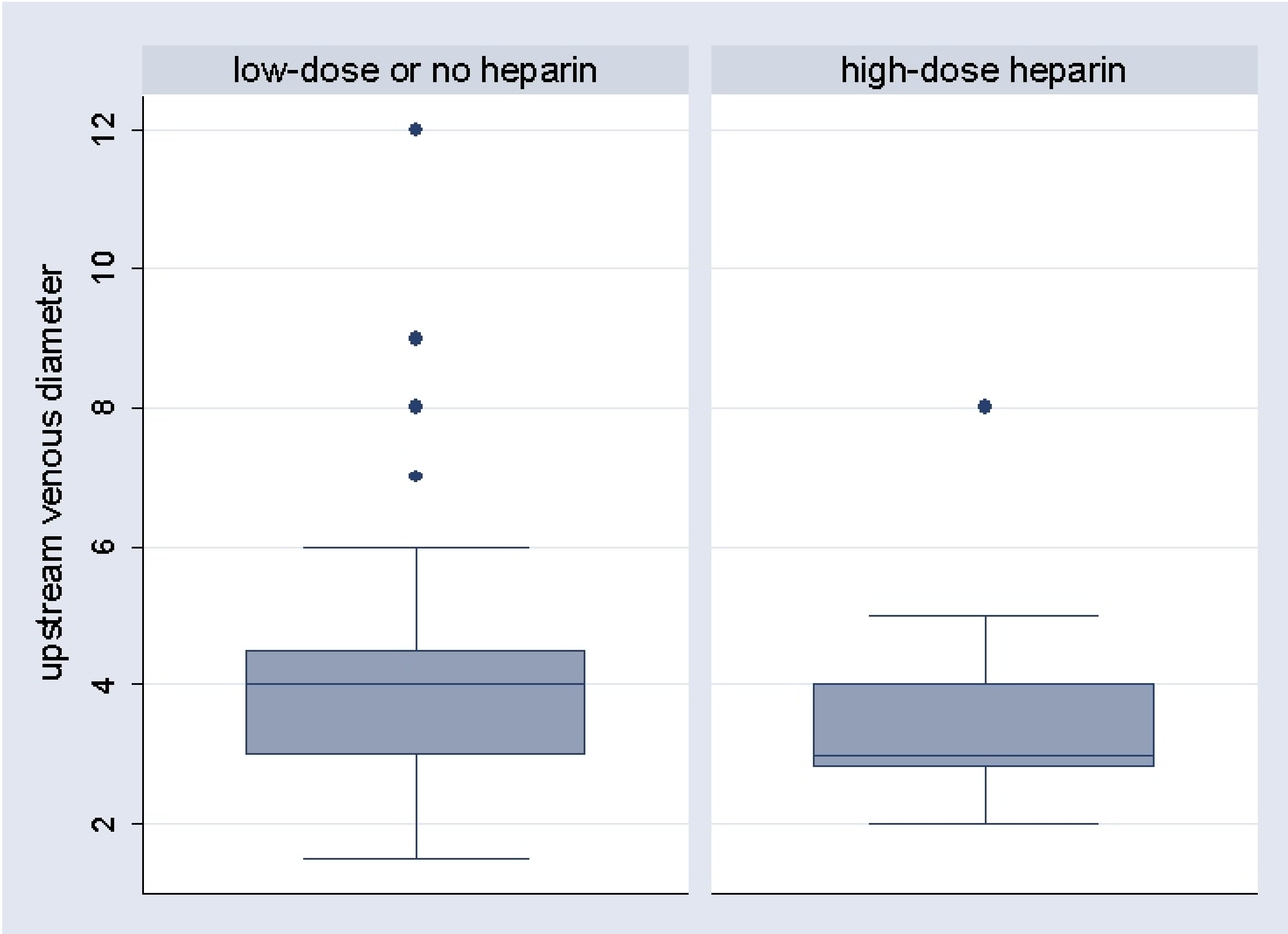
Propensity score as balancing score: show for 2 continuous variables

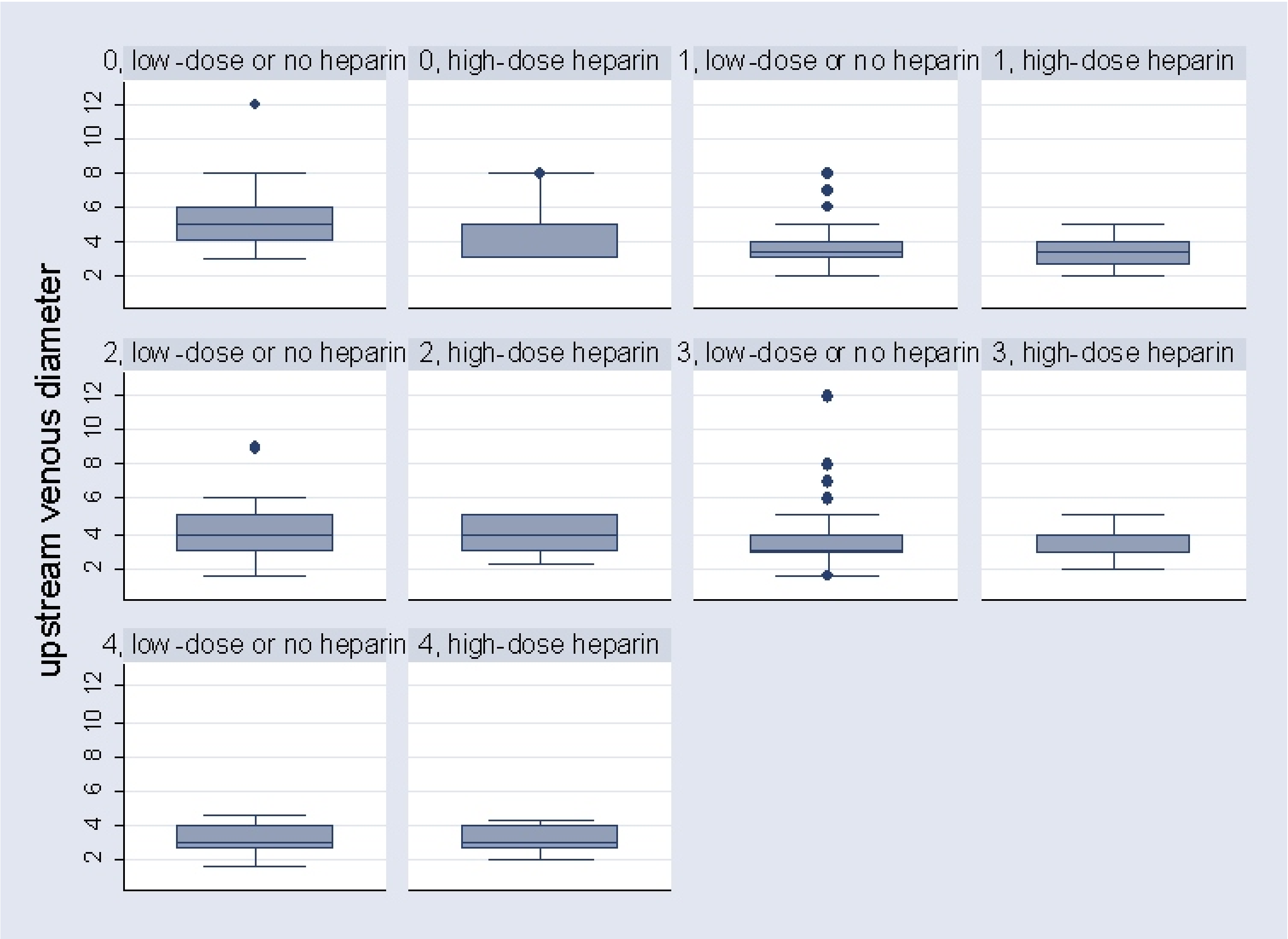
These variables were categorized for use in building the model

Should simultaneously balance each component of X

Crude comparison of age across heparin groups:


0, low-dose or no heparin
0, high-dose heparin
1, low-dose or no heparin
1, high-dose heparin
2, low-dose or no heparin
2, high-dose heparin
3, low-dose or no heparin
3, high-dose heparin
4, low-dose or no heparin
4, high-dose heparin
age
20
40
60
80
100


low-dose or no heparin
high-dose heparin
upstream venous diameter
2
4
6
8
10
12



0, low-dose or no heparin
0, high-dose heparin
1, low-dose or no heparin
1, high-dose heparin
2, low-dose or no heparin
2, high-dose heparin
3, low-dose or no heparin
3, high-dose heparin
4, low-dose or no heparin
4, high-dose heparin
upstream venous diameter
2 4 6 8 10 12

propensity score does good job of balancing covariates

how can one use propensity score to control for confounding?

Stratified analysis; estimation of common measures of effect

Regression on propensity score

standardization on propensity score

Model-based standardization

stratified analysis

```
. cs fa3 hepcat2, by(p_hep_pct)

      p_hep_pct |       RR       [95% Conf. Interval]    M-H Weight
-----------------+-------------------------------------------------
               0 |   1.145455      .8234145   1.593445      6.416667
               1 |   1.441176      1.029763   2.016959          4.25
               2 |   1.728571      1.087769   2.746869      4.468085
               3 |   1.340385      .8804873   2.040496      7.761194
               4 |   1.865132      .9757556   3.565151      4.676923
-----------------+-------------------------------------------------
           Crude |   1.210286       1.00311    1.46025
    M-H combined |   1.462469      1.191031   1.795768
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(4) =    3.319  Pr>chi2 = 0.5060
```

```
. cc fa3 hepcat2, by(p_hep_pct)

       p_hep_pct |       OR       [95% Conf. Interval]    M-H Weight
-----------------+-------------------------------------------------
               0 |        1.8      .3094041   19.01637      1.166667 (exact)
               1 |   4.529412      .5125458   213.2199        .53125 (exact)
               2 |        4.4      .8984246   28.21263      .9574468 (exact)
               3 |   1.983333      .6461413   6.261112      2.686567 (exact)
               4 |   2.933824      .9254734   9.661076      2.092308 (exact)
-----------------+-------------------------------------------------
           Crude |   1.565769      .9656131   2.550868               (exact)
    M-H combined |   2.715253      1.521295   4.846264
-------------------------------------------------------------------
Test of homogeneity (M-H)      chi2(4) =     1.27  Pr>chi2 = 0.8666

                   Test that combined OR = 1:
                                Mantel-Haenszel chi2(1) =     11.73
                                                Pr>chi2 =    0.0006
```

why do risk and odds ratios differ so much?

Outcome is not rare

crude analyses:

```
. cs fa3 hepcat2

                 | high-dose heparin      |
                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+----------
           Cases |        71         122  |        193
        Noncases |        42         113  |        155
-----------------+------------------------+----------
           Total |       113         235  |        348
                 |                        |
            Risk |  .6283186    .5191489  |   .5545977
                 |                        |
                 |      Point estimate    |  [95% Conf. Interval]
                 |------------------------+----------------------
 Risk difference |         .1091696       | -.0004648    .2188041
      Risk ratio |         1.210286       |   1.00311     1.46025
 Attr. frac. ex. |         .1737489       |  .0031003    .3151859
 Attr. frac. pop |          .063918       |
                 +-----------------------------------------------
                               chi2(1) =     3.68  Pr>chi2 = 0.0550
```

```
. cc fa3 hepcat2
                                                         Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+------------------------
           Cases |        71         122  |        193       0.3679
        Controls |        42         113  |        155       0.2710
-----------------+------------------------+------------------------
           Total |       113         235  |        348       0.3247
                 |                        |
                 |      Point estimate    |  [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         1.565769       |  .9656131    2.550868  (exact)
 Attr. frac. ex. |         .3613362       | -.0356115    .6079766  (exact)
 Attr. frac. pop |         .1329268       |
                 +-------------------------------------------------
                               chi2(1) =     3.68  Pr>chi2 = 0.0550
. cs fa3 hepcat2, by(p_hep_pct) s(stdd)

       p_hep_pct |       RR       [95% Conf. Interval]        Weight
-----------------+---------------------------------------------------
               0 |   1.145455     .8234145   1.593445           .198
               1 |   1.441176     1.029763   2.016959          .2112
               2 |   1.728571     1.087769   2.746869          .1551
               3 |   1.340385     .8804873   2.040496          .2211
               4 |   1.865132     .9757556   3.565151          .2145
-----------------+---------------------------------------------------
           Crude |   1.210286      1.00311    1.46025
    Standardized |   1.430448     1.193222   1.714838
```

standardized analysis (weights are population proportions)

```
. cc fa3 hepcat2, by(p_hep_pct) s(stdd)

       p_hep_pct |       OR       [95% Conf. Interval]        Weight
-----------------+--------------------------------------------------
               0 |        1.8      .3094041   19.01637          .198 (exact)
               1 |   4.529412      .5125458   213.2199         .2112 (exact)
               2 |        4.4      .8984246   28.21263         .1551 (exact)
               3 |   1.983333      .6461413   6.261112         .2211 (exact)
               4 |   2.933824      .9254734   9.661076         .2145 (exact)
-----------------+--------------------------------------------------
           Crude |   1.565769      .9656131   2.550868               (exact)
    Standardized |   2.893005      1.055495   7.929435
```

What is difference in estimand between standardization and Mantel-Haenszel (in principle)?

regression adjustments: categorical covariate

```
. xi: logistic fa3 hepcat2 i.p_hep_pct
i.p_hep_pct       _Ip_hep_pct_0-4     (naturally coded; _Ip_hep_pct_0 omitted)

Logistic regression                               Number of obs   =        303
                                                  LR chi2(5)      =      24.54
                                                  Prob > chi2     =     0.0002
Log likelihood = -193.43904                       Pseudo R2       =     0.0597

------------------------------------------------------------------------------
         fa3 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     hepcat2 |   2.727384    .805091     3.40   0.001     1.529261    4.864192
_Ip_hep_pc~1 |   .6756917   .2670562    -0.99   0.321     .3114045    1.466129
_Ip_hep_pc~2 |    .393963   .1667369    -2.20   0.028     .1718706    .9030449
_Ip_hep_pc~3 |   .3581631   .1413834    -2.60   0.009     .1652238    .7764064
_Ip_hep_pc~4 |   .1857047   .0778445    -4.02   0.000     .0816605     .422312
------------------------------------------------------------------------------
```

conditional logistic regression

```
. clogit fa3 hepcat2, group(p_hep_pct) or

note: multiple positive outcomes within groups encountered.
Iteration 0:   log likelihood = -186.93953
Iteration 1:   log likelihood = -182.34641
Iteration 2:   log likelihood = -182.32765
Iteration 3:   log likelihood = -182.32765

Conditional (fixed-effects) logistic regression   Number of obs   =        303
                                                  LR chi2(1)      =      12.23
                                                  Prob > chi2     =     0.0005
Log likelihood = -182.32765                       Pseudo R2       =     0.0324

------------------------------------------------------------------------------
         fa3 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     hepcat2 |   2.683382   .7856073     3.37   0.001     1.511744    4.763068
------------------------------------------------------------------------------
```

what is difference between conditional logisitic, logistic with categorical covariate?

- in model
- in estimation

regression correction for propensity score as continuous covariate

```
. logistic fa3 hepcat2 p_hep

Logistic regression                               Number of obs   =        303
                                                  LR chi2(2)      =      18.14
                                                  Prob > chi2     =     0.0001
Log likelihood = -196.63996                       Pseudo R2       =     0.0441

------------------------------------------------------------------------------
         fa3 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     hepcat2 |    2.52995   .7330599     3.20   0.001     1.433751    4.464264
       p_hep |   .0523271   .0427877    -3.61   0.000     .0105367    .2598668
------------------------------------------------------------------------------
```

what might be wrong with approach?

Model misspecification (for association of propensity score with outcome; or assumption that effect is same in all strata)

try changing form

```
. logistic fa3 hepcat2 p_hep ph2

Logistic regression                               Number of obs   =        303
                                                  LR chi2(3)      =      23.76
                                                  Prob > chi2     =     0.0000
Log likelihood = -193.83217                       Pseudo R2       =     0.0577

------------------------------------------------------------------------------
         fa3 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     hepcat2 |    2.70448   .8023675     3.35   0.001     1.511989    4.837478
       p_hep |    14.5265   36.68177     1.06   0.289     .1029768    2049.192
         ph2 |   .8518356   .0584799    -2.34   0.019     .7445937    .9745232
------------------------------------------------------------------------------
```

now more in line with categorical analyses

what else can one do in regression analysis with propensity score as a regressor?

Add other regressors

```
. xi: logistic fa3 i.hepcat2 binadmit  p_hep ph2
i.hepcat2         _Ihepcat2_0-1       (naturally coded; _Ihepcat2_0 omitted)

Logistic regression                               Number of obs   =        303
                                                  LR chi2(4)      =      27.43
                                                  Prob > chi2     =     0.0000
Log likelihood = -191.99649                       Pseudo R2       =     0.0667

------------------------------------------------------------------------------
         fa3 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _Ihepcat2_1 |   2.732512   .8149714     3.37   0.001      1.52297    4.902672
    binadmit |   .4874255   .1841326    -1.90   0.057     .2324628    1.022028
       p_hep |   15.40078   39.05357     1.08   0.281     .1069178    2218.376
         ph2 |   .8528906   .0588585    -2.31   0.021     .7449916    .9764168
------------------------------------------------------------------------------
```

little change

can add more covariates, etc.

when is propensity score adjustment advantageous?

What are advantages and disadvantages of propensity score adjustment?

If outcome binary/failure-time and rare, may be hard to control simultaneously for several confounders X using conventional approach; can include in propensity score model

Not dependent on modeling assumptions

If do diagnostic of looking at stratum-specific exposure frequency, can tell where not enough data to yield reasonable inference

Restricting to strata with enough data for inference changes estimand

Can lead to control for irrelevant covariates (related to exposure, not outcome)

Stratification will not generally eliminate bias; sometimes important

Interpretation: models condition on secondary characteristic (propensity score)

- important in noncollapsible models (e.g., logistic)

- effect modification (however, can include modification of effect by other covariates in model in conjunction with controlling for propensity score)

Standard modeling reliant on model form assumptions

How can one reduce bias?

Consider more flexible models

e.g., add polynomial terms, splines, smooths, etc. for other covariates

What is effect of strata with few subjects at one exposure level on variance of estimators?

- Stratified estimator (e.g., Mantel-Haenszel)
- Standardized estimator

Stratified estimator: little effect; does not change

Inverse variance weights:

$$\hat{\Phi} = \frac{\sum_i w_i \hat{\Phi}_i}{\sum_i w_i}$$

$$Var(\hat{\Phi}) = \frac{\sum_i w_i^2 Var(\hat{\Phi}_i)}{(\sum_i w_i)^2} = \frac{\sum_i w_i}{(\sum_i w_i)^2} = \frac{1}{\sum_i w_i}$$

Standardized estimator: increases variance substantially

Let $\mu_i^a = E(Y^a | X)$

$$Var(\hat{\mu}^a) = \frac{\sum_i w_i^2 Var(\hat{\mu}_i^a)}{(\sum_i w_i)^2}$$

Consequences of use of estimated instead of true propensity score: increased or decreased variance and bias? Why?

Decreased variance, decreased bias (conditional on observed propensity score)

Why?

Propensity score is ancillary statistic

Likelihood for observables (under ignorable treatment assignment):

$$pr(X;\alpha)pr(A|X;\beta)pr(Y|A,X;\gamma)$$

if parameters distinct, ancillary

conditioning on observed ancillaries leads to decreased variance

conditional on observed ancillary, estimator which uses expected/true ancillary biased (Robins, 1987)

γ^* treatment effect parameter

$\hat{\gamma}^*$ estimator which uses observed ancillary $\hat{\beta}$

$\hat{\gamma}^*$ estimator which uses true β /expected ancillary

$$Var(\hat{\gamma}^*) \leq Var(\tilde{\gamma}^*)$$

$$E(\tilde{\gamma}^* - \gamma | \hat{\beta}) \neq 0$$

$$E(\hat{\gamma}^* - \gamma | \hat{\beta}) = 0$$

Other methods using propensity scores: weighting

Inverse probability weighting:

Consider first estimating $E(Y^a)$

note that $\frac{E\{YI(A=a)|X\}}{pr(A=a|X)} = \frac{E\{Y^a I(A=a)|X\}}{pr(A=a|X)} =$ (consistency)

$$= \frac{E\{Y^a|X\}E\{I(A=a)|X\}}{pr(A=a|X)} \qquad \text{(ignorability)}$$

$$= \frac{E\{Y^a|X\}pr(A=a|X)}{pr(A=a|X)} = E(Y^a|X)$$

thus, can estimate $E(Y^a|X)$ as $\dfrac{\sum_{i:X=x} YI(A=a)/pr(A=a|X=x)}{\sum_{i:X=x} I(A=a)/pr(A=a|X=x)}$

and $E(Y^a)$ as $\dfrac{\sum_{i} YI(A=a)/pr(A=a|X=x)}{\sum_{i} I(A=a)/pr(A=a|X=x)}$

interpretation of weighting:

potential outcome Y^a is observed for proportion $pr(A=a|X)$ of subjects with covariates X

each subject with treatment level a stands in for several subjects with covariate level X (a total of $1/pr(A=a|X)$: self and $1/pr(A=a|X)-1$) subjects who are not treated but are otherwise comparable

rewrite

$$E(Y^a)=\sum_x pr(X=x)pr(Y|X=x,A=a)=\sum_x \frac{pr(X=x)pr(Y,A=a|X=x)}{pr(A=a|X=x)}$$

$$=\sum_x \frac{pr(X=x)pr(Y,A=a|X=x)}{pr(A=a|X=x)}$$

where X high-dimensional, can choose to model $pr(Y|X,A)$ or $pr(A|X)$

approach similar to survey sampling (Horvitz-Thompson), methods for missing data

use weights to jointly estimate outcomes under different potential outcomes

weights: $1/pr(A|X)$

what is distribution of A in pseudopopulation? (For binary treatment)

let $pr^*(\cdot|\cdot)$ denote probability in pseudopopulation

$$pr^*(A=a|X) = \frac{pr(A=a|X)/pr(A=a|X)}{\sum_{a'} pr(A=a'|X)/pr(A=a'|X)}$$ (Normalize through denominator)

$=0.5$ (binary A)

$=1/\tilde{A}$, where $\tilde{A}$ is # of possible values of treatment

What is $pr^*(X|A)$ in pseudopopulation?

$$pr^*(X|A=a) = \frac{pr(X|A=a)/pr(A=a|X)}{\sum_x pr(X=x|A=a)/pr(A=a|X=x)}$$

Note: $\frac{pr(X|A=a)}{pr(A=a|X)} = \frac{pr(X)pr(A=a|X)}{pr(A=a)pr(A=a|X)} = \frac{pr(X)}{pr(A=a)}$

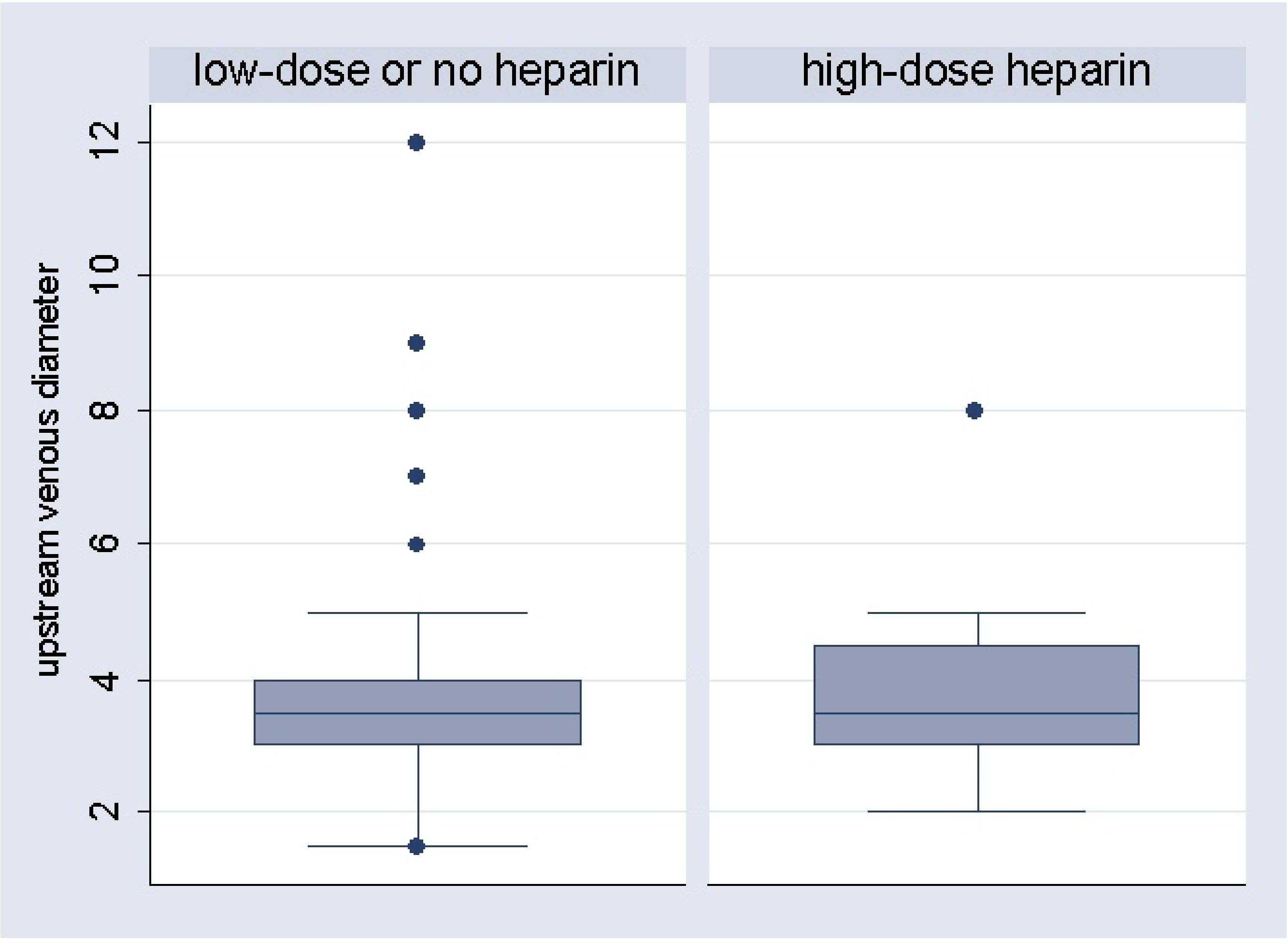
thus, $pr^*(X|A=a) = \frac{pr(X)/pr(A=a)}{\sum_x pr(X=x)/pr(A=a)} = pr(X)$

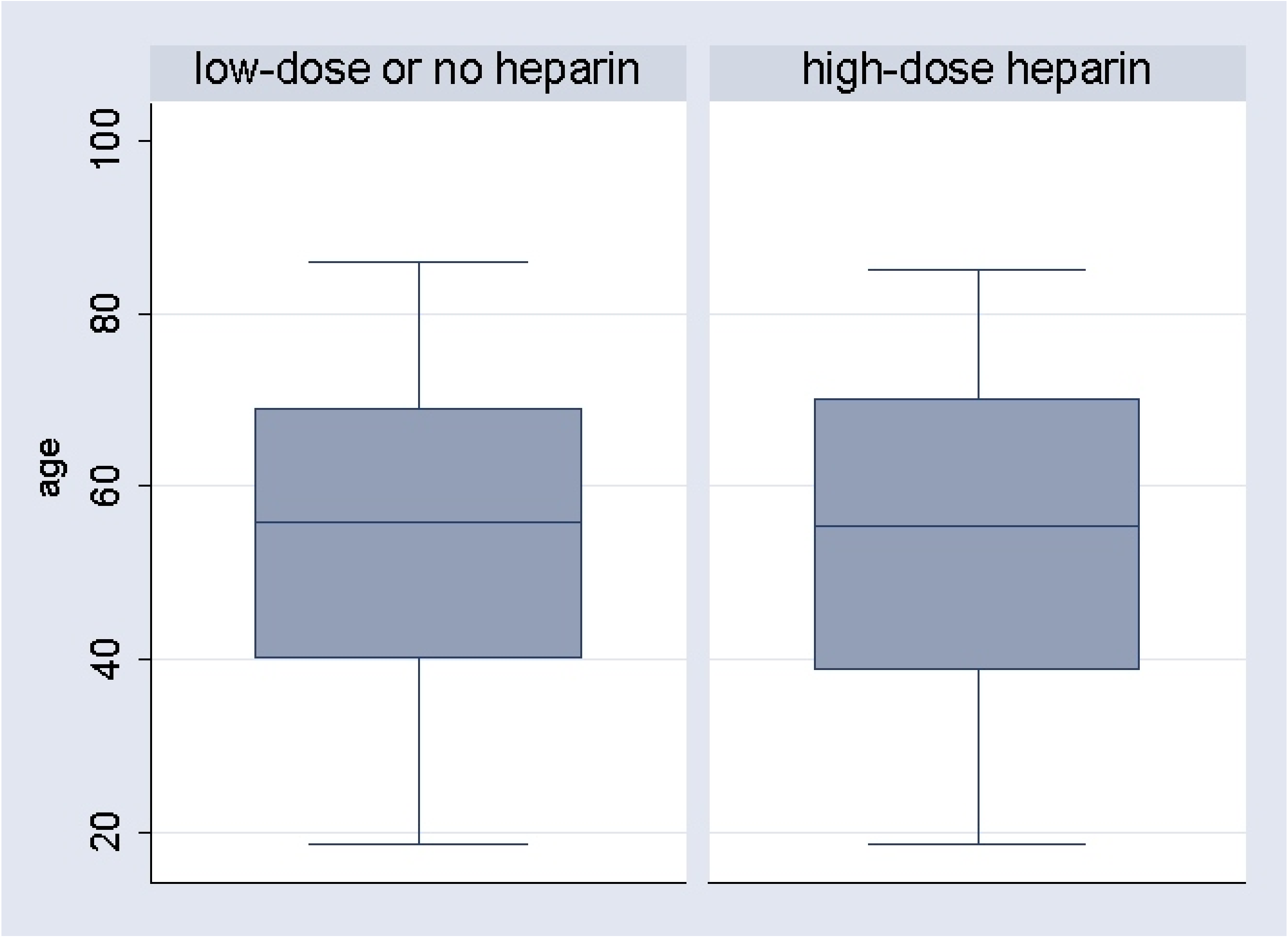
thus, in pseudopopulation, $A \perp X$

distribution of covariates balanced globally across different treatment groups

see Rosenbaum (1987)

AVF study: (weighted boxplots)




low-dose or no heparin
high-dose heparin
age
100
80
60
40
20

As with standard propensity score adjustment, balance on covariates + ignorability translates to unconditional balance on potential outcomes:

$pr^*(\underline{Y}|A)=pr^*(\underline{Y})=pr(\underline{Y})$

essentially, have created randomized trial in pseudopopulation

also have, for pseudopopulation,
$pr(Y^a|V)=pr^*(Y^a|V)=pr^*(Y^a|V,A=a)=pr^*(Y|A=a)$

can estimate parameters in MSM by estimating in weighted population

weight each subject by $1/pr(A|X)$

what happens to variance when probability of receiving some treatment in some groups defined by covariates X is small?

Variances can get large

Suppose have normal outcomes with variance σ^2: $\hat{\mu} \equiv \dfrac{\sum_i Y_i w_i}{\sum_i w_i}$

$$Var(\hat{\mu}) \equiv \sigma^2 \frac{\sum_i w_i^2}{(\sum_i w_i)^2}$$

equal weights: $Var(\hat{\mu}) \equiv \sigma^2 \frac{Nw^2}{(Nw)^2} = \sigma^2/N$

1 very large weight: $Var(\hat{\mu}) \equiv \sigma^2 \frac{w^2}{w^2} = \sigma^2$

How can effect of unequal weights be mitigated in estimating MSMs?

Consider weights of form $q(A,V)/pr(A|X)$

If these weights are used, do we get covariate balance in any subgroup of population?

Yes, have $pr^*(X|A,V)=pr^*(X|V)$, where $pr^*(\cdot|\cdot) \propto pr(\cdot|\cdot)\dfrac{q(A,V)}{pr(A|X)}$

What would be reasonable weights to use to reduce variance?

Consider case where $V=X$

Here, can use unweighted estimate, since model $E(Y^a|X)=E(Y|X,A=a)$ assuming model correctly specified

formulate in terms of $q(A,V)/pr(A|X)$

$q(A|V)=pr(A|V)$

in general, "stabilized weights" are $pr(A|V)/pr(A|X)$

consider pseudopopulation created by weighting

what are exposure probabilities in population?

$$pr^*(A|X) = \frac{pr(A|V)}{pr(A|X)} pr(A|X) = pr(A|V)$$

model: randomized trial with uneven randomization probabilities

covariates, potential outcomes balanced within strata of V

Tests of strongly ignorable treatment assignment

Structural nested models:

$Y^0 \perp A | X$

Logistic regression version:

$logit\{E(A|X, Y^0)\} = X\beta + Y^0\gamma$

In randomized trial, $\beta = 0$; in observational study, $\beta \neq 0$

$\gamma = 0$

for model like $Y^0 = Y - A\Psi$, again propose putative model $Y^0(\Psi) = Y - A\Psi$

Estimating equations:

$$\sum_i \{A - e(X)\} g(Y^0, X) = 0$$

will substitute p_X for $e(X)$

Can derive as score equations from logistic regression

or, note that, under strong ignorability, $Y^0 \perp A | X$, which implies that any function of Y^0, X will be independent of X

this, in turn, implies

$$E\{(A-p_X)g(Y^0,X)\} = E[E\{(A-p_X)g(Y^0,X)|X\}]$$

$$= E[E(A-p_X|X)E\{g(Y^0,X)|X\}]$$

$$= E[E(p_X-p_X|X)E\{g(Y^0,X)|X\}]$$

$$= 0$$

For strong model, will be true if substitute $Y^0(\Psi_0)$ for Y^0

How would one compute variance for test statistic (under null)?

Let $U(\Psi) = \sum_i \{A - p_X\} g\{Y^0(\Psi), X\} = \sum_i U_i(\Psi)$

Since $E\{U_i(\Psi)\} = 0$, $Var\{U_i(\Psi)\} = E\{U_i(\Psi)^2\}$

Can compute $Var\{U(\Psi)\}$ as $\sum_i U_i(\Psi)^2$

Alternatively, compute expected value from logistic regression

$\partial U(\Psi)/\partial\beta = -\sum_i p_x(1 - p_x) g\{Y^0(\Psi), X\}^2$

$I(\Psi) = -\partial U(\Psi)/\partial\beta$, take inverse of information matrix to get variance

How would one modify if estimating logistic regression?

Let $U(\beta) = \sum_i \{A - p_X\}X = \sum_i U_i(\beta)$

Let $U(\beta, \Psi) = \{U(\beta), U(\Psi)\}$

run through same argument (now multivariable/matrix argument)

Variance of $\hat{\Psi}$; how to derive

$$\square = U(\hat{\Psi}) \approx U(\Psi_0) + \frac{\partial U(\Psi)}{\partial \Psi}(\hat{\Psi} - \Psi_0)$$

$$U(\Psi_0) = -\frac{\partial U(\Psi)}{\partial \Psi}(\hat{\Psi} - \Psi_0)$$

$$Var\{U(\Psi_0)\} = E\{\frac{\partial U(\Psi)}{\partial \Psi}\} Var(\hat{\Psi}) E\{\frac{\partial U(\Psi)}{\partial \Psi}\}' = B(\Psi) Var(\hat{\Psi}) B(\Psi)'$$

$$Var(\hat{\Psi}) = B^{-1}(\Psi) Var\{U(\Psi_0)\} B^{-1}(\Psi)'$$

substitute to estimate variance

extensions to settings where estimate β straightforward

Variance depends on choice of function $g(\cdot)$

Essentially infinite # of estimators

How can one find estimator with good properties/smallest variance?

What is smallest variance possible?

Theory of semiparametric inference (Newey, 1990; Bickel et al., 1992)

will present some ideas briefly; complicated, mathematical; involves Hilbert spaces, etc.

Semiparametric information bound/analogous to Cramer-Rao bound:

minimum information of all parametric submodels

$Y = g(X) + A\beta + \epsilon$, ϵ i.i.d.

interested in inference for β that will be valid no matter what function $g(X)$, ϵ

true model: $g(X) = 2X$, $\epsilon \sim N(0,1)$, X scalar

what would be a submodel?

submodels:

submodel 1: $g(X) = 2X$, $\epsilon \sim N(0,1)$

submodel 2: $g(X) = X\alpha$, α unknown; $\epsilon \sim N(0,1)$

submodel 3: $g(X) = X\alpha$, α unknown; $\epsilon \sim N(0,\sigma^2)$, σ^2 unknown

submodel 4: $g(X) = \alpha_0 + X\alpha_1 + X\alpha_2^2$, α unknown; $\epsilon \sim N(0,\sigma^2)$, σ^2 unknown

for inference to be valid no matter what submodel is true, information for β in semiparametric model (making no assumptions about $g(X)$ or ϵ cannot be greater than information available in any true parameteric submodel

information bound is minimum information from all parameteric submodels some choice for $g(Y^0,X)$ will lead to attaining semiparametric information bound

other choices will lead to consistent but inefficient estimators of Ψ_0

optimal choice:

let $l(\Psi,\theta;X,A,Y)$ denote the log-likelihood for Ψ, θ based on the observable data

θ denotes all other parameters in the joint distribution of the observables

let $S_\Psi(\Psi,\theta;X,A,Y) \equiv \partial l(\Psi,\theta;X,A,Y)/\partial\Psi$

optimal function:

$$g^{opt}(Y^0,X) = E\{S_\Psi(\Psi,\theta;X,A,Y)|X,A=1,Y^0\} - E\{S_\Psi(\Psi,\theta;X,A,Y)|X,A=0,Y^0\}$$

are expectations meaningful here?

derivation based on theory of semiparametric inference

likelihood (based on deterministic/rank-preserving model):

$$L(\Psi,\theta;X,A,Y) = \frac{\partial Y^0(\Psi)}{\partial Y} L\{\Psi,\theta;X,A,Y^0(\Psi)\}$$ (change-in variables)

In simple model, $Y^0(\Psi) = Y - A\Psi$, so $\frac{\partial Y^0(\Psi)}{\partial Y} = 1$

$$L\{\Psi,\theta;X,A,Y^0(\Psi)\} = pr(X;\theta)pr(A|X;\theta)pr\{Y^0(\Psi)|X,A;\theta\}$$

to take partial with respect to Ψ, need only consider last term

$$\frac{\partial \ln[pr\{Y^0(\Psi)|X;\theta\}]}{\partial\Psi} = \frac{\partial \ln[pr\{Y^0(\Psi)|X;\theta\}]}{\partial Y^0(\Psi)}\frac{\partial Y^0(\Psi)}{\partial\Psi}$$

for simple model, $\frac{\partial Y^0(\Psi)}{\partial\Psi} = -A$

$\frac{\partial \ln[pr\{Y^0(\Psi)|X;\theta\}]}{\partial Y^0(\Psi)}$ depends on distribution of $f(Y^0|X)$

suppose that $Y^0 \sim N\{\mu(X), \sigma^2(X)\}$

$$f(Y^0|X) = \frac{1}{\sigma(2\pi)^{0.5}} \exp[-\frac{\{Y^0(\Psi)-\mu(X)\}^2}{2\sigma^2}]$$

$$\ln f(Y^0|X) = r(\sigma) - \frac{\{Y^0(\Psi)-\mu(X)\}^2}{2\sigma^2}$$

$$\frac{\partial \ln f(Y^0|X)}{\partial \Psi} = -\frac{Y^0(\Psi)-\mu(X)}{\sigma^2}\frac{\partial Y^0(\Psi)}{\partial \Psi} = -\frac{Y^0(\Psi)-\mu(X)}{\sigma^2}A$$

$$E\{\frac{\partial \ln f(Y^0|X)}{\partial \Psi}|X, A=1, Y^0\} - E\{\frac{\partial \ln f(Y^0|X)}{\partial \Psi}|X, A=0, Y^0\} = -\frac{Y^0(\Psi)-\mu(X)}{\sigma^2}$$

interpret:

optimal function: residual from regression of Y^0 on X

proportional to inverse of residual variance; suppose that σ^2 varies with X

departures of Y^0 from predicted value more consequential if variance small optimal function q^{opt} often depends on unknown parameters of joint distribution

can estimate those parameters (θ) from “data” $Y^0(\Psi), X$

e.g., in previous example, assume $\mu(X) = X\beta$

estimate β from least squares regression of $Y^0(\Psi)$ on X

asymptotically, no efficiency loss from estimating those parameters

called adaptive estimation

if model specification leading to estimates of θ wrong, will get inefficient but consistent estimation

Not much used for this setting

nonetheless, valuable; why?

semiparametric model (for continuous outcome);

need not specify association of X with Y^a to obtain consistent estimates

extends nicely to more complex settings (multiple/time-varying treatments, noncompliance in randomized trials, instrumental variables estimation)

Other inferential method based on tests of ignorable treatment assignment:

Permutation-test based

How could this be done?

Advantages/disadvantages

Performing permutation tests:

stratify on propensity score

test:
- derive overall test statistic
- test based on permutation distribution of test statistic; more complicated
- estimation based on inverting test

advantage:
- not dependent on asymptotic assumptions

disadvantage:
- requires categorization of confounders/propensity score

Can consider models for causal effect with multiple parameters:

e.g., $Y^0 = Y + A\Psi_a + AX\Psi_{ax}$

interpret

use vector function $q(x, y^0)$ of same dimension as parameter vector $\boldsymbol{\Psi}(\equiv\{\Psi_a, \Psi_{ax}\})$

optimal function obtained by taking vector of partial derivatives with respect to $\boldsymbol{\Psi}$

Other settings:

nonbinary exposures:

categorical (possibly ordered)

continuous

how would methods sketched above relate to these types of exposures?

in general, no scalar balancing score

exceptions:

some models for ordinal outcomes (treatments):

e.g., ▯▯▯▯▯ $pr(A \geq a|X)\} = a\alpha_a + X\beta$ (proportional odds model; β same for all levels a)

other models work as well

some models for continuous outcomes, common error distribution

e.g., $A = X\beta + \epsilon, \ \epsilon \sim N(0,\sigma^2)$

in both cases, $X\beta$ is balancing score otherwise, have vector generalized propensity score

for categorical treatments, have vector of independent probabilities (number of categories -1)

for continuous treatments, more complicated, depends on model

if have continuous outcome normally distributed with both mean and variance dependent on X, stratum-specific mean μ_X and standard deviation σ_X^2 are balancing score/generalized propensity score $e^*(X)$

$A \perp X | e^*(X)$

generalization of methods to non-binary treatments

standard control (regression, stratified analysis):

now in general have vector summary to control for;

- stratification more difficult

- may be more dependent on assumptions in modeling association of $e^*(X)$ and outcome weighted estimation

for categorical treatments, no need for dimension reduction

for continuous treatments, may use density in inverse weights

tests of ignorable treatment assignment

may proceed similarly in principle

Propensity scores in case-control studies; not worked out fully

How can one estimate propensity score in case-control studies with known sampling fraction?

Use weighted estimation

provides consistent estimate of population propensity score

unfortunately, get funny small sample behavior

distortion of degree of effect modification by propensity score / artifacual effect modification even if there truly is none

can use a sort of resampling to estimate degree of bias, but (limited simulations) underestimates this survival studies: need to deal with left truncation, right censoring appropriately

left truncation can lead to problems estimating propensity score (see next section)

censoring: for standard methods to be valid, require censoring to be independent of failure conditional on variables included in model

thus, if adjusting for propensity score, require that censoring independent of failure among subjects with particular value of $e(X)$, exposure level A

independence not implied by independence of censoring conditional on X,A

alternate type of confounder score: risk score

formulate risk score in fashion analogous to propensity score

$\mu(X) = E(Y|X)$

or $\mu^a(X) = E(Y^a|X)$

problems with each

$\mu(X)$ "confounded"

$$\mu(X) = E(Y|X) = E\{E(Y^A|X,A)|A\} = \sum_a pr(A=a|X)E(Y^a|X)$$

$\mu^a(X)$ not unique; for binary A, vector with 2 elements

if ranks of $\mu^1(X),\mu^0(X)$ same for each subject, then can reduce dimension to 1

balancing score of sorts: $f(X|Y^a,\mu^a)=f(X|\mu^a)$

requires some additional conditions, like constant error structure distribution for Y^a given X

not true for binary outcome

not verifiable

use of estimated scores: is this problematic for naive variance calculations?

Yes. Essentially same as regressing Y on X and A and treating coefficients for X as known