A review of modern methods of estimating the size of health disparities

May 24, 2017

Emil Coman[1]
Helen Wu[2]

[1] UConn Health Disparities Institute, [2] UConn Health

Health Disparities (HD): It's just about comparing two groups

Goals

1. Simplify and reposition common analytic methods
2. Compare methods to estimate HDs
3. Suggest cross-pollinations
4. Encourage disparities investigations

One intuitive model of Health Disparities

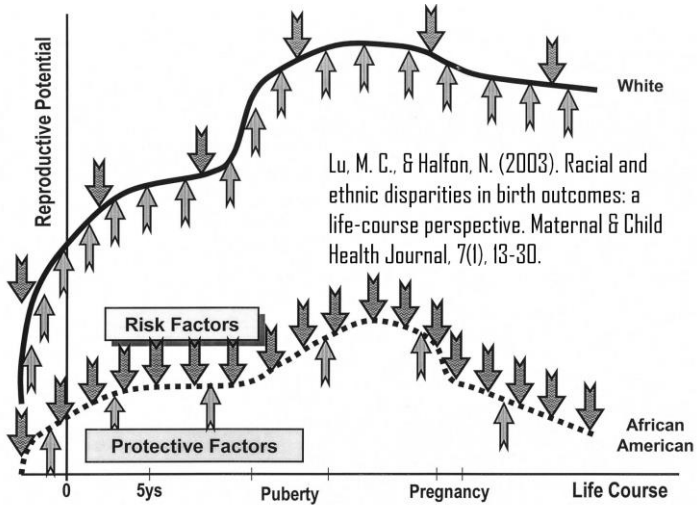


Fig. 1. How differential exposures to risk factors (downward arrows) and protective factors (upward arrows) over the life course affect developmental trajectories and contribute to disparities in birth outcomes. The lower reproductive potential of African American women, relative to White women, results from their cumulative exposure to more risk factors and less protective factors across the life span, particularly during sensitive periods of development.

Lu, M. C., & Halfon, N. (2003). Racial and ethnic disparities in birth outcomes: a life-course perspective. *Maternal & Child Health Journal, 7(1), 13-30.*

One intuitive model of Health Disparities

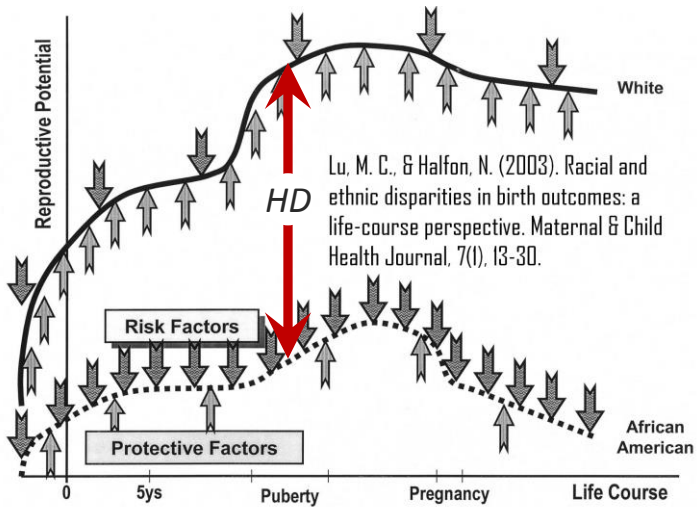
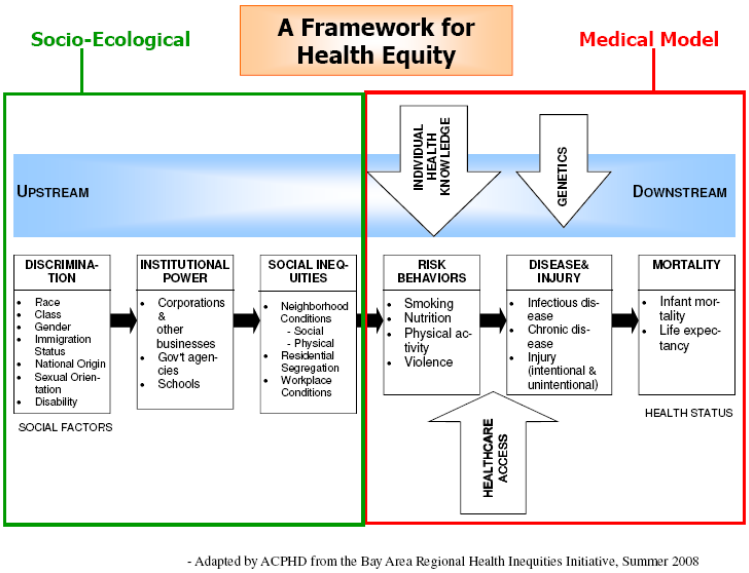


Fig. 1. How differential exposures to risk factors (downward arrows) and protective factors (upward arrows) over the life course affect developmental trajectories and contribute to disparities in birth outcomes. The lower reproductive potential of African American women, relative to White women, results from their cumulative exposure to more risk factors and less protective factors across the life span, particularly during sensitive periods of development.

Another intuitive model of HDs



- Adapted by ACPHD from the Bay Area Regional Health Inequities Initiative, Summer 2008

Tony Iton "Framework for Health Equity", referenced in the text: *Tackling Health Inequities Through Public Health Practice: Theory to Action*. Edited by Richard Hofrichter and Rajiv Bhatia. Based on a project by the National Association of County and City Health Officials. Oxford University Press, 2010 (page 380).
Iton, A., & Shrimali, B. P. (2016). Power, Politics, and Health: A New Public Health Practice Targeting the Root Causes of Health Equity. *Maternal and Child Health Journal, 20(8), 1753-1758. doi:10.1007/s10995-016-1980-6*

A testable model of HDs

Black vs. White → HD → Anxiety

Others → Anxiety

Caveats:

1. Causality: can race 'cause' health?
2. Compare the comparable
3. Control for 'covariates'

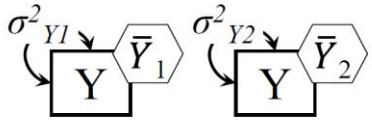
VanderWeele TJ, & Hernán MA. (2012). Causal effects and natural laws: towards a conceptualization of causal counterfactuals for non-manipulable exposures with application to the effects of race and sex. In Berzuini C, Dawid P, & Bernardinelli L (Eds.), *Causality: Statistical Perspectives and Applications (pp. 101–113). West Sussex, UK John Wiley & Sons*.
VanderWeele, T. J., & Robinson, W. R. (2014). On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology, 25(4), 473-484.*

Modeling options for HDs

1. Independent samples t-test
2. Anova
3. Regression
4. Instrumental Variable regression
5. SEM
6. Matching methods
7. +

Modeling options for HDs

1. Independent samples T-test is a 2-group model



Y_1 eq. (3a.1)

Y_2 eq. (3a.2)

Figure 3a: The t-test model

Note: The independent samples t-test is testing the hypothesis: $\bar{Y}_1 = \bar{Y}_2$; this two-group setup allows for inclusion of group specific covariates; the two equations are simply one variable for each group, but across-group constraints are possible, like $\sigma^2_{Y1} = \sigma^2_{Y2}$.

Coman, E. N., Suggs, L. S., Iordache, E., Coman, M. A., & Fifield, J. (2015). A Review of Graphical Approaches to Common Statistical Analyses. The Omnipresence of Latent Variables in Statistics *International Journal of Clinical Biostatistics and Biometrics, 1(1), 1-9.*

Modeling options for HDs

1. Independent samples T-test is a 2-group model

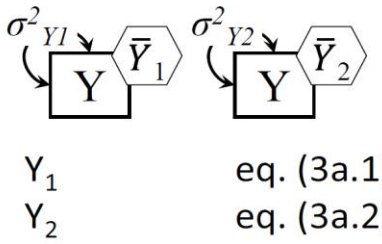
```
ttest y, by(binary)
pwmean y , over(binary) effects
```

2. Anova – similar for 2 groups

```
anova y binary
```

- Allows for covariates though

```
anova y binary c.x1 c.x2
```



Y_1 eq. (3a.1)

Y_2 eq. (3a.2)

Figure 3a: The t-test model

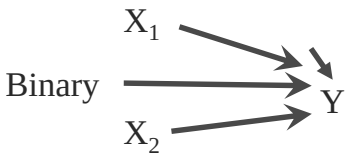
Note: The independent samples t-test is testing the hypothesis: $\bar{Y}_1 = \bar{Y}_2$; this two-group setup allows for inclusion of group specific covariates; the two equations are simply one variable for each group, but across-group constraints are possible, like $\sigma^2_{Y1} = \sigma^2_{Y2}$.

Coman, E. N., Suggs, L. S., Iordache, E., Coman, M. A., & Fifield, J. (2015). A Review of Graphical Approaches to Common Statistical Analyses. The Omnipresence of Latent Variables in Statistics *International Journal of Clinical Biostatistics and Biometrics, 1(1), 1-9.*

Modeling options for HDs

3. Regression

```
reg y binary  x1 x2
```



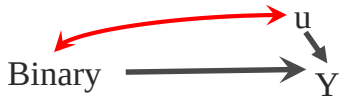
But: ? Can race cause health?

VanderWeele TJ, & Hernán MA. (2012). Causal effects and natural laws: towards a conceptualization of causal counterfactuals for non-manipulable exposures with application to the effects of race and sex. In Berzuini C, Dawid P, & Bernardinelli L (Eds.), *Causality: Statistical Perspectives and Applications (pp. 101–113). West Sussex, UK John Wiley & Sons*.

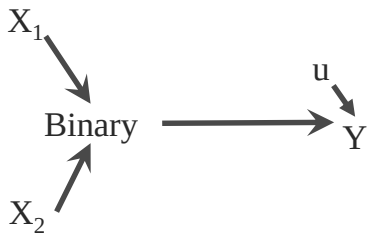
VanderWeele, T. J., & Robinson, W. R. (2014). On the causal interpretation of race in regressions adjusting for confounding and mediating variables. *Epidemiology, 25(4), 473-484.*

Modeling options for HDs

4. IV regression [Cameron]
- Binary correlated with u



- Need IV->X but not IV->Y



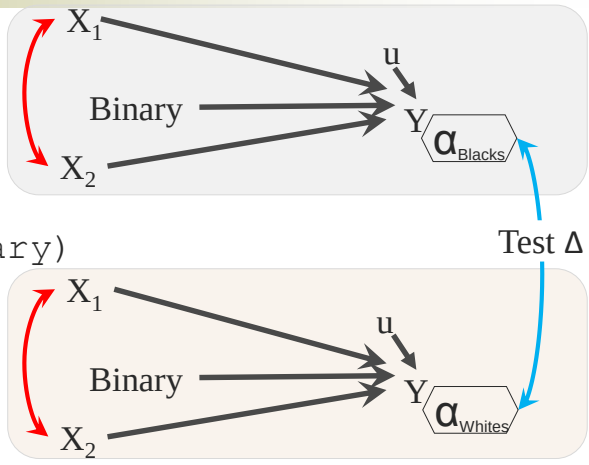

```
ivregress 2sls y
(binary = x1 x2), first
```

- Tests the IVs:

```
estat endogenous
```


Modeling options for HDs

5. SEM
2-group model, compare intercepts (needs centering)




```
sem y <- x1 x2, group(binary)
ginvariant(covex)
```

Can relax the equal Y variances assumption

Modeling options for HDs

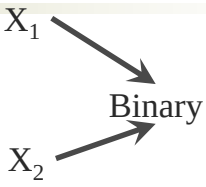
6. Matching options

```
teffects psmatch (y)
(binary  x1 x2 )
```

```
psmatch2 binary  x1 x2 ,
outcome (y) common ate
```

```
attnd y binary  x1 x2
```

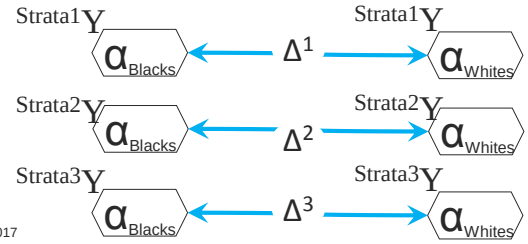
Stage 1



Stage 2

$Match\ in\ strata\ on\ \widehat{Binary}$

Stage 3



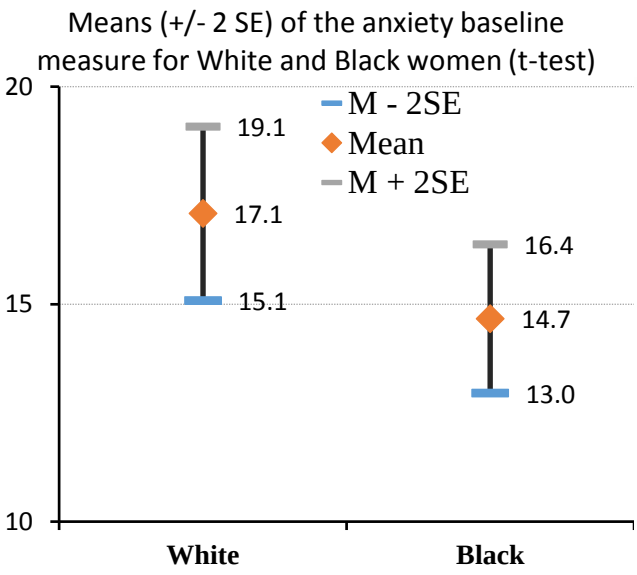
Modeling options for HDs

1. T-test

2. Anova – identical

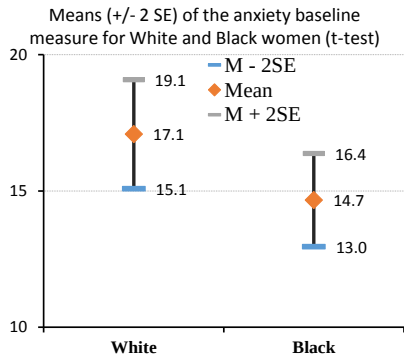
- Allows for covariates though

```
anova y binary c.x1 c.x2
```

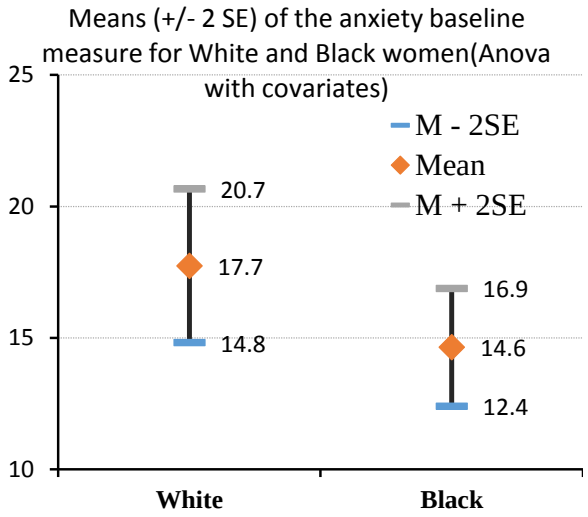


Modeling options for HDs

T-test

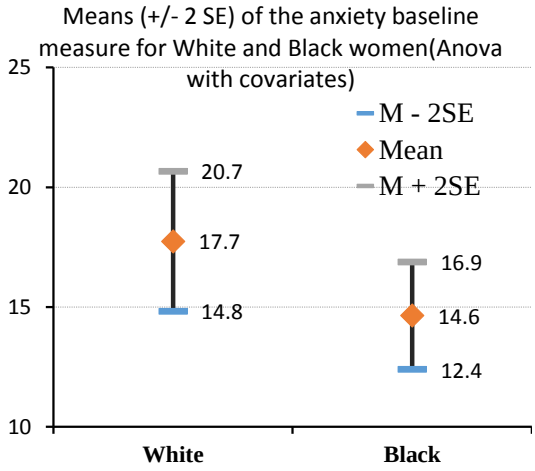


Anova – with covariates

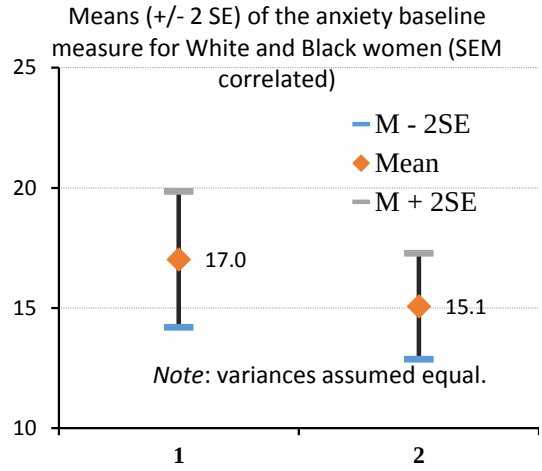


Modeling options for HDs

Anova – with covariates



SEM with *correlates*



Modeling options for HDs

SEM with *covariates* [0=White; 1=Black]; variances freed

```
sem y <- x1 x2, group(binary) ginvariant(covex)
              |                 OIM
              |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------+------------------------------------------------------------------
Structural    |
  y <-        |
          x1  |
           0  |   .0199704    2.16228     0.01   0.993     -4.21802    4.257961
           1  |    2.61248   1.673379     1.56   0.118    -.6672832    5.892244
          x2  |
           0  |    .164855   .3766247     0.44   0.662    -.5733159    .9030258
           1  |   .4905938   .3001652     1.63   0.102    -.0977191    1.078907
       _cons  |
           0  |   13.43964   8.259071     1.63   0.104    -2.747845    29.62712
           1  |   3.422483   6.895766     0.50   0.620    -10.09297    16.93794
--------------+------------------------------------------------------------------
     var(e.y) |
           0  |   60.78133   13.59112                      39.21351    94.21166
           1  |   84.67747   14.74046                      60.19947    119.1086
------means centered at .050 (X1 neighborhood disorder) and 22.91y (X2=age)-------
```


Modeling options for HDs

teffects psmatch

```
. teffects psmatch (y) (binary  x1 x2 )

Treatment-effects estimation                   Number of obs      =        106
Estimator      : propensity-score matching     Matches: requested =          1
Outcome model  : matching                                     min =          1
Treatment model: logit                                        max =          1
------------------------------------------------------------------------------
             |               AI Robust
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ATE          |
      binary |
   (1 vs 0)  |  -3.169811   2.228166    -1.42   0.155    -7.536936    1.197313
------------------------------------------------------------------------------
```

Modeling options for HDs

psmatch2

```
psmatch2 binary  x1 x2 ,  outcome (y) common ate  Probit regression
Number of obs     =       106
                                                  LR chi2(2)        =     12.08
                                                  Prob > chi2       =    0.0024
Log likelihood = -64.210562                       Pseudo R2         =    0.0860
------------------------------------------------------------------------------
      binary |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .6428318   .2054421     3.13   0.002     .2401726    1.045491
          x2 |   .0468275   .0366658     1.28   0.202    -.0250361    .1186911
       _cons |  -.7222502   .8205248    -0.88   0.379    -2.330449    .8859489
------------------------------------------------------------------------------

 psmatch2: |   psmatch2: Common
 Treatment |        support
assignment | Off suppo  On suppor |     Total
-----------+----------------------+----------
 Untreated |         0         40 |        40
   Treated |        11         55 |        66
-----------+----------------------+----------
     Total |        11         95 |       106
```


Modeling options for HDs

psmatch2

```
psmatch2 binary  x1 x2 ,  outcome (y) common ate  Probit regression
Number of obs     =       106


        Variable     Sample |    Treated     Controls   Difference          S.E.   T-stat
-----------------------------+------------------------------------------------------------
               y  Unmatched | 15.0757576       17.025  -1.94924242    1.8091337    -1.08
                        ATT |       13.8  18.9272727  -5.12727273   2.19360704    -2.34
                        ATU |     17.025         12.5       -4.525            .        .
                        ATE |                           -4.87368421            .        .
-----------------------------+------------------------------------------------------------
```

Modeling options for HDs

attnd

```
attnd y binary  x1 x2  , comsup boot reps(100) dots logit  detail
Note: the common support option has been selected
 The region of common support is [.30252502, .89642859]
 The distribution of the pscore is
                        Pr(binary)
-------------------------------------------------------------
      Percentiles      Smallest
 1%     .3173758        .302525
 5%     .3990624       .3173758
10%     .4397769       .3197376       Obs                 141
25%     .4943597       .3332007       Sum of Wgt.         141

50%     .6010811                      Mean           .6028369
                       Largest        Std. Dev.      .1374292
75%     .7131706        .854728
90%     .7813184       .8586264       Variance       .0188868
95%     .8165104       .8644204       Skewness      -.0160976
99%     .8644204       .8964286       Kurtosis        2.2178
 The program is searching the nearest neighbor of each treated unit.
 This operation may take a while.
```


Modeling options for HDs

attnd

```
attnd y binary  x1 x2  , comsup boot reps(100) dots logit  detail


ATT estimation with Nearest Neighbor Matching method
(random draw version)
Analytical standard errors
-------------------------------------------------------------
n. treat.    n. contr.         ATT     Std. Err.          t
-------------------------------------------------------------

       85           23      -0.524        2.008     -0.261

-------------------------------------------------------------
```

Modeling options for HDs

ivregress

```
ivregress 2sls y (binary = x1 x2 ), first

First-stage regressions
-----------------------
                                                 Number of obs     =        106
                                                 F(   2,    103)    =       6.05
                                                 Prob > F          =     0.0033
                                                 R-squared         =     0.1051
                                                 Adj R-squared     =     0.0878
                                                 Root MSE          =     0.4652
 ------------------------------------------------------------------------------
      binary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   .2176514   .0672757     3.24   0.002     .0842259    .3510769
          x2 |   .0148822   .0124655     1.19   0.235    -.0098403    .0396046
       _cons |   .2805711    .281939     1.00   0.322    -.2785883    .8397305
------------------------------------------------------------------------------
```


Modeling options for HDs

ivregress

```
ivregress 2sls y (binary = x1 x2 ), first

Instrumental variables (2SLS) regression          Number of obs   =        106
                                                  Wald chi2(1)    =       1.57
                                                  Prob > chi2     =     0.2098
                                                  R-squared       =          .
                                                  Root MSE        =     10.126

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      binary |   7.847662   6.257221     1.25   0.210    -4.416267    20.11159
       _cons |   10.92504   4.018222     2.72   0.007      3.04947    18.80061
------------------------------------------------------------------------------

Instrumented:  binary
Instruments:   x1 x2
```

Modeling options for HDs

ivregress

```
estat endogenous

  Tests of endogeneity
  Ho: variables are exogenous

  Durbin (score) chi2(1)          =  3.51188  (p = 0.0609)
  Wu-Hausman F(1,103)             =  3.52942  (p = 0.0631)
```

"If the test statistic is significant, the variables must b e treated as endogenous " URL

```
estat overid
  Tests of overidentifying restrictions:

  Sargan (score) chi2(1) =  .872447  (p = 0.3503)
  Basmann chi2(1)        =  .854791  (p = 0.3552)
```

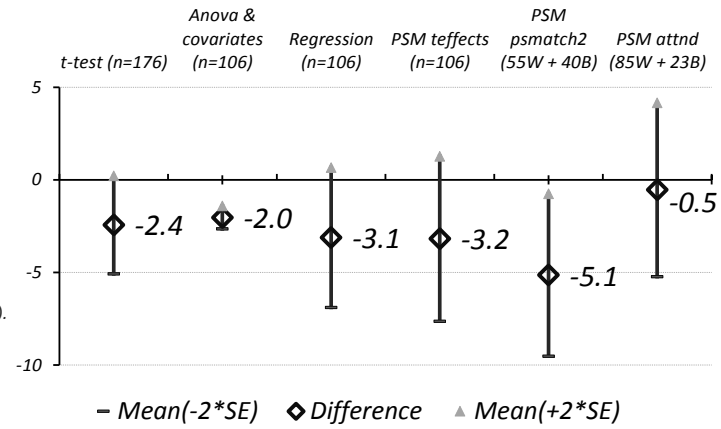
"A statistically significant test statistic indicates that the instruments may not be valid" URL

Modeling options for HDs

Comparisons

Comparison of racial/ethnic (R/E) differences (92 White- 145 Black) in Anxiety baseline scores, by estimation method

All models used age and neighborhood stress as: covariates (regression); propensity predictors (PSM); instruments for R/E (IV).



Some methods are 'better' than others in Identifying health disparities.

SEM is not shown, the test of HD is a Wald (or chi-squared test).

Conclusions – email for >: comanus@gmail.com

- Some methods are ‘better’ than others in identifying health disparities.
- Comparisons encourage analysts to think about ‘the place’ of the variables in the analytical model.
- One general insight: we can trust estimates when that are somewhat consistent across methods.