# Project Summary

UNC EPID 722 – Sydney Allison Jones & Ann Von Holle

April 26, 2016

# Topics

- Data generation
- Data analysis
- Data analysis follow-up

# Data generation

- Generate time to event using National Ambulatory Medical Care Survey NAMCS data:
- Introduce
  - *Non-positivity*
    - Exclude persons with hypertension who were not treated
  - *Confounding*
  - *Treatment effect heterogeneity*
    - Low-potency statins had no effect
- Subset to a 10% random sample of the non-users
  - *To make data set a more manageable size.*

# Generate time to event for CVD hospitalization and all-cause mortality

- Time to event will follow a Weibull distribution
- T $\sim$ Weibull($\lambda, k$)
  - *scale* = $\lambda$
  - *shape* = $k$
    - Weibull density function: $f(t; \lambda, k) = k\lambda t^{k-1} e^{-\lambda t^k}$
    - Exponential density function: $f(x; \lambda) = \lambda e^{-\lambda t}$
      - Equivalent to Weibull using shape = 1.
- Censoring follows an exponential distribution with $\lambda$ = 0.005

# CVD hospitalization

- Time to hospitalization, T $\sim$ Weibull($\lambda, k$):
  - $\eta_1 = \beta_1 \cdot \text{high dose} + \beta_2 \cdot \text{low dose} + \beta_3 \cdot \text{age} + \beta_4 \cdot \text{hyplipid} + \beta_5 \cdot \text{htn} + \beta_6 \cdot \text{diabetes}$
  - *scale = $\lambda$ = 0.011*
  - *shape = k = 1*
  - *Survival time corresponding to Cox-Weibull model:* $T = \left(-\frac{log(U)}{\lambda exp(\eta_1)}\right)^{1/k}$ *(1)*
- Parameter values
  - $\beta_1$ *= log(0.5) = -0.69*
  - $\beta_2$ *= 0*
  - $\beta_3 = \beta_4 = \beta_5 = \beta_6$ *= log(1.5) = 0.41*

# All-cause mortality

- Time to death
  - $\eta_2 = \beta_1 \cdot \text{age} + \beta_2 \cdot \text{smoke} + \beta_3 \cdot \text{obese}$
  - *scale for the Weibull distribution = 0.012*
  - *shape = k = 1*
  - *Survival time corresponding to Cox-Weibull model:* $T = \left(-\frac{log(U)}{\lambda exp(\eta_2)}\right)^{1/k}$ *(1)*
- Parameter Values
  - $\beta_1$ *= log(1.1) = 0.1*
    - Age = ((years-53)/10)
  - $\beta_2$ *= log(1.5) = 0.41*
  - $\beta_3$ *= log(1.4) = 0.34*

# Data generation key points

- *True* hazard ratio = 0.801
  - *log(HR) = -0.222; standard error = 0.093.*
- **No** variables associated with drop out
- **No** variables associated with missing data in effect model
- **No** product term between any variable and time to event
  - *proportional hazards assumption should hold*

# Data analysis

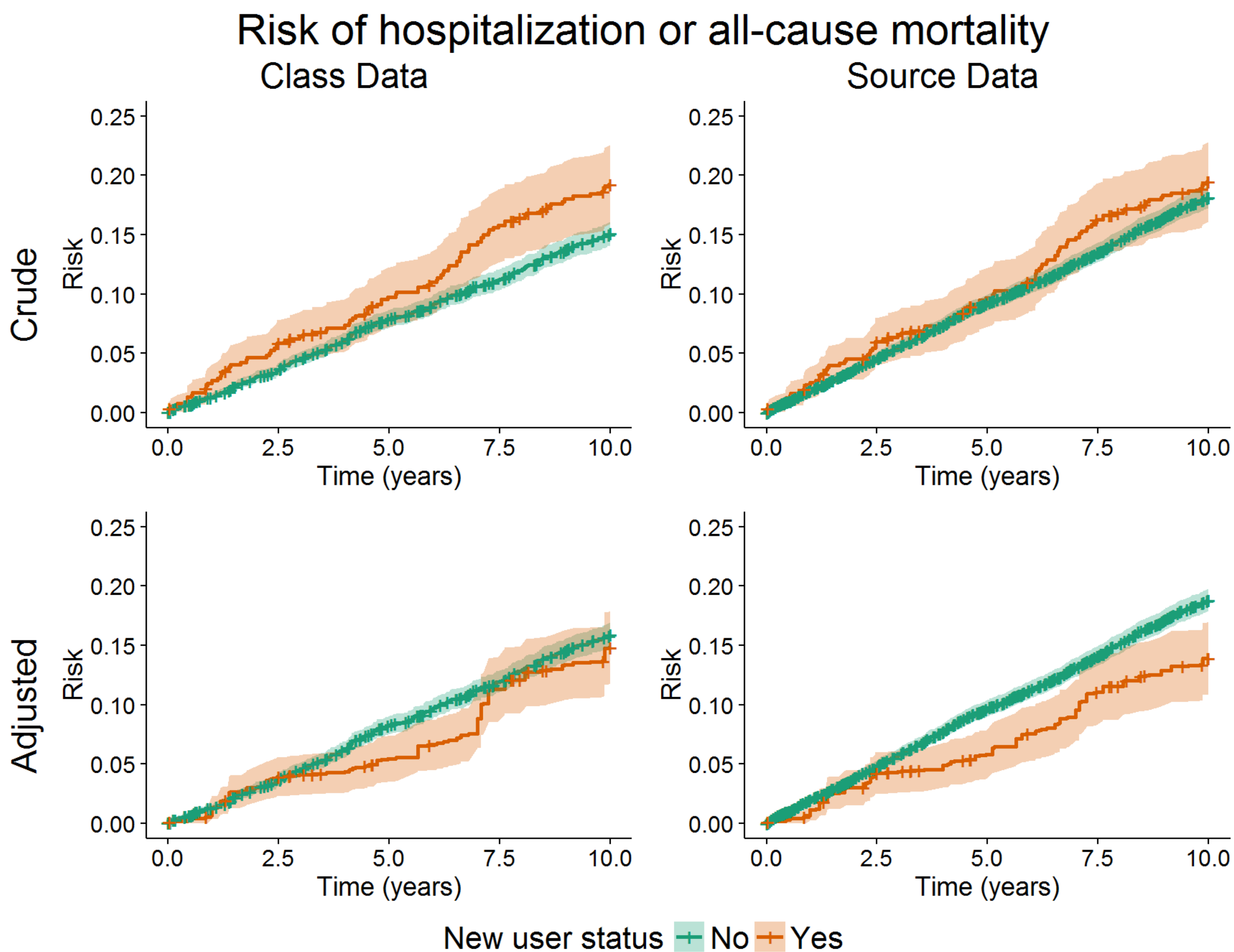
# Our analyses on entire class data set

HR and 95% CI (treated vs not treated)

| | Method for missing | | |
|---|---|---|---|
| Confounding + Selection bias handling | Omit BP | Impute | Missing weight |
| Crude | 1.34 (1.09, 1.65) | 1.34 (1.09, 1.65) | 1.31 (1.05, 1.64) |
| IPTW | 0.78 (0.54, 1.12) | 0.78 (0.54, 1.12) | 0.92 (0.62, 1.37) |
| IPTW + censor weight | 0.78 (0.54, 1.12) | 0.78 (0.54, 1.12) | 0.92 (0.62, 1.37) |

- Covariates for:
  - *IPTW model: age, male, diabetes, white, obese, smoke, hyplipid [sbp and dbp when using MI or missing weights]*
  - *Censoring weight model: age, male, diabetes, white, obese, smoke, hyplipid [sbp and dbp when using MI or missing weights]*
  - *Missing weight model: newuser, age, male, diabetes, white, obese, smoke, hyplipid, htn, event*

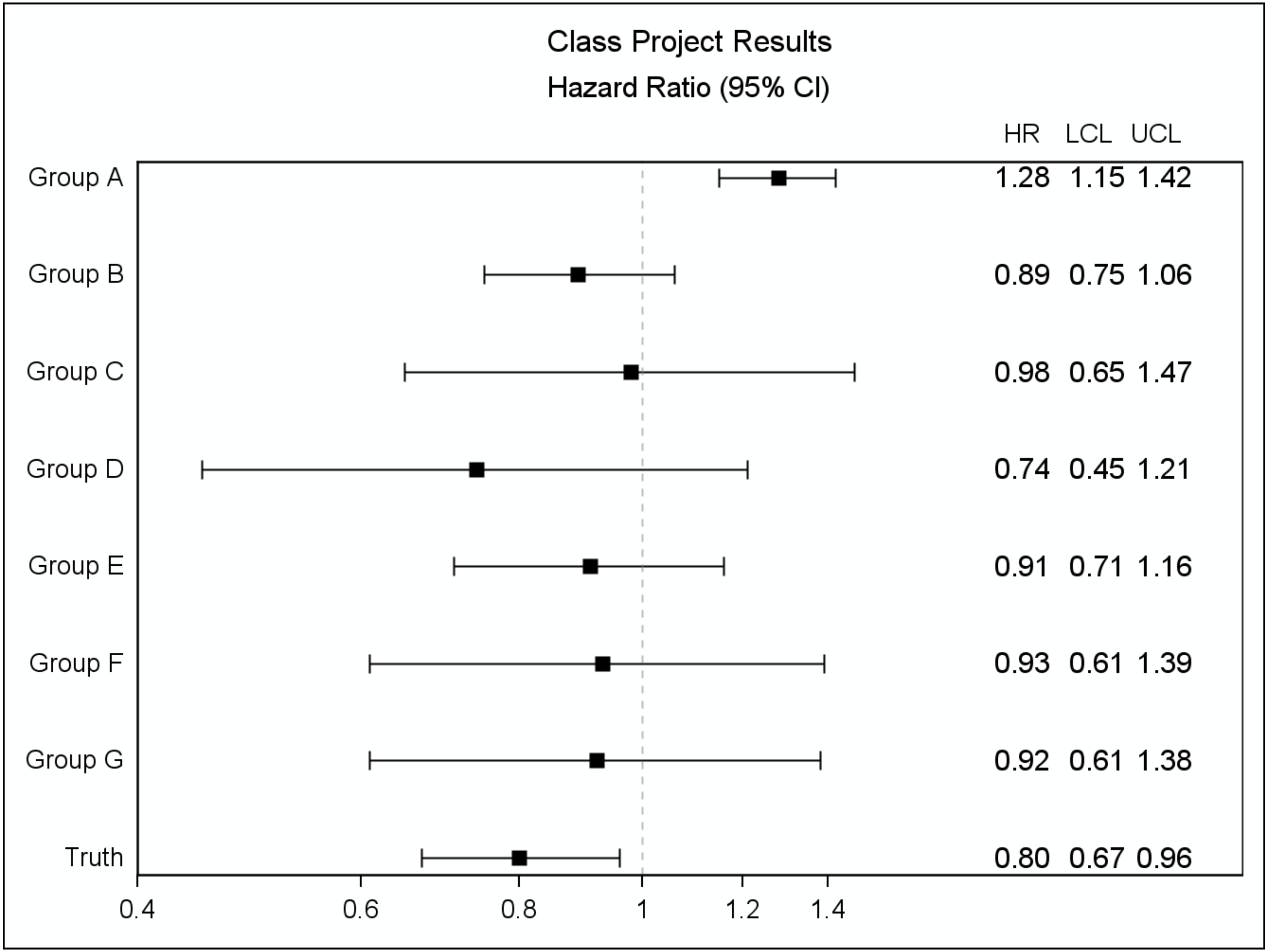
# Our survival curves by treatment status and data set type



# Summary of group analyses

- Approaches used
  - *Missing data: inverse probability weighting (5 groups) or multiple imputation (2 groups)*
  - *Censoring: time-varying inverse probability weighting (7 groups)*
  - *Wide range of variables included in censoring weights*
  - *Confounding: inverse probability weighting (7 groups)*
  - *Most groups included: age, diabetes, DBP, SBP, obese, race, sex, and smoking*
  - *No groups included HTN in the confounding weight model*
- Results presented:
  - *Hazard ratio (5 groups)*
  - *Risk difference (2 groups)*

# Class group results


Class Project Results
Hazard Ratio (95% CI)
HR LCL UCL
Group A 1.28 1.15 1.42
Group B 0.89 0.75 1.06
Group C 0.98 0.65 1.47
Group D 0.74 0.45 1.21
Group E 0.91 0.71 1.16
Group F 0.93 0.61 1.39
Group G 0.92 0.61 1.38
Truth 0.80 0.67 0.96
0.4
0.6
0.8
1
1.2
1.4

# Class group example: Table 2

**Table 2.** Association of Statin Initiation With CVD Hospitalization or Death in the National Ambulatory Medical Care Survey, 2005-2009

| Exposure | No. Events | Total N | 10-year Risk | 95% CI[a] | 10-year RD | 95% CI[a] |
|---|---|---|---|---|---|---|
| Crude analysis | | | | | | |
| Non statin users | 755 | 5,146 | 0.15 | 0.14, 0.16 | 0 | Reference |
| Statin initiators | 105 | 552 | 0.19 | 0.16, 0.23 | 0.04 | 0.01, 0.08 |
| Weighted analysis[b] | | | | | | |
| Non statin users | 429.72 | 2,790.12 | 0.16 | 0.14, 0.18 | 0 | Reference |
| Statin initiators | 42.77 | 299.64 | 0.15 | 0.10, 0.20 | -0.01 | -0.07, 0.05 |

Abbreviations: N, number of observations; RD, risk difference; CI, confidence interval; CVD, cardiovascular disease; BP, blood pressure

[a] Confidence intervals were obtained using the standard deviation from 100 bootstrap resamples.

[b] Stabilized inverse probability weights were used to weight the complete cases ($n$=3,135) to control for: missing BP in the full sample (adjusting for statin initiation, hypertension, diabetes, hyperlipidemia, race, sex, obesity, smoking, and age); confounding (adjusting for diabetes, hyperlipidemia, race, sex, obesity, smoking, age, diastolic BP, and systolic BP); and time-varying censoring (adjusting for statin initiation, diabetes, hypertension, hyperlipidemia, race, sex, obesity, smoking, age, diastolic BP, and systolic BP).

# Class group example: Figure

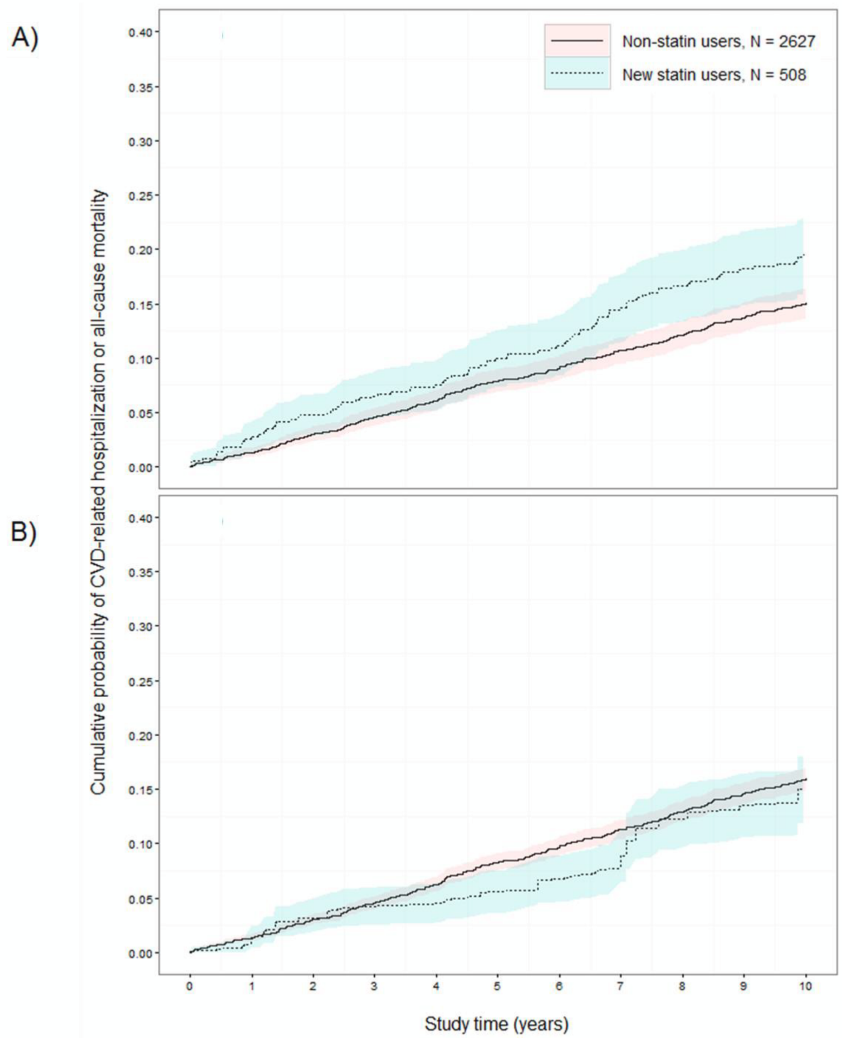



**FIGURE 1**. Crude (A) and adjusted (B) cumulative incidence and robust 95% confidence intervals of 10-year CVD-related hospitalization or all-cause mortality, by statins initiation at baseline in the National Ambulatory Medical Care Survey, 2000-2004. Restricted to 3135 participants with data on all covariates. Full model (B) adjusted for missing data, dropout, and for confounding by age, sex, race, blood pressure, diabetes, hyperlipidemia, obesity and cigarette use using inverse-probability weighting.

# A few points given the model generating the data

# Censoring weights don't (shouldn't) matter

HR and 95% CI (treated vs not treated)

| | Method for missing | | |
|---|---|---|---|
| Confounding + Selection bias handling | Omit BP | Impute | Missing weight |
| Crude | 1.34 (1.09, 1.65) | 1.34 (1.09, 1.65) | 1.31 (1.05, 1.64) |
| IPTW | 0.78 (0.54, 1.12) | 0.78 (0.54, 1.12) | 0.92 (0.62, 1.37) |
| IPTW + censor weight | 0.78 (0.54, 1.12) | 0.78 (0.54, 1.12) | 0.92 (0.62, 1.37) |

# Accounting for missing data shouldn't matter

HR and 95% CI (treated vs not treated)

| | Method for missing | | |
|---|---|---|---|
| Confounding + Selection bias handling | Omit BP | Impute | Missing weight |
| Crude | 1.34 (1.09, 1.65) | 1.34 (1.09, 1.65) | 1.31 (1.05, 1.64) |
| IPTW | 0.78 (0.54, 1.12) | 0.78 (0.54, 1.12) | 0.92 (0.62, 1.37) |
| IPTW + censor weight | 0.78 (0.54, 1.12) | 0.78 (0.54, 1.12) | 0.92 (0.62, 1.37) |

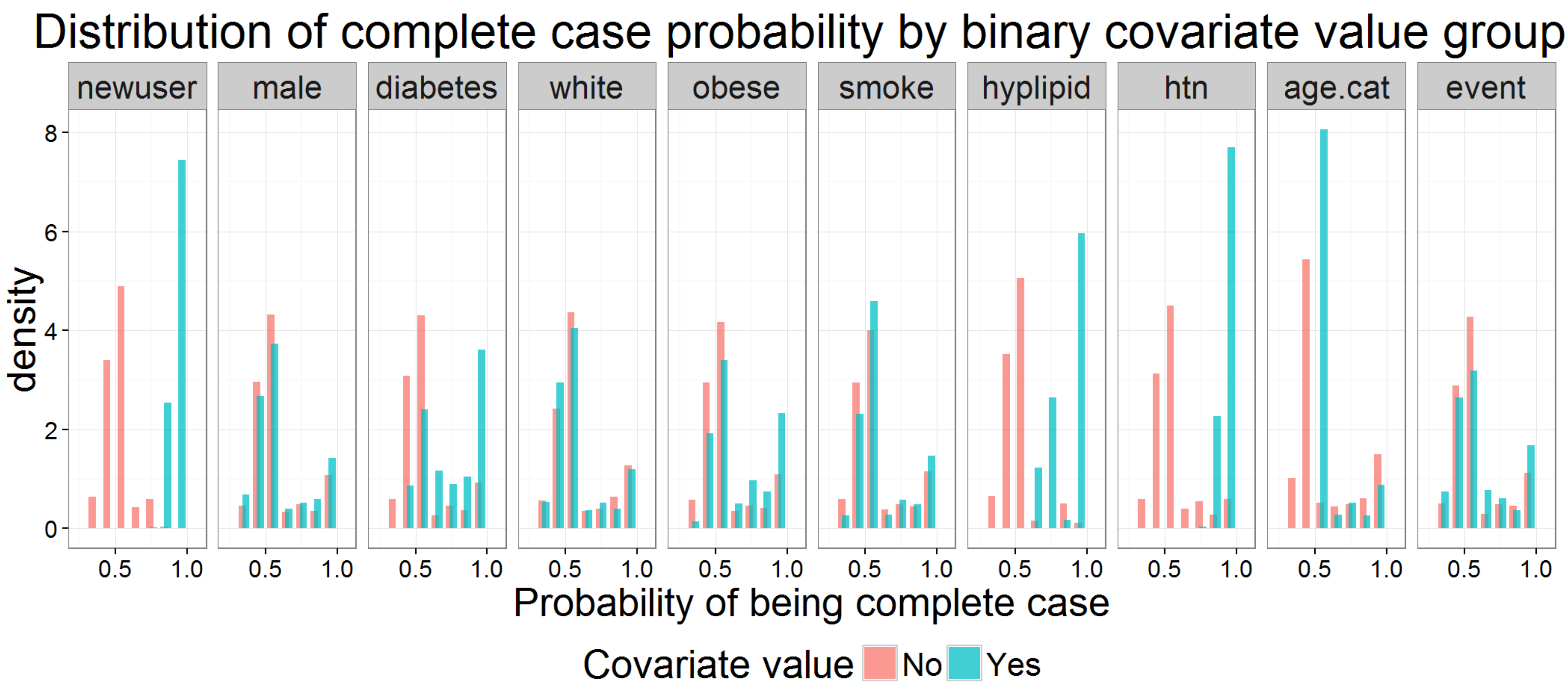
# Why are there differences in treatment effect when using missing weights?

HR and 95% CI (treated vs not treated)

| | Method for missing | | |
|---|---|---|---|
| Confounding + Selection bias handling | Omit BP | Impute | Missing weight |
| Crude | 1.34 (1.09, 1.65) | 1.34 (1.09, 1.65) | 1.31 (1.05, 1.64) |
| IPTW | 0.78 (0.54, 1.12) | 0.78 (0.54, 1.12) | 0.92 (0.62, 1.37) |
| IPTW + censor weight | 0.78 (0.54, 1.12) | 0.78 (0.54, 1.12) | 0.92 (0.62, 1.37) |

# What's happening with the missing weights?

- People with adverse conditions (hypertension, etc..) are more likely to be a complete case.



# What's happening with the missing weights?

- People with hypertension more likely to benefit from treatment than people with no hypertension.

HR and 95% CI (treated vs not treated)

| | Method for missing |
|---|---|
| Confounding + Selection bias handling | Omit BP |
| **HTN=1 subset (Source data)** | |
| Crude | 0.75 (0.57, 0.98) |
| IPTW | 0.60 (0.42, 0.87) |
| IPTW + censor weight | 0.60 (0.41, 0.86) |

# Non-Positivity

- Several of you noticed the non-positivity
- What can you do about it?

# Non-Positivity

## What is the impact on our inferences?

HR and 95% CI (treated vs not treated)

| | Method for missing | | |
|---|---|---|---|
| Confounding + Selection bias handling | Omit BP | Impute | Missing weight |
| **Class data** | | | |
| Crude | 1.34 (1.09, 1.65) | 1.34 (1.09, 1.65) | 1.31 (1.05, 1.64) |
| IPTW | 0.78 (0.54, 1.12) | 0.78 (0.54, 1.12) | 0.92 (0.62, 1.37) |
| IPTW + censor weight | 0.78 (0.54, 1.12) | 0.78 (0.54, 1.12) | 0.92 (0.62, 1.37) |
| **HTN=0 subset** | | | |
| Crude | 1.19 (0.88, 1.6) | 1.19 (0.88, 1.6) | 1.05 (0.75, 1.48) |
| IPTW | 0.72 (0.43, 1.2) | 0.72 (0.43, 1.2) | 0.71 (0.39, 1.27) |
| IPTW + censor weight | 0.72 (0.43, 1.2) | 0.72 (0.43, 1.2) | 0.71 (0.39, 1.27) |
| **Source data** | | | |
| Crude | 1.09 (0.89, 1.33) | 1.09 (0.89, 1.33) | 1.06 (0.86, 1.32) |
| IPTW | 0.72 (0.52, 0.99) | 0.72 (0.52, 0.99) | 0.74 (0.53, 1.03) |
| IPTW + censor weight | 0.72 (0.52, 0.99) | 0.72 (0.52, 0.99) | 0.74 (0.54, 1.03) |

# References

1. Bender R, Augustin T, Blettner M. Generating survival times to simulate cox proportional hazards models. *Statistics in Medicine* [electronic article]. 2005;24(11):1713–1723. (http://doi.wiley.com/10.1002/sim.2059)

2. Seaman S, White I. Inverse Probability Weighting with Missing Predictors of Treatment Assignment or Missingness. *Communications in Statistics - Theory and Methods* [electronic article]. 2014;43(16):3499–3515. (http://www.tandfonline.com/doi/abs/10.1080/03610926.2012.700371)

3. Seaman SR, White IR, Copas AJ, et al. Combining Multiple Imputation and Inverse-Probability Weighting. *Biometrics* [electronic article]. 2012;68(1):129–137. (http://doi.wiley.com/10.1111/j.1541-0420.2011.01666.x)

4. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Statistical Methods in Medical Research* [electronic article]. 2013;22(3):278–295. (http://smm.sagepub.com/cgi/doi/10.1177/0962280210395740)