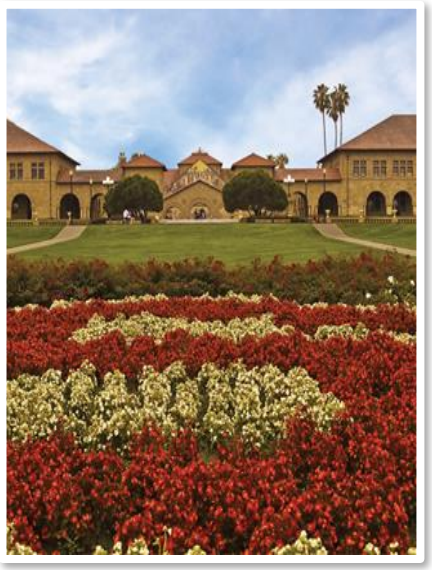
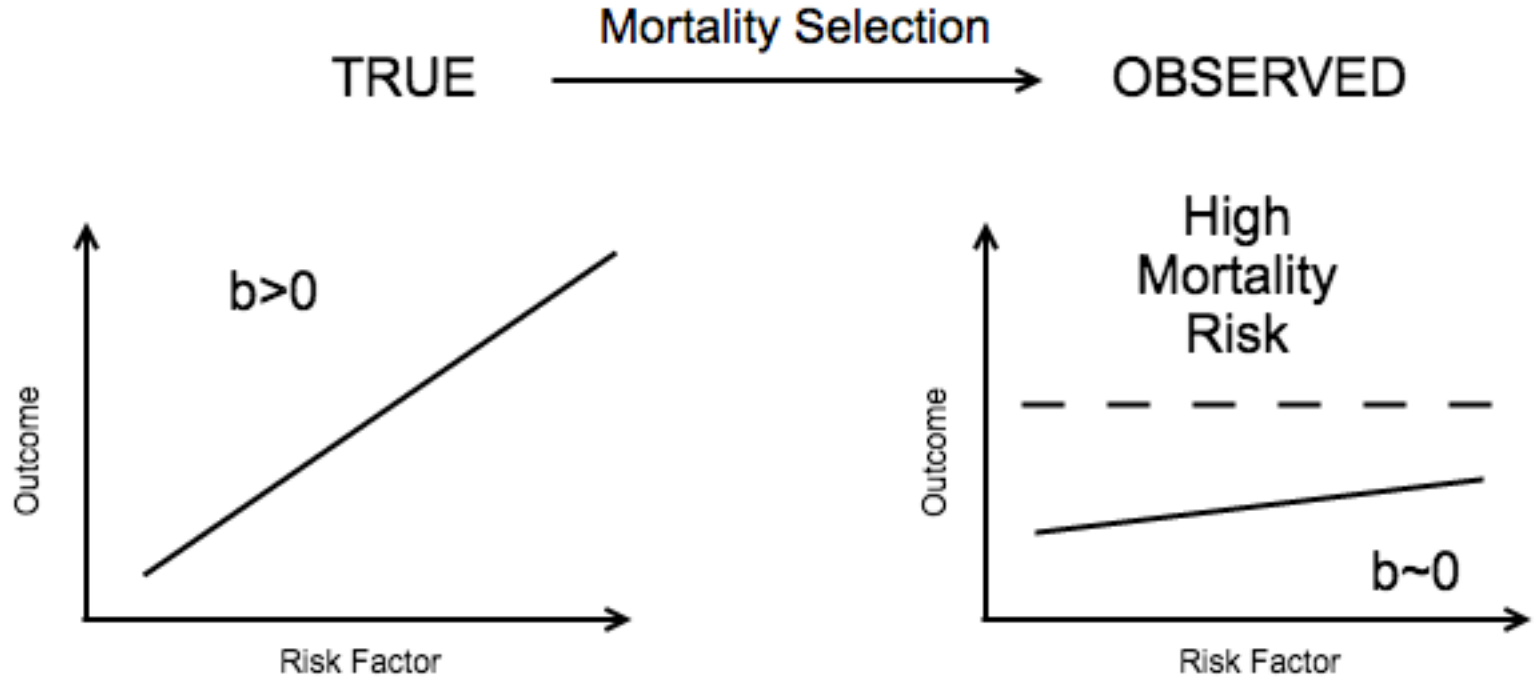
Mortality selection in a genetic sample and implications for association studies

BEN DOMINGUE



Mortality **Selection**



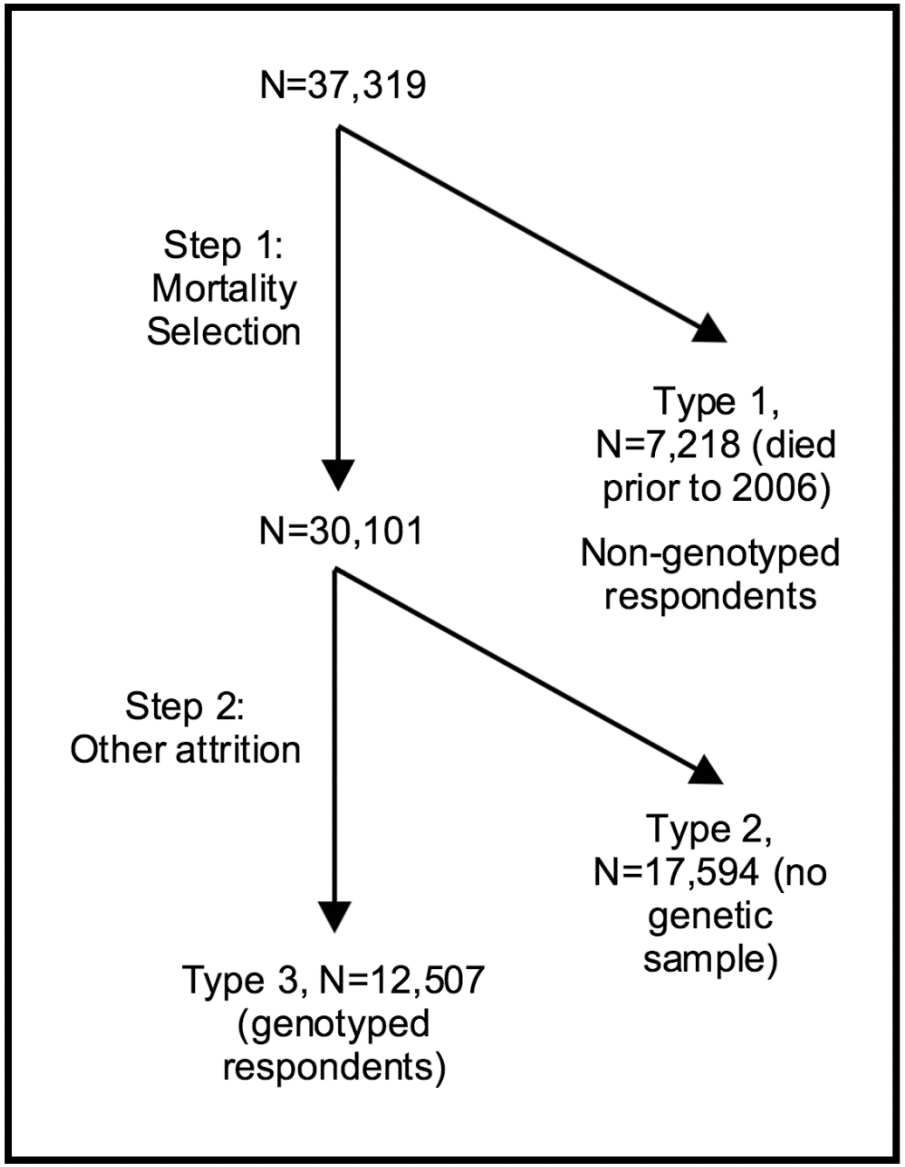
HRS | HEALTH AND RETIREMENT STUDY
A Longitudinal Study of Health, Retirement, and Aging
Sponsored by the National Institute on Aging

- Started in 1992
- Mortality Selection → “healthier, wealthier, and wiser” (Zajacova & Burgard)

- Genotyping in 2006/2008

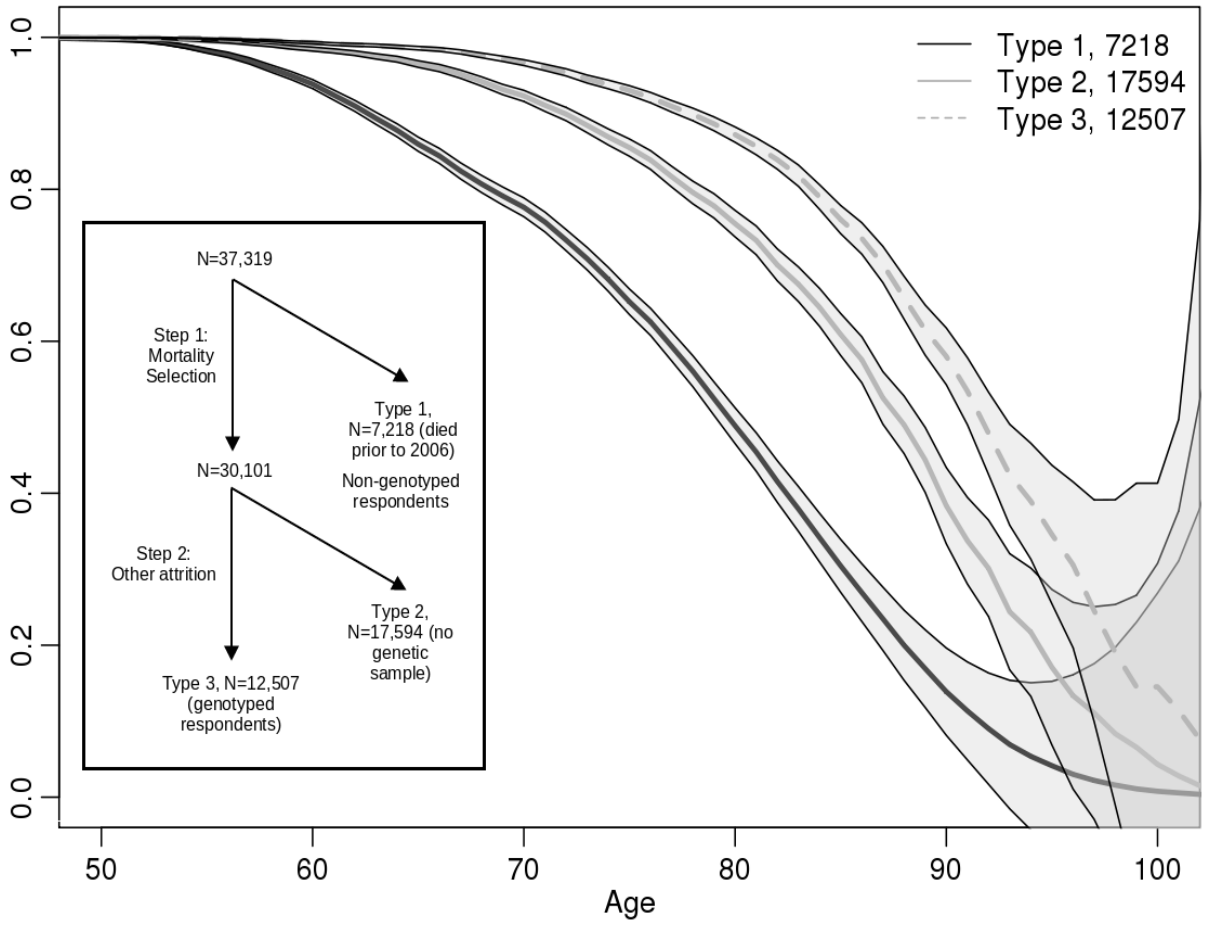
Two Steps to be in genetic sample

- Step 1: Mortality Selection
- Step 2: Genotyping in 2006/2008



Both steps are interesting

- But steps aren't due to same selective forces.
- We focus on step 1 (mortality selection).
  - IGNORE selection into genetic sample for those who lived long enough.

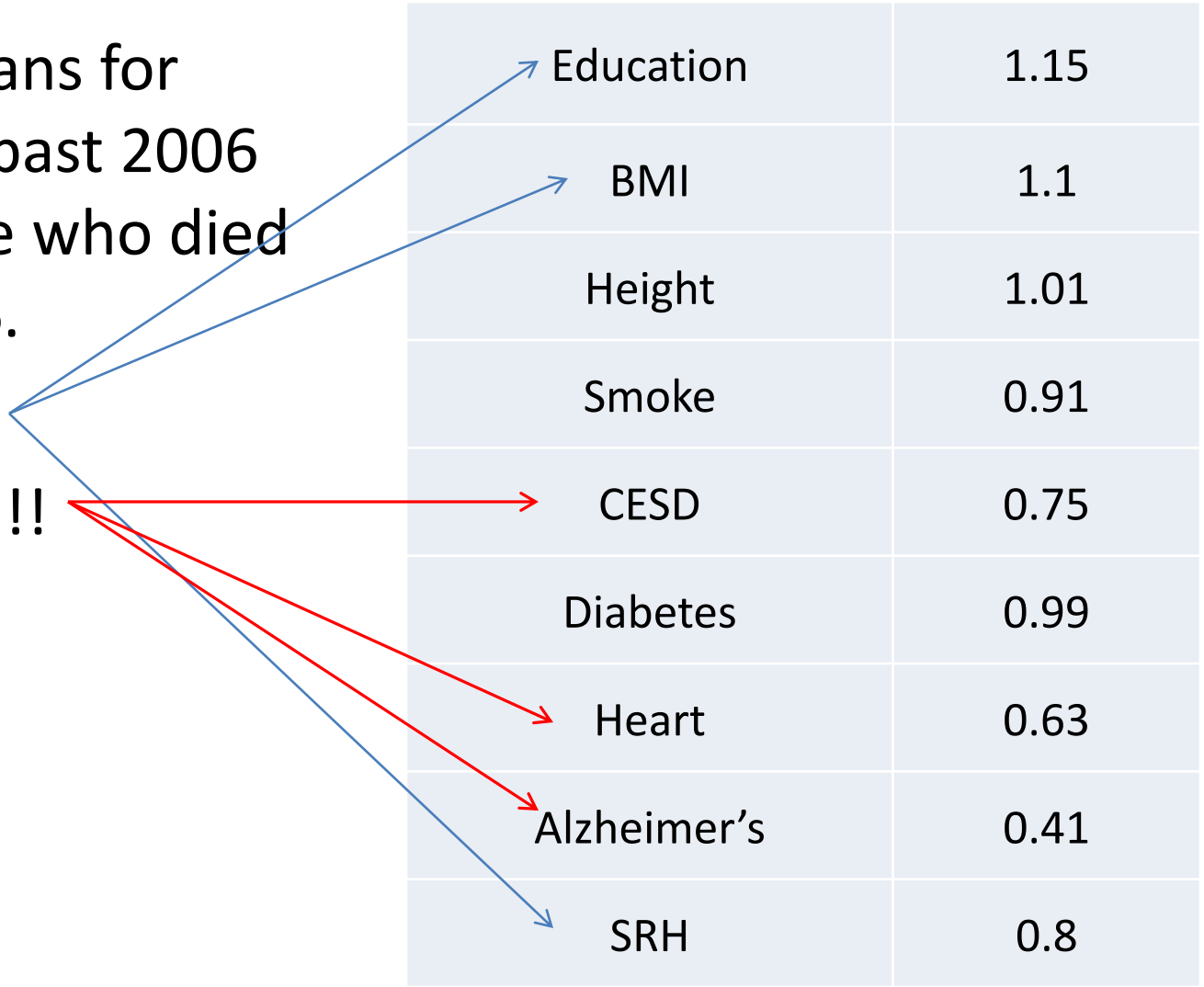


Two Questions

- How effectively can we model mortality selection?
- How might information from these models change inference about genetic associations?

Motivating models for mortality

- Ratio of means for those lived past 2006 versus those who died before 2006.
- Differences.
- Differences!!!

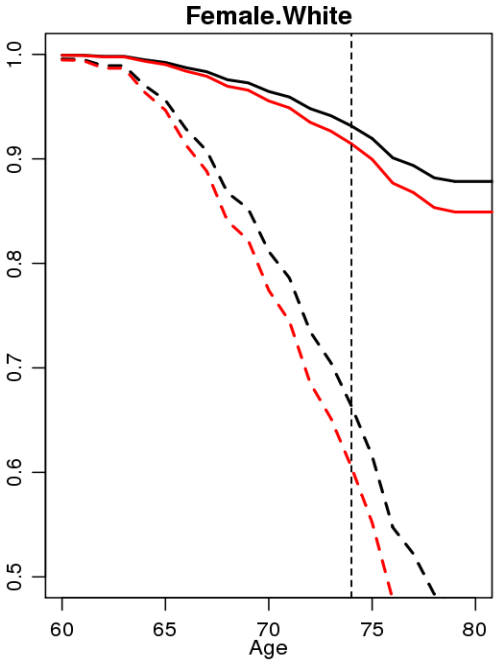


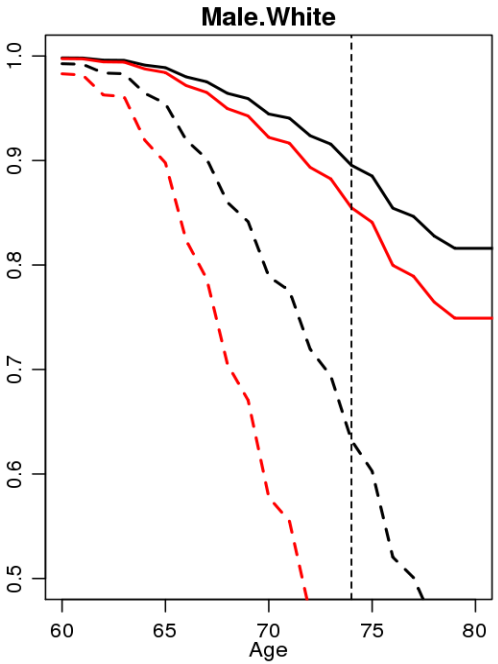
Education	1.15
BMI	1.1
Height	1.01
Smoke	0.91
CESD	0.75
Diabetes	0.99
Heart	0.63
Alzheimer's	0.41
SRH	0.8

Modeling Early Death

- Birth year is very explanatory.
- Health conditions add to this.
- But we can't improve much past that.

AUC	F Non-White	M Non-White	F White	M White
Birth year	0.867	0.874	0.843	0.829
+ health conditions	0.888	0.893	0.879	0.864
+ birth year interactions	0.89	0.892	0.88	0.864
Random Forest	0.875	0.877	0.88	0.862
N	5454	3967	11488	9170





- Cox Survival Models based on genotype status, age at first interview, birth year, and interaction of genotype status with both.
- Black 1930 birth; Red 1945 birth
- Solid genotyped; dashed-non-genotyped

Can we reduce differences in survival between genotyped and non-genotyped using our model for mortality?

Next, if we adjust association estimates to be representative of average treatment effects how do they compare to raw estimates?

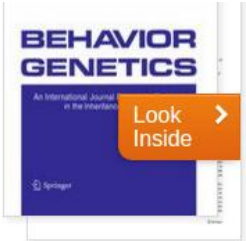
Association studies

- Examine genetic associations with outcomes.
  - BMI, Height, Education, Smoking
  - Polygenic Scores
  - Static (time-invariant) and dynamic (time-dependent)
- Before/after inverse propensity weighting.

First online: 30 July 2015

Cohort Effects in the Genetic Influence on Smoking

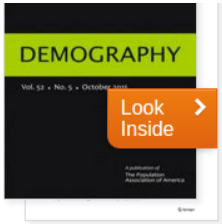
Benjamin W. Domingue ✉ , Dalton Conley, Jason Fletcher, Jason D. Boardman



First online: 29 August 2015

The Genome-Wide Influence on Human BMI Depends on Physical Activity, Life Course, and Historical Period

Guang Guo ✉ , Hexuan Liu, Ling Wang, Haipeng Shen, Wen Hu



May matter at margins for understanding static associations

- Mortality selection in height is MAR.
- Potential bias in estimates of association with smoking.

	Naïve		Weighted		**Ratio (weighted estimate/naïve estimate**
	Est	SE	Est	SE	
BMI	0.243	0.01	0.238	0.01	**0.98**
Height	0.307	0.01	0.304	0.013	**0.99**
Education	0.167	0.01	0.17	0.01	**1.016**
Smoke	0.114	0.011	0.109	0.011	**0.955**

Definite area of interest for dynamic models

- Definite potential for bias in estimation of unweighted dynamic models.

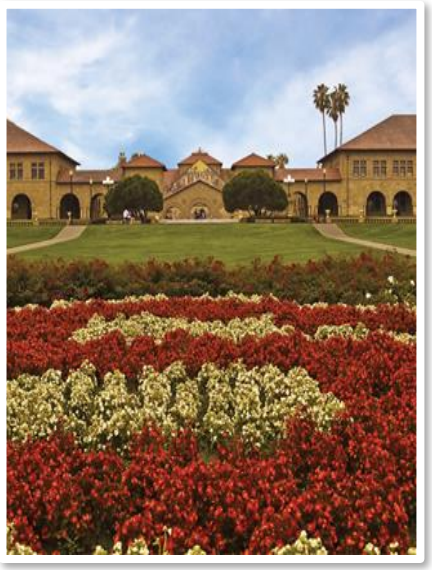
		Ratio (weighted estimate/naïve estimate)
BMI	G	1.001
	t	1.025
	G*t	0.976
Height	G	0.996
	t	1.11
	G*t	0.991
Education	G	1.011
	t	1.082
	G*t	1.187
Smoking	G	0.971
	T	-30.682
	G*t	1.298

Conclusions

- How effectively can we model mortality selection?
  - **Models reduce difference in survival between genotyped and non-genotyped.**
- How might this information change inference about certain types of associations?
  - **Dynamics of polygenic score associations may be underestimate if not corrected for mortality selection.**
- Implications for GWAS.
  - Mortality Selection could bias GWAS results in a manner similar to that of population stratification.

Thanks!

bdomingue@stanford.edu



Collaborators: Dan Belsky, Amal Harrati, Dalton Conley, Jason Boardman and David Weir