

# History of Low Hourly Wage and All-Cause Mortality Among Middle-aged Workers

Kuang Li, Xiaowei Zeng, Mingqian Zhang

UC Davis & Fudan University

December 7, 2023

# Agenda

## 01 Introduction

Background, main objective, previous work.

## 02 Exploratory Analysis

Data preprocessing and description, KM curves.

## 03 Models and Results

Cox model, parametric model, Aalen's additive model.

## 04 Conclusions

Summary, limitations.



# Agenda

## 01 Introduction

Background, main objective, previous work.

## 02 Exploratory Analysis

Data preprocessing and description, KM curves.

## 03 Models and Results

Cox model, parametric model, Aalen's additive model.

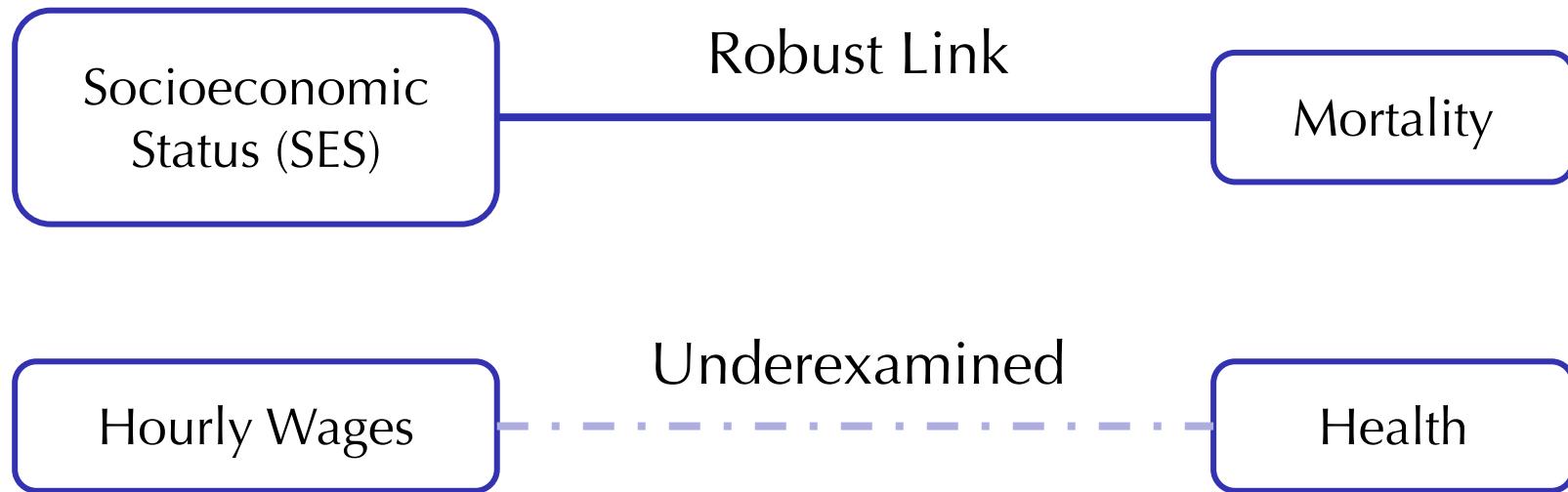
## 04 Conclusions

Summary, limitations.



# Background and Prior Work

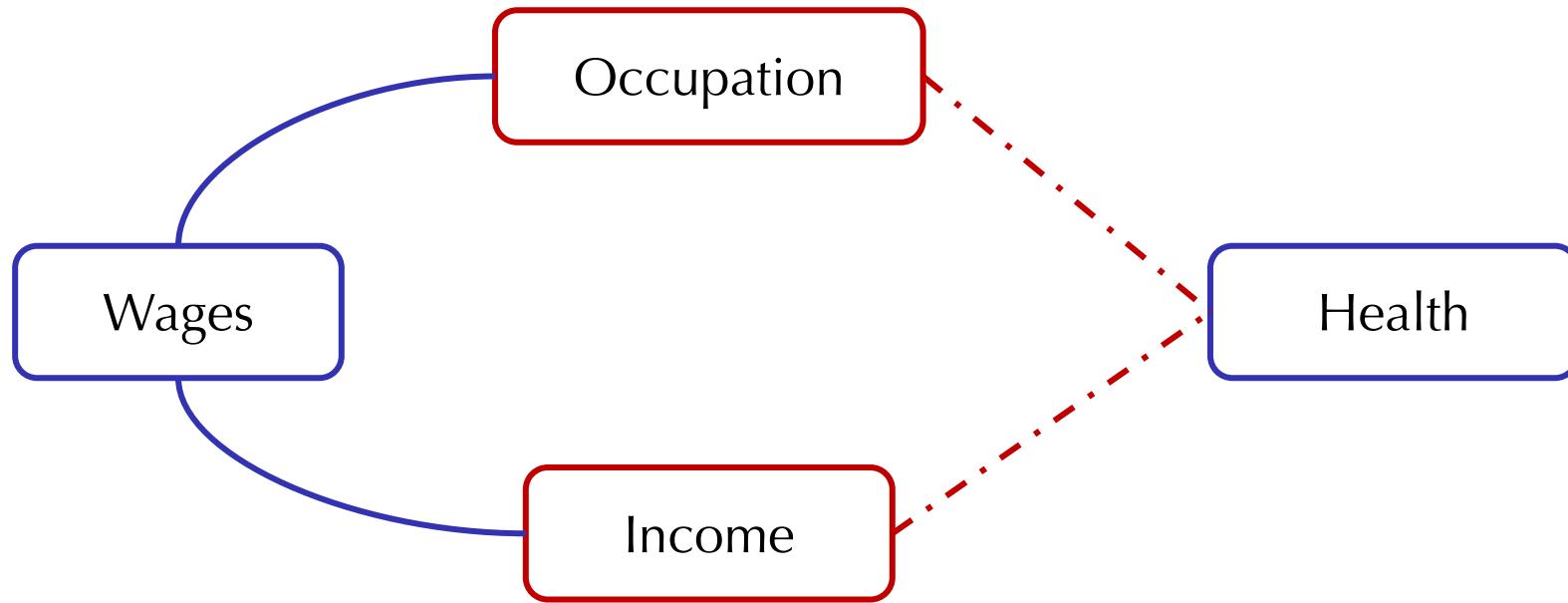
- Earning a low wage is an increasingly recognized public health concern.
- Prior research shows...



- Little research exists on the long-term health consequences of sustained low-wage earning.

# Why wages?

- The importance of wages might be reflected in...



- If causal, policies such as minimum wage laws could directly impact hourly wage, which could potentially improve health and health inequality according to the research.

# Objectives

- Research question:

Whether the **association** of low hourly wage history and all-cause mortality exist among middle-aged (35-50) workers.

- Please note that this study is...

Causal Study



Association Study



- The study is based on data from the Health and Retirement Study (HRS), which collects data every two years.

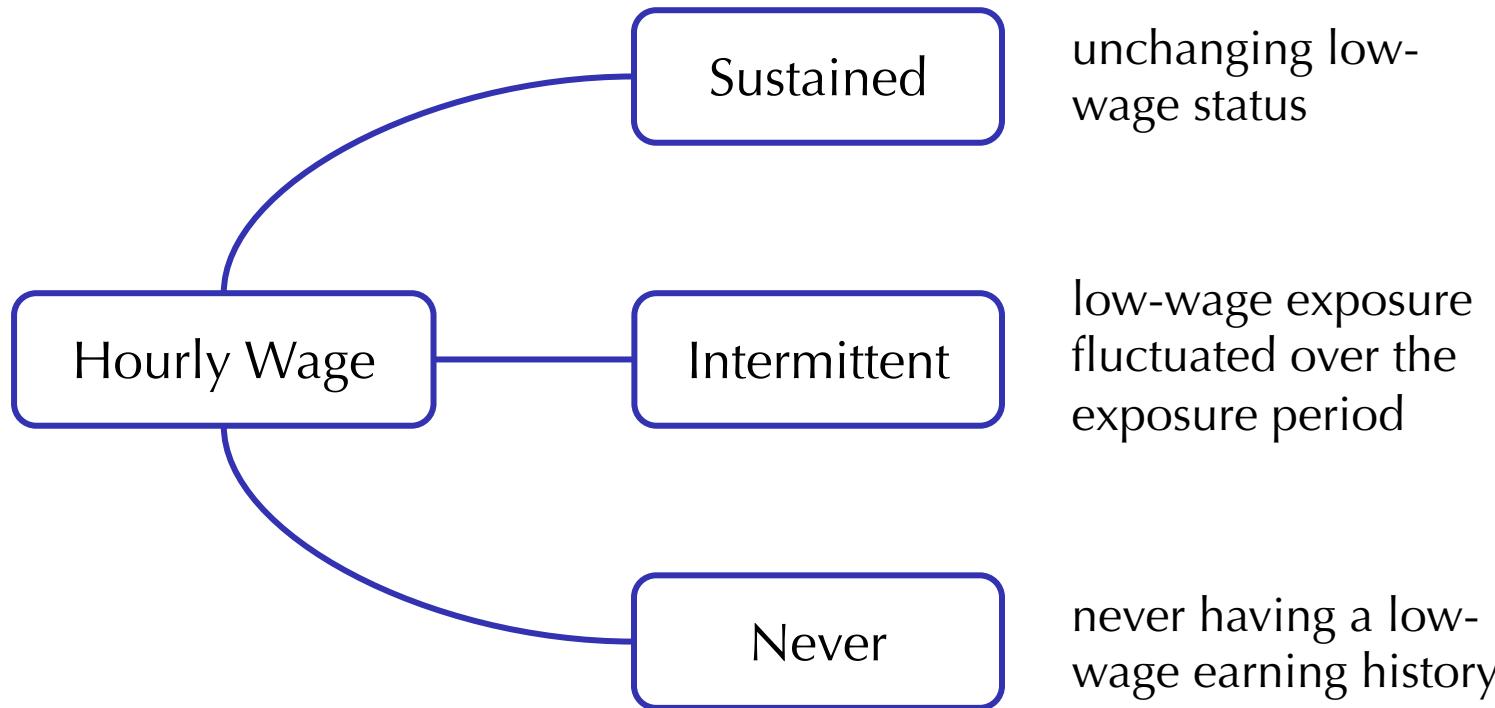


# Definition of Low Wage

- Definition 1
  - Full-time, full-year work (2087 hours)
  - Annual earnings lower than the federal poverty line.
- Definition 2
  - Full-time, full-year work (1750 hours)
  - Annual earnings lower than the federal poverty line.
- Definition 3 (**Used in this study**)
  - An hourly wage lower than 2/3 of the federal median wage for the corresponding year.

# History of Low Hourly Wage

- Categorization:



- We mainly conclude from the **sustained** group rather than the intermittent group.

# Agenda

## 01 Introduction

Background, main objective, previous work.

## 02 Exploratory Analysis

Data preprocessing and description, KM curves.

## 03 Models and Results

Cox model, parametric model, Aalen's additive model.

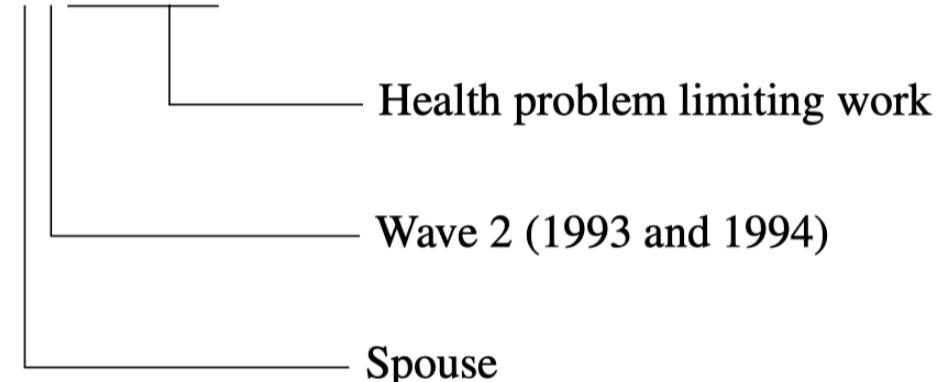
## 04 Conclusions

Summary, limitations.



# Dataset Information

- The Health and Retirement Study (HRS) is a longitudinal household survey conducted by the Institute for Social Research at the University of Michigan.
- The RAND HRS Longitudinal File 2020 (V1) includes 15 waves of Core Interview data (biennially 1992-2020).
  - Wave: The ?th time of data collection.
  - Variable naming pattern: S2 HLTHLM



# Data Cleaning

- Based on RAND HRS data (1992-2020), we ...
- Select 192 variables from 17013 variables in raw data, then encode them into **21** variables.
- Select participants reporting wage  $\geq 6$  waves in 15 waves.
- Select participants aged 35-50 in the beginning of study.
- Finally, **3460** of 42406 participants are chosen.

hhidpn	event	status	wghist	wave_num	prop_pt	employ	age	race	gender	edu	pedu	religion	live	marriage	wealth	insur_gov	insur_com	bmi	self_health	smoke
3010	21.666666666		1 intermittent	7	0 fluctuated		84	nhwhite	1.male	<=12years	>12years	1.protestant	4.west	married	1280149.518	sometimes	sometimes	26.13636363	fair/poor	never
3020	23.833333333		1 sustained	6	0 fluctuated		82	nhwhite	2.female	>12years	>12years	1.protestant	4.west	married	1270951.612	sometimes	never	28.891666666	good/very gc	never
10013040	30	0 never		8	0.2 fluctuated		73	nhwhite	2.female	>12years	>12years	3.jewish	1.northeast	married	828546.4953	sometimes	sometimes	24.766666666	good/very gc	ever
10038010	30	0 never		9	0.06666667 fluctuated		84	nhwhite	1.male	>12years	>12years	3.jewish	1.northeast	married	3808126.074	sometimes	sometimes	23.946666666	good/very gc	never
10038040	30	0 never		12	0.2 fluctuated		77	nhwhite	2.female	>12years	>12years	3.jewish	1.northeast	married	3808126.074	sometimes	sometimes	22.42	good/very gc	ever
10050010	30	0 never		12	0 stable		79	nhother	2.female	>12years	>12years	1.protestant	1.northeast	single	1139876.300	never	always	25.65	good/very gc	never
10059030	30	0 never		11	0.133333333 stable		92	nhwhite	1.male	>12years	4.none/no pr	1.northeast	married	13242771.28	sometimes	sometimes	23.34	good/very gc	never	
10109030	30	0 never		10	0 fluctuated		91	nhwhite	1.male	>12years	<=12years	2.catholic	1.northeast	married	5381677.878	sometimes	sometimes	28.59285714	good/very gc	never
102299010	30	0 never		6	0 fluctuated		80	nhwhite	1.male	>12years	>12years	5.other	5.move	single	96922.80033	never	always	27.15714285	good/very gc	ever
10394010	30	0 never		11	0 stable		80	nhwhite	1.male	>12years	>12years	2.catholic	1.northeast	married	800316.9921	sometimes	always	27.3	good/very gc	ever
10394040	21.75	1 intermittent		9	0.545454545 fluctuated		75	nhwhite	2.female	>12years	>12years	2.catholic	1.northeast	married	800316.9921	sometimes	never	19.92727272	good/very gc	never
10397010	30	0 never		13	0 fluctuated		79	nhwhite	1.male	>12years	>12years	2.catholic	1.northeast	married	1439821.329	sometimes	sometimes	27.5	good/very gc	never
10397040	21.083333333	1 never		7	0.09090909 fluctuated		77	nhwhite	2.female	>12years	<=12years	2.catholic	1.northeast	married	1140506.841	sometimes	sometimes	27.41818181	good/very gc	ever
10451010	30	0 never		7	0 fluctuated		82	nhother	2.female	>12years	>12years	2.catholic	5.move	married	476071.9854	sometimes	sometimes	23.566666666	good/very gc	never
10451020	30	0 never		7	0 fluctuated		80	nhother	1.male	>12years	<=12years	2.catholic	5.move	married	476071.9854	sometimes	sometimes	26.64	good/very gc	ever
10458030	30	0 never		8	0.07692308 fluctuated		76	nhwhite	2.female	>12years	>12years	2.catholic	1.northeast	married	401753.1310	sometimes	sometimes	25.29230769	good/very gc	never
10475010	25.916666666	1 intermittent		6	0.07692308 fluctuated		89	nhwhite	2.female	>12years	>12years	1.protestant	1.northeast	single	1236934.963	sometimes	sometimes	20.93846153	good/very gc	ever
10481010	30	0 never		6	0 fluctuated		80	nhwhite	1.male	>12years	<=12years	2.catholic	5.move	married	364851.8547	sometimes	sometimes	31.946666666	good/very gc	ever
10481020	30	0 intermittent		11	0.4 fluctuated		80	nhwhite	2.female	>12years	>12years	2.catholic	5.move	married	364851.8547	sometimes	sometimes	28.393333333	good/very gc	never
10565020	27.833333333	1 never		7	0 fluctuated		88	nhwhite	1.male	<=12years	<=12years	2.catholic	1.northeast	married	141177.8103	sometimes	sometimes	25.67857142	good/very gc	never
10571010	30	0 intermittent		9	0 fluctuated		80	nhblack	2.female	<=12years	<=12years	1.protestant	1.northeast	married	256155.4428	sometimes	sometimes	30.580000000	fair/poor	never
10577010	30	0 intermittent		0	0.266666666 fluctuated		91	nhblack	2.female	<=12years	<=12years	2.catholic	5.move	single	32461.48050	sometimes	sometimes	20.900000000	good/very gc	ever

# Variable Definitions

## Variable definitions.

Var.	Definition	Var.	Definition
hhidpn	person id	edu	personal education years
event	death time	pedu	maximum parental education years
status	1=death, 0=censored	religion	religion belief
wghist	history of low hourly wage	live	living area
wave_num	wave number	marriage	married or not married
prop_pt	proportion of part-time job	wealth	household wealth/savings
employ	employment stability	insur_gov	whether covered by gov. insurance
age	age in 2020	insur_com	with or without employer's insurance
race	Hispanic, Non-Hispanic	bmi	BMI index
gender	male, female	self_health	self-reported health
		smoke	never or ever smoke

# Data Description

## ● Descriptive statistics by history of low hourly wage:

<b>Characteristic</b>	<b>never, N = 1,695<sup>1</sup></b>	<b>intermittent, N = 1,442<sup>1</sup></b>	<b>sustained, N = 323<sup>1</sup></b>
wave_num	7 (6, 8)	7 (6, 9)	7 (6, 8)
prop_pt	0.00 (0.00, 0.08)	0.00 (0.00, 0.17)	0.13 (0.00, 0.33)
<b>employ</b>			
stable	929 (55%)	610 (42%)	120 (37%)
fluctuated	766 (45%)	832 (58%)	203 (63%)
age	68.0 (64.0, 73.0)	71.0 (67.0, 75.0)	70.0 (66.0, 74.5)
<b>race</b>			
nhwhite	1,321 (78%)	1,015 (70%)	156 (48%)
hispan	126 (7.4%)	170 (12%)	89 (28%)
nhblack	179 (11%)	213 (15%)	66 (20%)
nhother	69 (4.1%)	44 (3.1%)	12 (3.7%)
<b>gender</b>			
female	887 (52%)	929 (64%)	249 (77%)
male	808 (48%)	513 (36%)	74 (23%)

# Data Description

## ● Descriptive statistics by history of low hourly wage:

<b>Characteristic</b>	<b>never, N = 1,695<sup>1</sup></b>	<b>intermittent, N = 1,442<sup>1</sup></b>	<b>sustained, N = 323<sup>1</sup></b>
<b>edu</b>			
<=12years	452 (27%)	632 (44%)	248 (77%)
>12years	1,243 (73%)	810 (56%)	75 (23%)
<b>pedu</b>			
<=12years	376 (22%)	513 (36%)	190 (59%)
>12years	1,319 (78%)	929 (64%)	133 (41%)
<b>religion</b>			
none	221 (13%)	132 (9.2%)	31 (9.6%)
catholic	462 (27%)	388 (27%)	100 (31%)
jewish	46 (2.7%)	15 (1.0%)	4 (1.2%)
other	43 (2.5%)	30 (2.1%)	6 (1.9%)
protestant	923 (54%)	877 (61%)	182 (56%)

# Data Description

## ● Descriptive statistics by history of low hourly wage:

<b>Characteristic</b>	<b>never, N = 1,695<sup>1</sup></b>	<b>intermittent, N = 1,442<sup>1</sup></b>	<b>sustained, N = 323<sup>1</sup></b>
<b>live</b>			
northeast	301 (18%)	159 (11%)	22 (6.8%)
midwest	382 (23%)	347 (24%)	63 (20%)
move	119 (7.0%)	122 (8.5%)	19 (5.9%)
<b>south</b>	<b>495 (29%)</b>	<b>544 (38%)</b>	<b>171 (53%)</b>
west	398 (23%)	270 (19%)	48 (15%)
<b>marriage</b>			
married	1,332 (79%)	1,121 (78%)	228 (71%)
single	363 (21%)	321 (22%)	95 (29%)
wealth	399,921 (162,188, 843,061)	215,019 (87,260, 508,485)	76,668 (28,928, 181,698)
<b>insur_gov</b>			
never	610 (36%)	298 (21%)	61 (19%)
always	31 (1.8%)	28 (1.9%)	13 (4.0%)
sometimes	1,054 (62%)	1,116 (77%)	249 (77%)

# Data Description

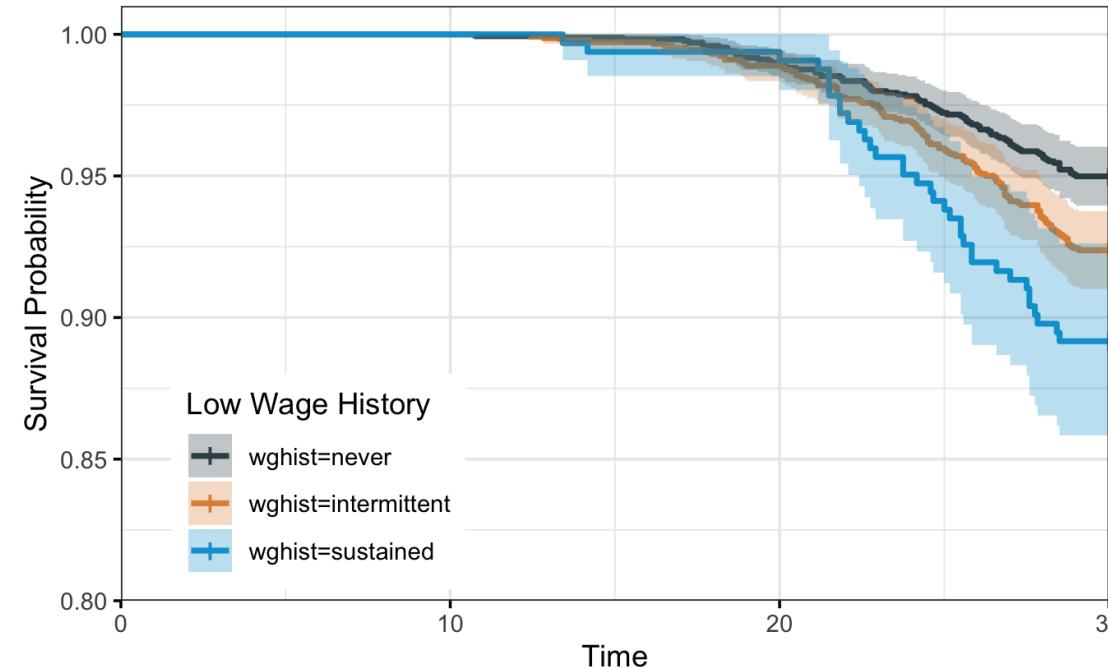
## ● Descriptive statistics by history of low hourly wage:

<b>Characteristic</b>	<b>never, N = 1,695<sup>1</sup></b>	<b>intermittent, N = 1,442<sup>1</sup></b>	<b>sustained, N = 323<sup>1</sup></b>
<b>insur_com</b>			
never	134 (7.9%)	263 (18%)	133 (41%)
always	470 (28%)	107 (7.4%)	10 (3.1%)
sometimes	1,091 (64%)	1,072 (74%)	180 (56%)
<b>bmi</b>			
	27.8 (24.7, 31.7)	28.1 (24.9, 31.9)	28.3 (25.2, 31.8)
<b>self_health</b>			
good/very good/excellent	1,569 (93%)	1,298 (90%)	241 (75%)
fair/poor	126 (7.4%)	144 (10.0%)	82 (25%)
<b>smoke</b>			
never	806 (48%)	635 (44%)	147 (46%)
ever	889 (52%)	807 (56%)	176 (54%)

<sup>1</sup> Median (IQR); n (%)

# KM Curves

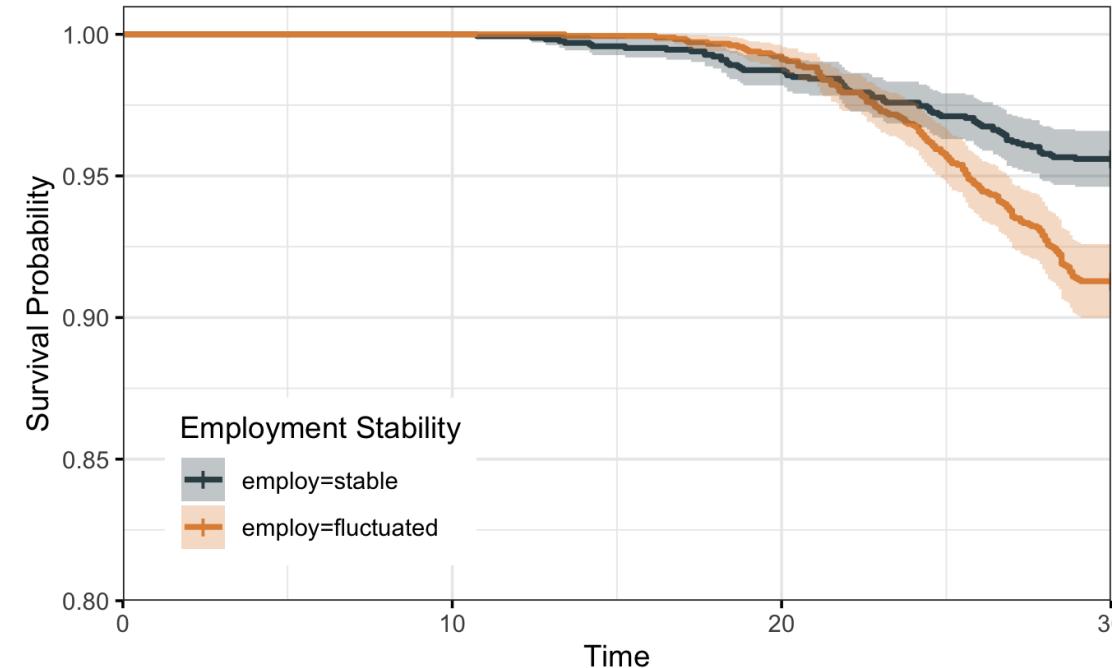
- All-cause mortality ~ history of low hourly wage.
- Interestingly, the sustained level curve slightly crosses the other two curves between 17-22 years.



- Log-rank test: p-value < 0.001\*\*\*.

# KM Curves

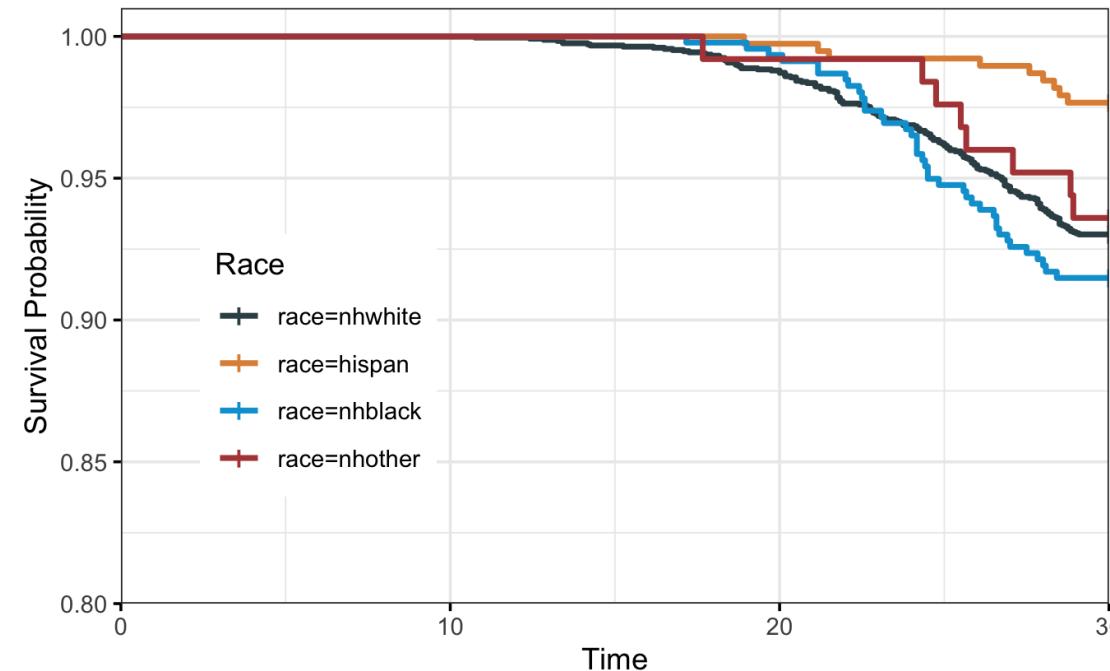
- All-cause mortality ~ employment stability.



- Log-rank test: p-value < 0.001 \*\*\*.

# KM Curves

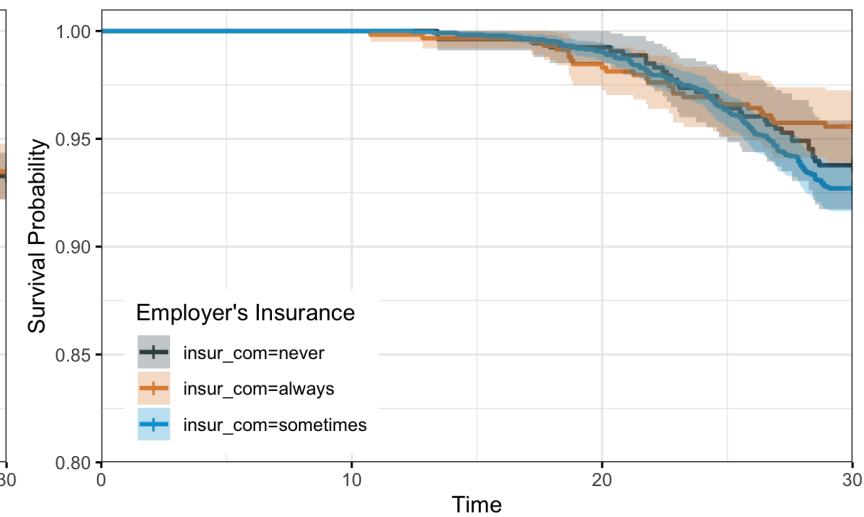
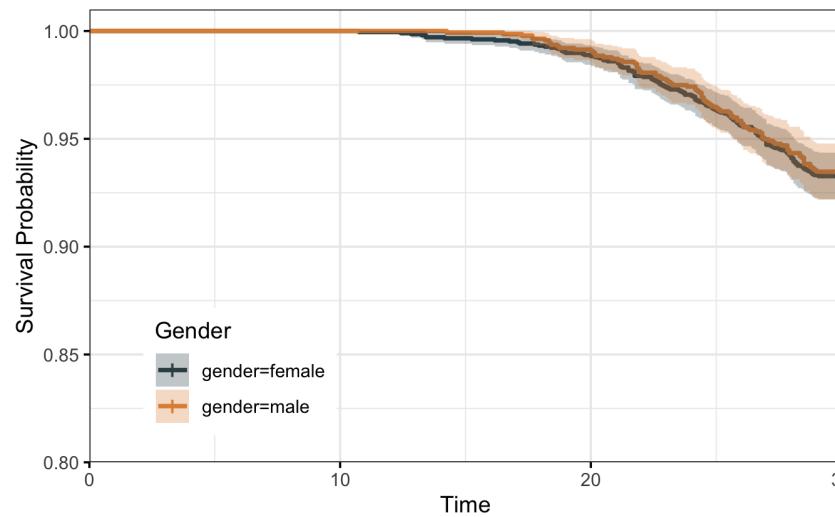
- All-cause mortality ~ race.



- Log-rank test: p-value = 0.002\*\*\*.

# KM Curves

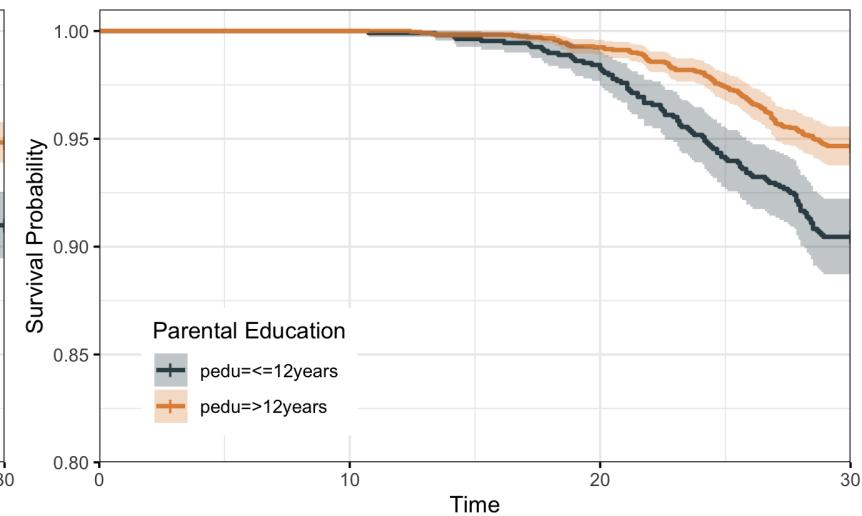
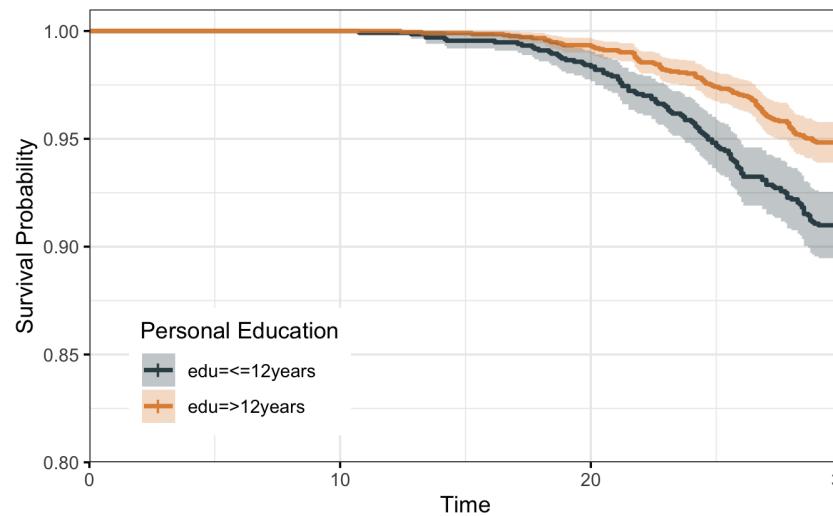
- All-cause mortality ~ gender / employer's insurance.



- Log-rank test: p-value = 0.8, 0.05\*.

# KM Curves

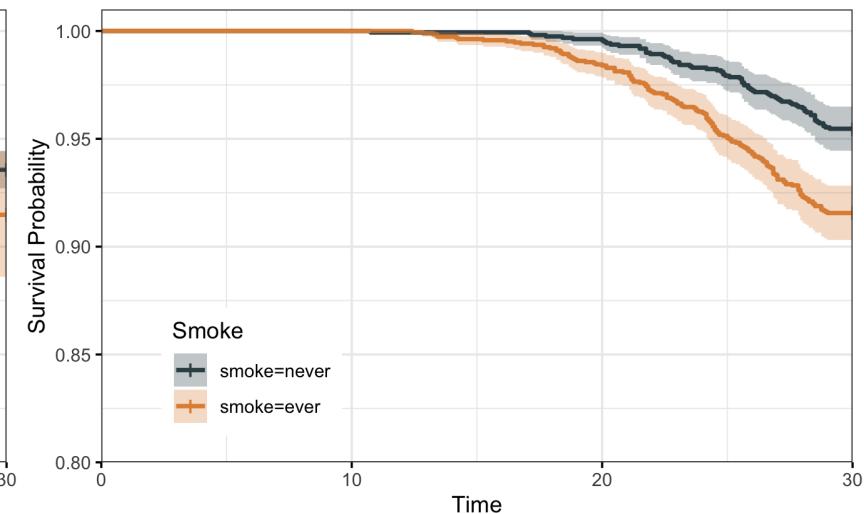
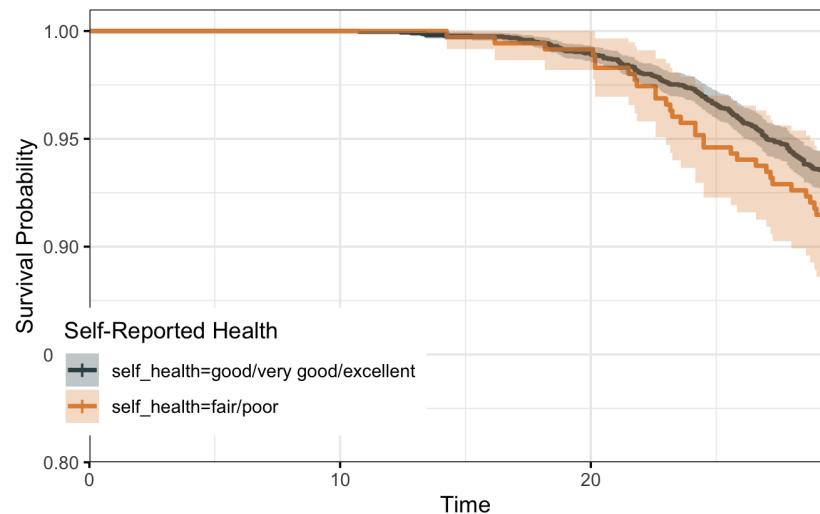
- All-cause mortality ~ education / parental education.



- Log-rank test: p-value < 0.001 \*\*\*.

# KM Curves

- All-cause mortality ~ self-reported health / smoke.



- Log-rank test: p-value = 0.1, <0.001\*\*\*.

# Agenda

## 01 Introduction

Background, main objective, previous work.

## 02 Exploratory Analysis

Data preprocessing and description, KM curves.

## 03 Models and Results

Cox model, parametric model, Aalen's additive model.

## 04 Conclusions

Summary, limitations.



# Outline

- Models are fit to examine the association between low wage and mortality on both **multiplicative** and **additive** scale.

Scale	Multiplicative	Additive
Model	Cox PH model Parametric model	Aalen's additive hazards model
Estimates	hazard ratio	excess deaths
PH	Cox ✓; AFT ✗	✗

- Schoenfeld residuals** are used to assess proportional hazards (PH) assumption. AIC and Lasso are used for variable selection.

## Cox PH Model

# Cox Proportional Hazards Model

# Model Selection: Lasso

- BIC: 3488
- AIC: 3437

**Table 2.** Cox Proportional Hazard Regression using LASSO as feature selection.

Covariate	Estimate	SE	p
wghist: intermittent	1.19 (0.89, 1.60)	0.15	0.24
wghist: sustained	1.78 (1.15, 2.76)	0.22	0.00
age	1.23 (1.19, 1.28)	0.02	0.00
race: hispan	0.32 (0.16, 0.66)	0.36	0.00
race: nhblack	1.12 (0.77, 1.62)	0.19	0.55
race: nother	1.02 (0.49, 2.14)	0.38	0.95
pedu: >12years	0.63 (0.48, 0.84)	0.14	0.00
live: midwest	1.04 (0.67, 1.60)	0.22	0.88
live: move	0.69 (0.36, 1.29)	0.32	0.24
live: south	1.05 (0.70, 1.57)	0.21	0.82
live: west	0.73 (0.43, 1.22)	0.26	0.23
insurance gov: always	0.09 (0.02, 0.37)	0.73	0.00
insurance gov: sometimes	0.20 (0.14, 0.28)	0.18	0.00
self_health: fair/poor	1.63 (1.09, 2.46)	0.21	0.02
smoke: ever	1.73 (1.31, 2.29)	0.14	0.00

# Model Selection: Stepwise

- Compared with Lasso model selection technique, stepwise removes the 'living region' variable.
- BIC: 3471
- AIC: 3434

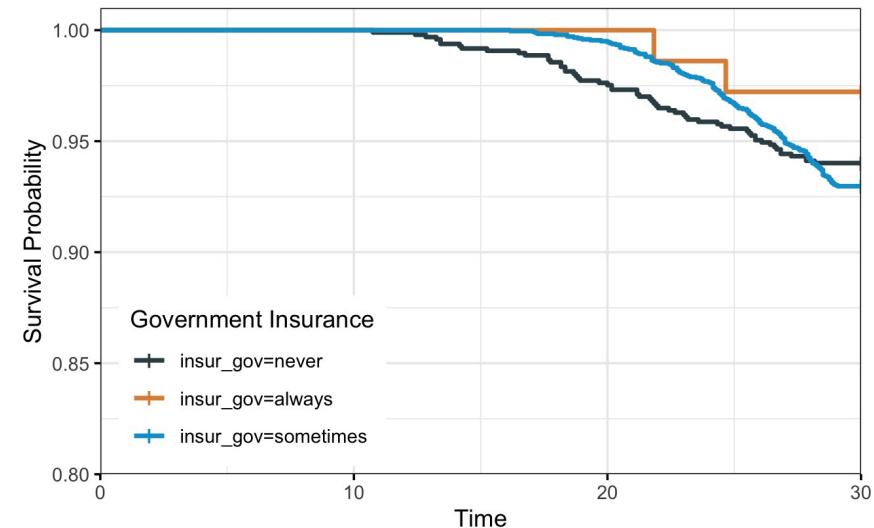
**Table 3.** Cox Proportional Hazard Regression using stepwise as feature selection.

Covariate	Estimate	SE	p
wghist: intermittent	1.21 (0.90, 1.62)	0.15	0.20
wghist: sustained	1.86 (1.21, 2.87)	0.22	0.00
age	1.23 (1.19, 1.28)	0.02	0.00
race: hispan	0.29 (0.14, 0.58)	0.36	0.00
race: nhblack	1.16 (0.80, 1.68)	0.19	0.43
race: nother	0.86 (0.42, 1.76)	0.37	0.68
pedu: >12years	0.62 (0.47, 0.82)	0.14	0.00
insurance gov: always	0.09 (0.02, 0.37)	0.73	0.00
insurance gov: sometimes	0.20 (0.14, 0.27)	0.18	0.00
self_health: fair/poor	1.62 (1.08, 2.43)	0.21	0.02
smoke: ever	1.70 (1.28, 2.24)	0.14	0.00

# Model Selection: Stepwise

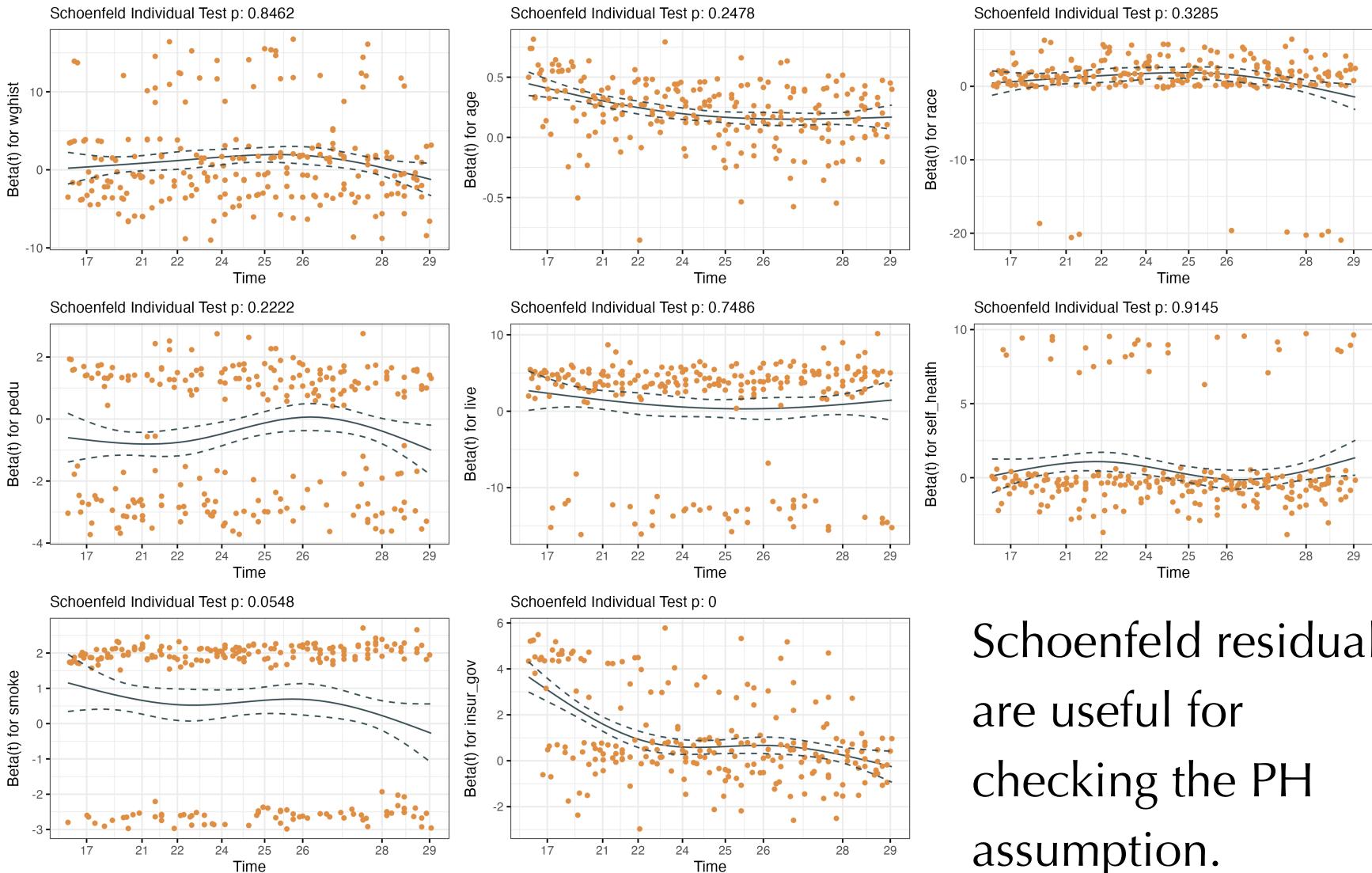
- Test the proportionality of hazards.
- The “insur\_gov” variable does not satisfy PH assumption.

	chisq	df	p
wghist	0.323	2	0.851
age	1.916	1	0.166
race	3.454	3	0.327
pedu	1.325	1	0.250
insur_gov	28.648	2	6.0e-07
self_health	0.039	1	0.843
smoke	3.646	1	0.056
GLOBAL	61.502	11	4.9e-09



- Why we don't choose to use the stratified model?

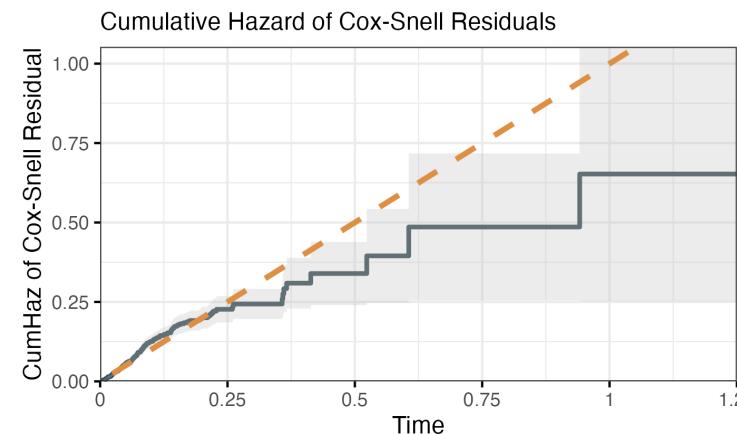
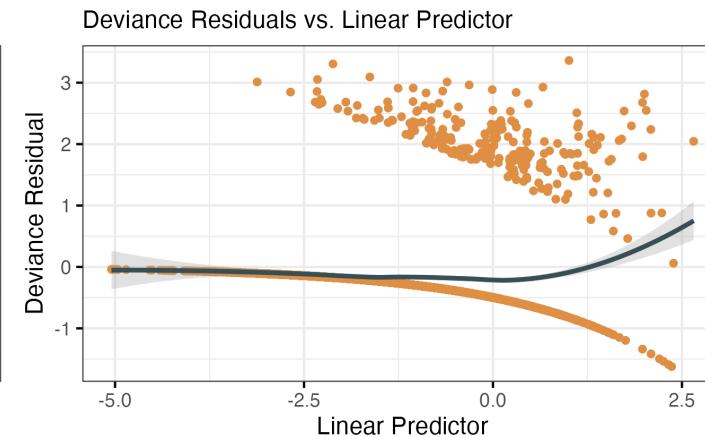
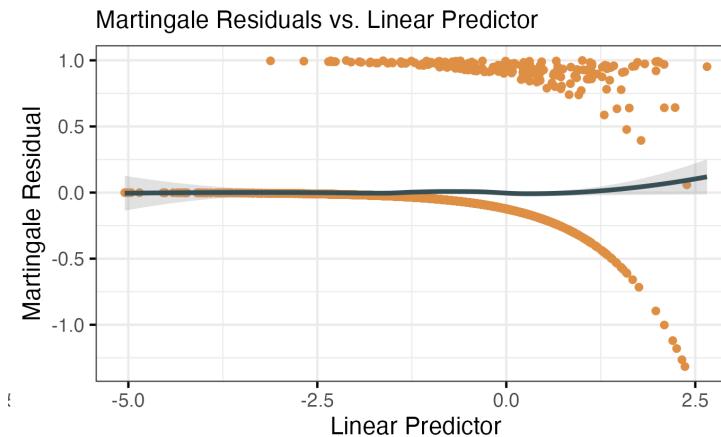
# Schoenfeld Residuals



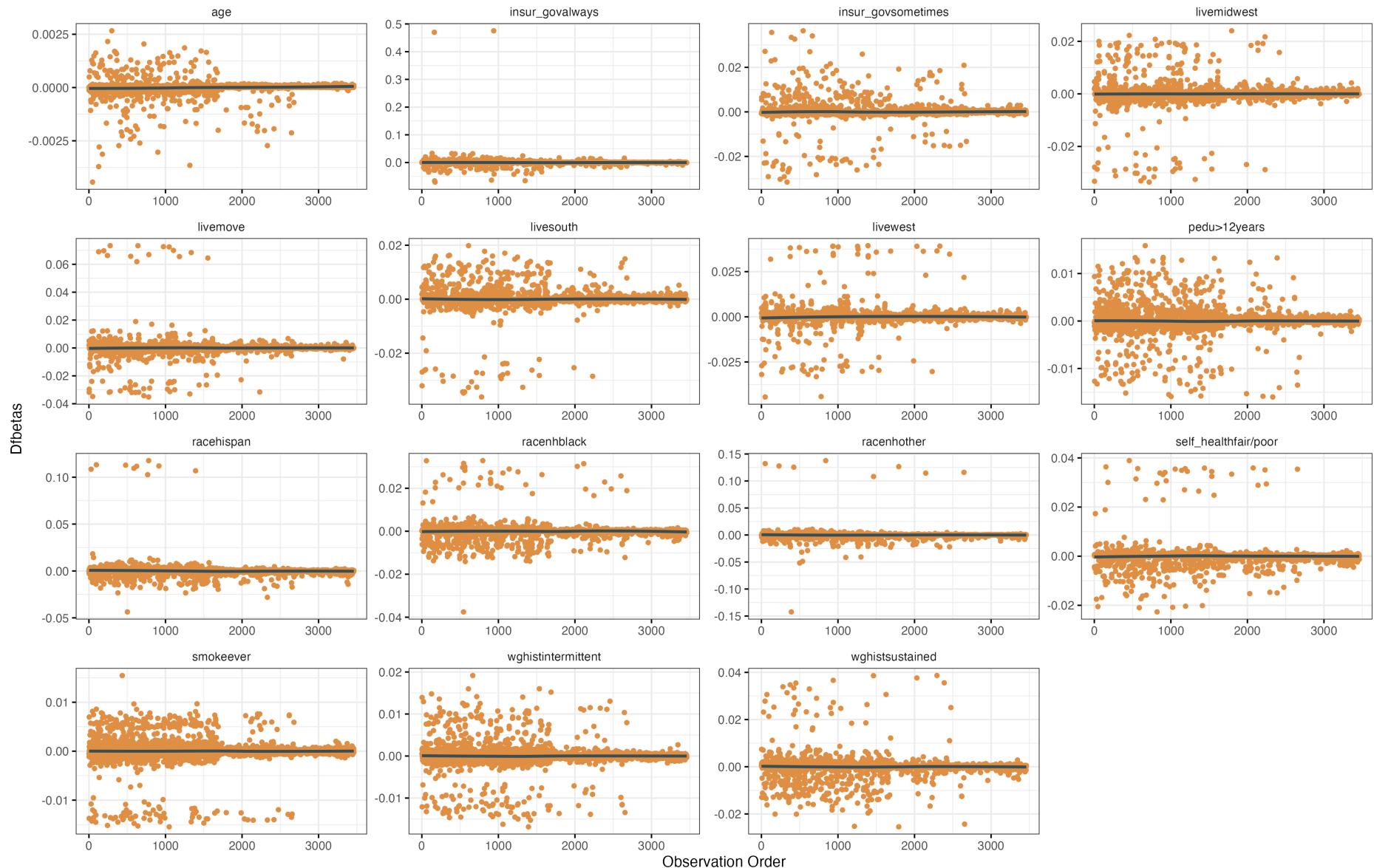
Schoenfeld residuals  
are useful for  
checking the PH  
assumption.

# Martingale Residuals and Deviance Residuals

- Martingale residuals, deviance residuals and dfbetas can help for outlier detection together.



# Dfbetas



# Findings

- If a worker has sustained low wage through the time have an **86.3% higher** hazard compared to workers who never have low wage history.
- Hispanic workers have **72.1% lower** compared to Non-Hispanic Whites.
- If a worker has insurance always covered by the government, it has **91.2% lower** in risk compared to those without coverage.
- If a worker has ever smoke through the time, the risk is **69.5% higher** than those who do not smoke at all.

# Parametric Model

# Parametric Model

# Parametric Model

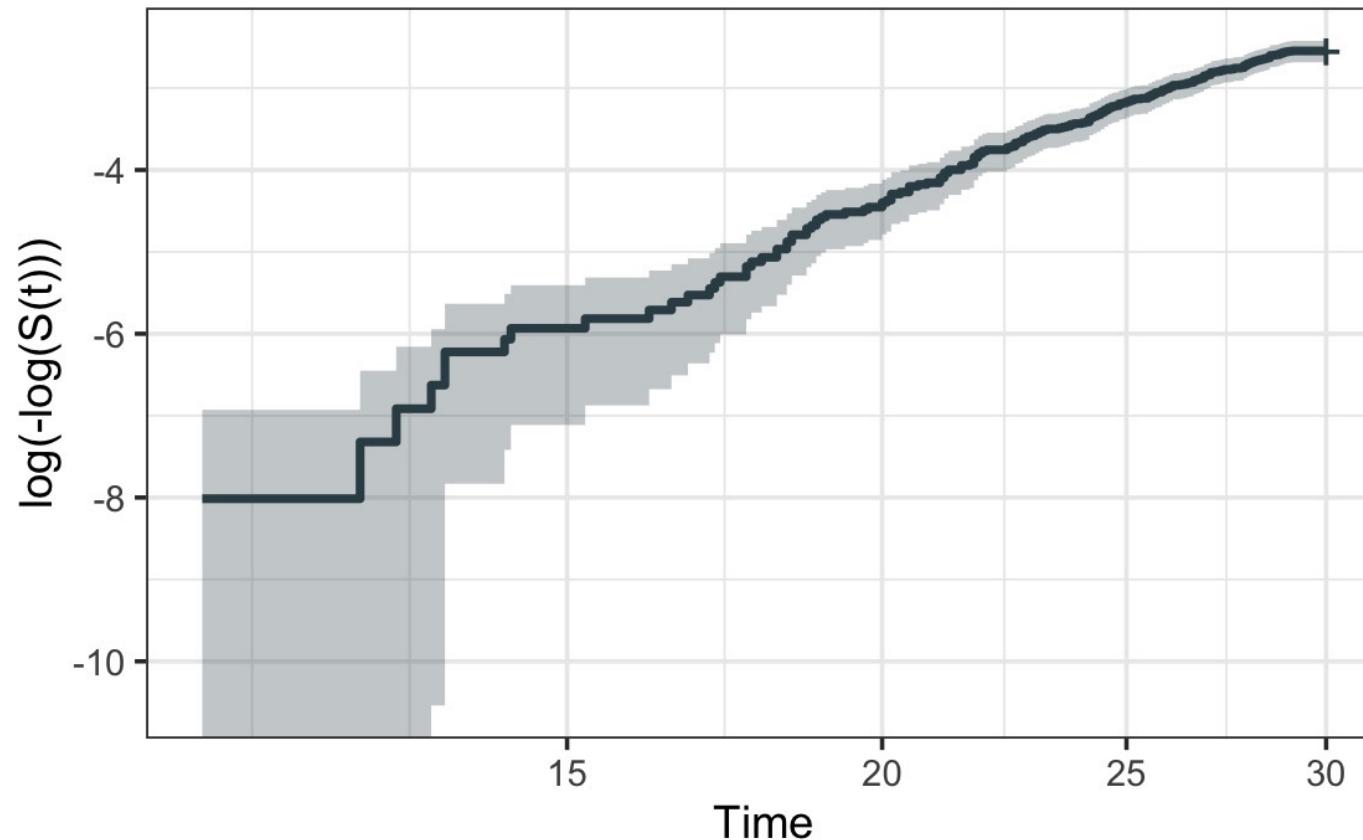
- We first construct four parametric models with the previously selected variables, and get their AICs.

**Table 4.** AIC for Parametric Models

Name	AIC
Weibull	2669.582
Log-logistic	2658.936
Log-normal	2635.860
Exponential	2999.149

# Parametric Model

- Cloglog plot.



# Weibull Distribution

- Model results of Weibull parametric model.

**Table 5.** Results of Weibull regression model.

Covariate	Value	SE	p
wghist: intermittent	-0.044	0.035	0.200
wghist: sustained	-0.145	0.052	0.005
age	-0.048	0.005	0.000
race: hispan	0.289	0.084	0.001
race: nhblack	-0.036	0.043	0.407
race: nother	0.037	0.085	0.665
pedu: >12years	0.111	0.034	0.001
insurance gov: always	0.562	0.172	0.001
insurance gov: sometimes	0.383	0.047	0.000
self_health: fair/poor	-0.112	0.048	0.020
smoke: ever	-0.123	0.034	0.000

# Parametric Model

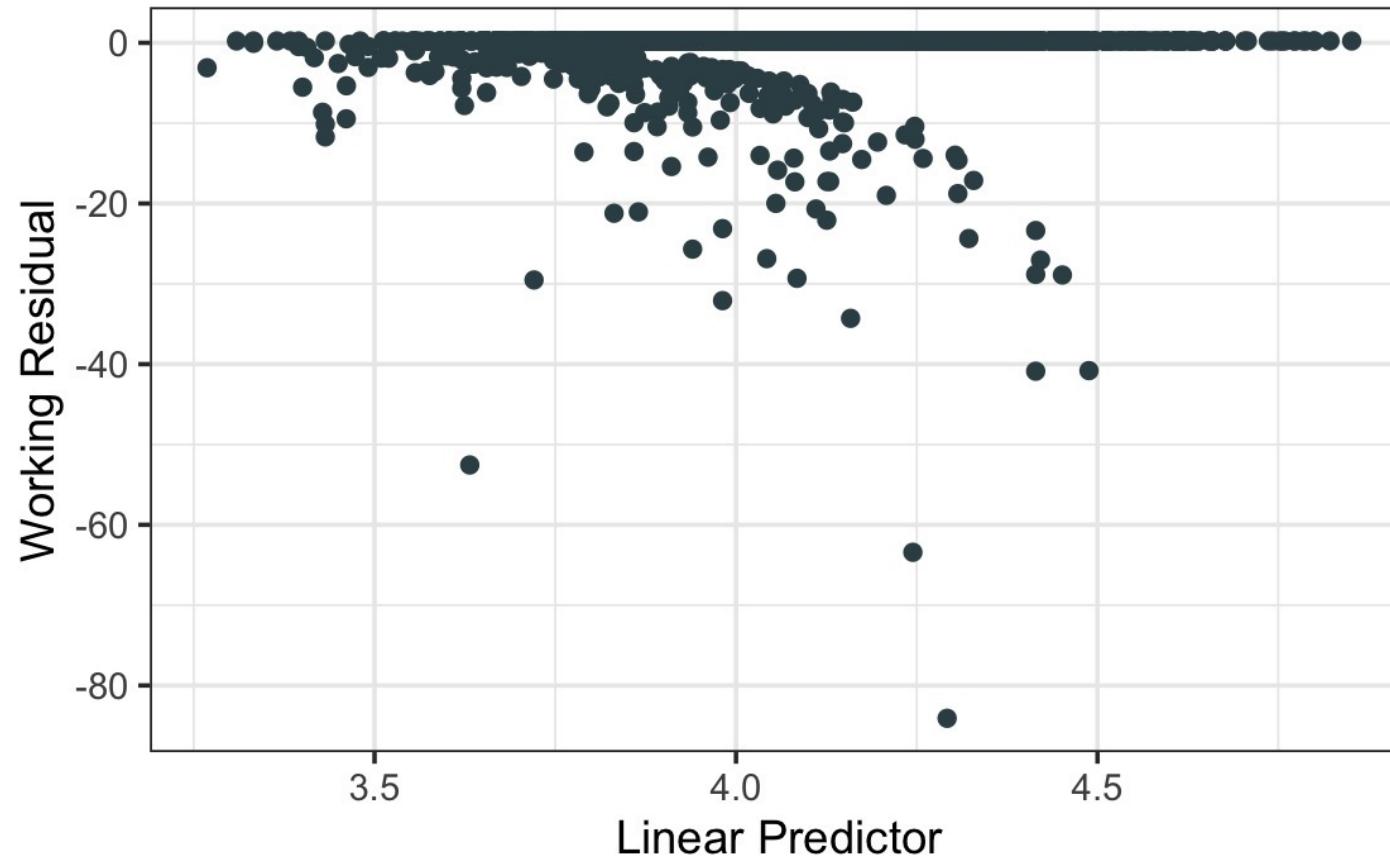
- Convert Weibull.

**Table 6.** Weibull convert results with 95% CI.

Covariate	Hazard ratio	LB	UB
wghist: intermittent	1.21	0.90	1.62
wghist: sustained	1.87	1.21	2.88
age	1.23	1.19	1.28
race: hispan	0.29	0.14	0.57
race: nhblack	1.17	0.81	1.69
race: nhotherr	0.85	0.42	1.75
pedu: >12years	0.62	0.47	0.82
insurance gov: always	0.09	0.02	0.37
insurance gov: sometimes	0.19	0.13	0.27
self_health: fair/poor	1.63	1.08	2.44
smoke: ever	1.70	1.28	2.25

# Residual Analysis for Parametric Models

- Working Residuals vs. Linear Predictor.



# Residual Analysis for Parametric Models

- Dig into Outliers.

Time	Wghist	Self health	Pedu	Insur_Gov
19	never	excellent	>12years	sometimes
11	never	excellent	>12years	never

# Findings

- If a worker has sustained low wage through the time have an **86.8% higher** hazard compared to workers who never have low wage history.
- Hispanic workers have **71.4% lower** compared to Non-Hispanic Whites.
- If a worker has insurance always covered by the government, it has **91.2% lower** in risk compared to those without coverage.
- If the maximum education level of a worker's parents is more than 12 years, the risk is **38.2% lower** than those with less or equal to 12 years parental education level.

# Aalen Additive Hazards Model

# Aalen Additive Hazards Model

# Aalen Additive Hazards Model

- Previous Cox PH model assumes...
  - The effects of the covariates are to act **multiplicatively** on an unknown baseline hazard function.
  - The risk coefficients are unknown constants whose value **do not change over time**.
- Aalen Additive Hazards Model assumes...
  - The covariates act in an **additive** manner.
  - The unknown risk coefficients are **functions of time** so that the effect of a covariate may **vary over time**.
  - $\lambda(t) = Y(t)(X^T(t)\beta(t) + Z^T(t)\gamma)$

# Additive Model 1

- We first construct a nonparametric model with the previously selected variables, assuming that they have time-varying effect.

	Test for Significance	Test for time-variant effects	
(Intercept)	9.06	0***	0***
wghist: intermittent	1.73	0.61	0.31
wghist: sustained	3.23	0.01**	0***
age	9.58	0***	0***
race: hispan	5.30	0***	0***
race: nhblack	2.92	0.06	0.14
race: nother	3.12	0.04*	0.30
pedu: >12years	3.84	0.01**	0.07
insur_gov: always	5.72	0***	0***
insur_gov: sometimes	6.80	0***	0***
self_health: fair/poor	2.29	0.22	0.23
smoke: ever	4.84	0***	0***

# Additive Model 1

- The race, parental education and self-reported health variables show no significant time-varying effects, so we can **simplify** the model by reducing the number of nonparametric component, say by wrapping both variables with *const* in the function *aalen*.

```
aalen(surv ~ wghist + age + const(race) + const(pedu) + insur_gov +
      const(self_health) + smoke, data=data, n.sim=300, start.time=0)
```

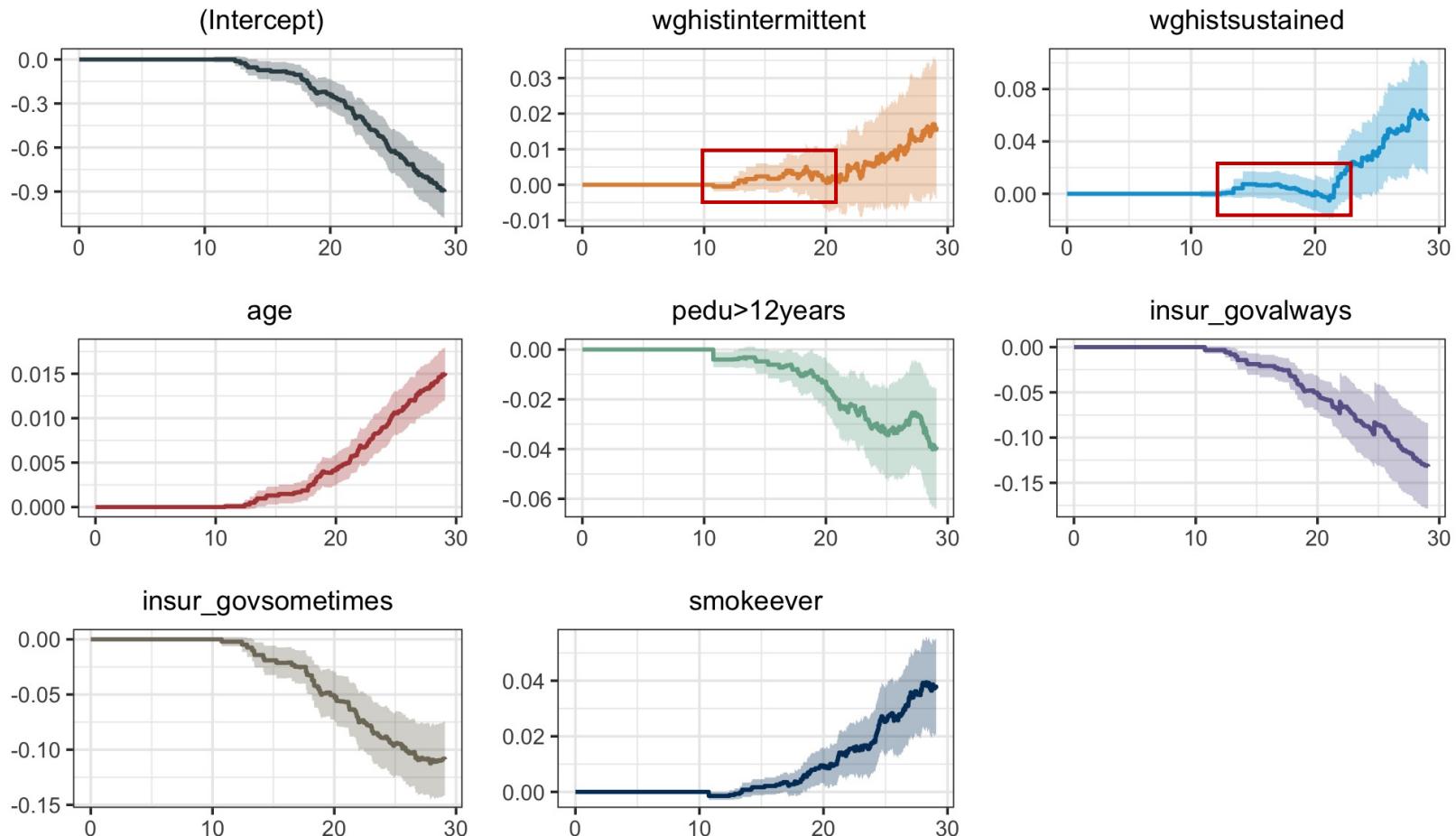
# Additive Model 2

- We then construct a semiparametric Aalen additive model, where effects of race, parental education and self-reported health are assumed to be constant and the remaining covariate effects are allowed to be time varying.

	Excess Death (per 10000 person-years)	p-value
race: hispanic	-19.9 (-27.9, -11.9)	<0.001***
race: nhblack	5.5 (-4.6, 15.6)	0.33
race: nother	-0.5 (-15.9, 14.9)	0.96
pedu: <=12	-13.0 (-20.9, -5.1)	0.002***
health: fair/poor	11.0 (-0.7, 22.7)	0.07

# Additive Model 2

- Estimated cumulative regression coefficients for variables with time-varying effects.



# Additive Model 3

- We adjust our assumption of time-variant effect of history of low hourly wage to a constant effect and then construct another semiparametric Aalen additive model.

	Excess Death (per 10000 person-years)	p-value
wghist: intermittent	5.3 (-1.0, 11.6)	0.11
wghist: sustained	18.6 (5.0, 32.2)	0.005**
race: hispanic	-19.8 (-27.8, -11.8)	<0.001***
race: nhblack	5.5 (-4.6, 15.7)	0.33
race: nother	-0.4 (-15.8, 15.0)	0.96
pedu: <=12	-13.0 (-20.9, -5.1)	0.002***
health: fair/poor	11.1 (-0.6, 22.8)	0.06

# Findings

- Workers with sustained low hourly wage have **0.19% higher** estimated excess death rate than those always with high wage, which means they experienced **18.6 more deaths** per 10000 person-years.
- The history of low hourly wage has less impact on death risks in the first 20 years, but then increases sharply.
- Workers with intermittent low hourly wage do not perform significant difference from those always with high wage.
- Compared to non-Hispanic Whites, Hispanic workers are associated with **19.9 less** deaths per 10000 person-years.

# Agenda

## 01 Introduction

Background, main objective, previous work.

## 02 Exploratory Analysis

Data preprocessing and description, KM curves.

## 03 Models and Results

Cox model, parametric model, Aalen's additive model.

## 04 Conclusions

Summary, limitations.



# Conclusions

- **Sustained** low-wage earning is significantly associated with elevated mortality risk and excess deaths.
- If causal, our findings can suggest that social and economic policies that improve the financial standing of low-wage workers (eg, minimum wage laws) could improve mortality outcomes.

# Limitations

- Selection bias.
- Omitted Variable Bias / Confounders.

# THANKS!

Kuang Li, Xiaowei Zeng, Mingqian Zhang

UC Davis & Fudan University

December 7, 2023