

Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Social and Behavioral Determinants of Health

An Empirical Application: Kaiser Permanente

Janelle Downing

Center for Health and Community
University of California, San Francisco

janelle.downing@ucsf.edu

June 30, 2017

Study Population

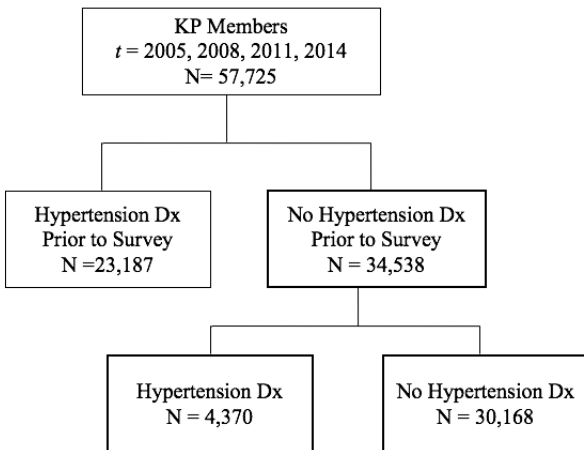
Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition



Research Question

Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition

- Does inclusion of social and behavioral variables improve prediction of hypertension in a clinical population without hypertension at baseline?
- Does a single, composite measure of social and behavioral variables perform better than inclusion of all variables?

Definitions of SBD variables

Methods

Descriptives

Logistic Regression

Survival Analysis

Attrition

- 1) **High School Graduate or Less** was defined as having attained a highest education of 12th grade or less;
- 2) **Financial Strain** was defined as worrying a great deal about one's family's financial security; 3) **Stressed** was defined as being stressed, tense, or anxious most or much of the time during the past 12 months; 4) **Depressed** was defined as being depressed for at least 2 weeks in the past year or taking an antidepressant medicine; 5) **No exercise** was defined as exercising once a month or less; 6) **Current smoker** was identified by answering smoking cigarettes now, even occasionally, in the affirmative; 7) **Heavy Drinker** was identified as consuming 3 or more drinks per day for females and 4 or more drinks per day for males on drinking days; 8) **Partnered** was identified as being married or in a committed relationship; 9) **Intimate Partner Violence** was identified as fearing for the safety of yourself, your family, or friends because of anger or threats of a current or former spouse, partner, or boyfriend/girlfriend in the last 12 months; 10) **Concentrated Neighborhood Poverty** is defined as residing in a block group where 30 percent or more of the households are at or below the federal poverty line.

Distribution of SBD Score

Methods

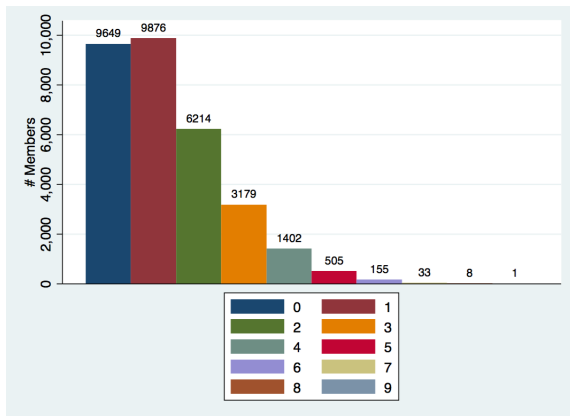
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Only members with no missing SBD variables, n= 31,022



Distribution of SBD Score

Methods

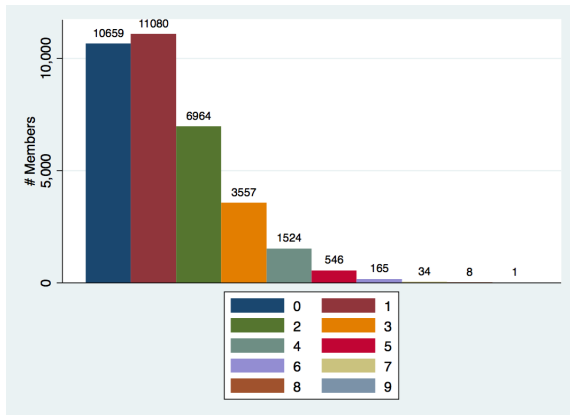
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

All members



Approach

Methods

Descriptives

Logistic Regression

Survival Analysis

Attrition

1 Logistic regression (ignoring censoring)

- 1 Test effect modification by obesity, sex, and age
- 2 Compare predictive ability of models with area under the ROC curve

2 Time-to-Event / Survival Analysis

- 1 Compare predictive ability of survival models with Harrells C and Somers D Concordance

3 Address attrition

- 1 Is attrition random?
- 2 Inverse probability weighting

Models to Predict Hypertension

Methods

Descriptives

Logistic Regression

Survival Analysis

Attrition

$$\Pr(Y = 1 \mid X) = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)} \quad (1)$$

$$\Pr(Y = 1 \mid \text{SBD}, X) = \frac{\exp(\beta_0 + \beta_1 \text{SBD} + \beta_2 X)}{1 + \exp(\beta_0 + \beta_1 \text{SBD} + \beta_2 X)} \quad (2)$$

$$\Pr(Y = 1 \mid \text{Score}, X) = \frac{\exp(\beta_0 + \beta_1 \text{Score} + \beta_2 X)}{1 + \exp(\beta_0 + \beta_1 \text{Score} + \beta_2 X)} \quad (3)$$

Note: All models include robust standard errors clustered at the facility-level. Ignores attrition and time to diagnosis.

- Y is diagnosis of hypertension
- X are traditional predictors: female, age, bmi, race/ethnicity, survey year
- SBD are social and behavioral variables: education, financial strain, stress, depression, exercise, smoking, heavy drinking, partnered, ipv, residence in poor neighborhood
- $Score$ is a sum of the binary SBD variables (0-10)

The Area Under an ROC Curve

Methods

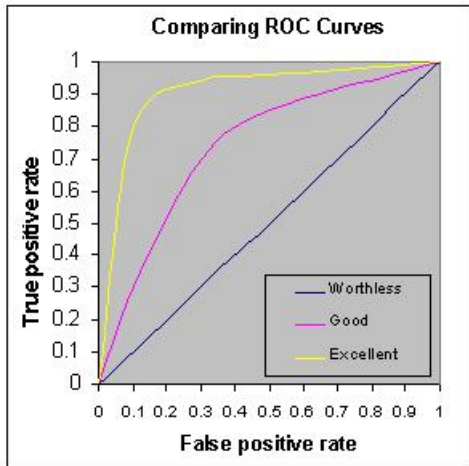
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Measures accuracy of model



Time-to-Event / Survival Analysis

Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition

- Kaplan-Meier (survival function): the probability of being disease-free at least to time t .
- Cox (hazard function): conditional probability of getting the disease at time t having survived to that time.
- Harrell's C / Somers D: measures of concordance that approximate area under the ROC

Cox: A Review

Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition

- Like a multiple regression model
- Enables the difference between survival times of particular groups of patients to be tested while allowing for other factors.
- Response (dependent) variable is the 'hazard'. Probability of getting hypertension given that members have been disease-free up to that point.

Characteristics by Hypertension Dx

Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Table: Differences in Member Characteristics

	No Hypertension		Hypertension		Difference	
	mean	sd	mean	sd	b	t
Female	0.59	0.49	0.51	0.50	0.08***	(9.37)
Age	47.54	16.27	60.77	14.43	-13.23***	(-55.71)
Body Mass Index	26.80	5.32	29.12	6.14	-2.32***	(-22.45)
Non-Hispanic black	0.05	0.22	0.06	0.25	-0.01***	(-3.65)
Less than Bachelors Degree	0.50	0.50	0.63	0.48	-0.13***	(-16.55)
Financial Strain	0.27	0.44	0.22	0.42	0.04***	(5.93)
High Stress	0.17	0.38	0.13	0.34	0.04***	(7.20)
Depressed	0.12	0.32	0.12	0.32	0.00	(0.39)
No Exercise	0.14	0.34	0.16	0.37	-0.03***	(-4.72)
Current Smoker	0.08	0.26	0.09	0.29	-0.01**	(-3.07)
Heavy Drinker	0.07	0.26	0.07	0.26	-0.00	(-0.06)
Partnered	0.24	0.43	0.28	0.45	-0.04***	(-4.93)
Intimate Partner Violence	0.02	0.14	0.02	0.14	0.00	(0.00)
High Poverty Block-Group	0.06	0.23	0.07	0.25	-0.01**	(-2.88)
SBD Index	1.32	1.28	1.41	1.32	-0.09***	(-4.21)
Observations	30168		4370		34538	

Odds of Hypertension: Naive

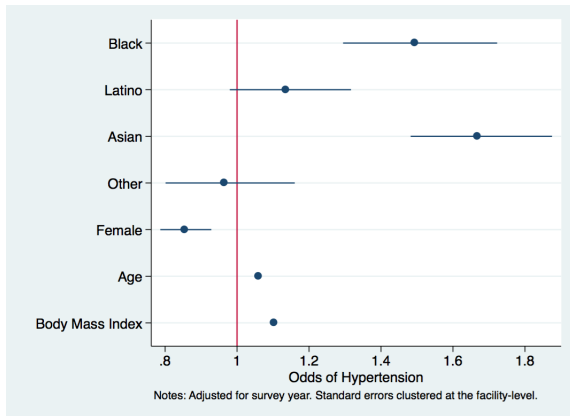
Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition



Odds of Hypertension: SBD

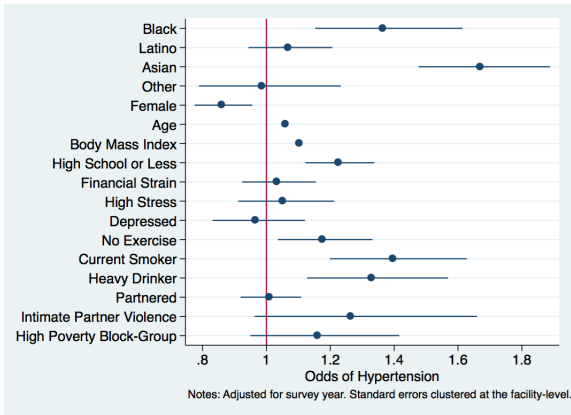
Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition



Odds of Hypertension: SBD Index

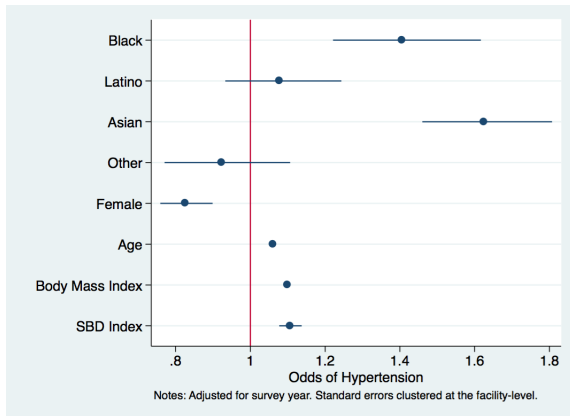
Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition



Odds of Hypertension: SBD Index

Methods

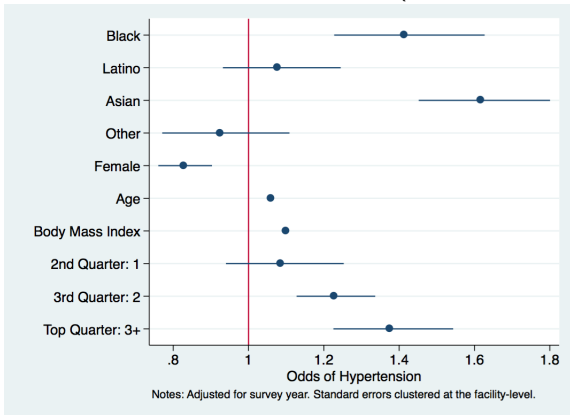
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

*SBD index is categorical, into quarters (ref is bottom quarter)



Odds of Hypertension

Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition

	(1) Naive		(2) SBD Variables		(3) Score	
Hypertension Dx						
White	1	(.)	1	(.)	1	(.)
Black	1.493***	(0.000)	1.400***	(0.000)	1.406***	(0.000)
Latino	1.136*	(0.043)	1.090	(0.213)	1.117	(0.105)
Asian	1.667***	(0.000)	1.730***	(0.000)	1.720***	(0.000)
Other	0.964	(0.753)	1.000	(0.997)	1.003	(0.980)
Female	0.855***	(0.000)	0.852***	(0.000)	0.821***	(0.000)
Age	1.060***	(0.000)	1.061***	(0.000)	1.061***	(0.000)
Body Mass Index	1.103***	(0.000)	1.103***	(0.000)	1.101***	(0.000)
High School or Less			1.233***	(0.000)		
Financial Strain			1.060	(0.238)		
High Stress			1.070	(0.277)		
Depressed			0.965	(0.571)		
No Exercise			1.128*	(0.030)		
Current Smoker			1.353***	(0.000)		
Heavy Drinker			1.356***	(0.000)		
Partnered			1.012	(0.792)		
Intimate Partner Violence			1.196	(0.206)		
High Poverty Block-Group			1.121	(0.152)		
SBD Index					1.112***	(0.000)
Observations	30985		27929		27929	

Exponentiated coefficients; p -values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Area Under the ROC

Methods

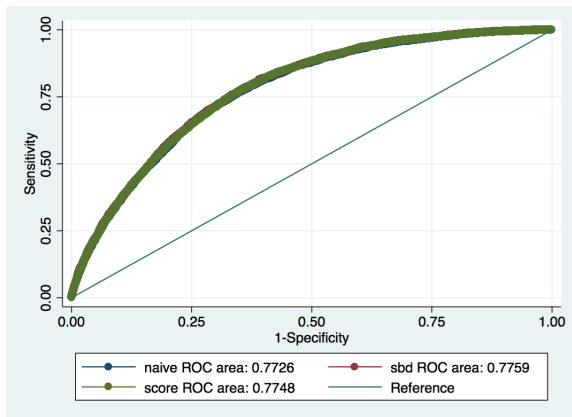
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Test for equality (DeLong 1988) $p < 0.01$



Effect Modification

Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Tested using fully interacted model with each hypothesized interaction term

- No effect modification by obesity
- Effect modification by sex and age (all interaction terms less than 0.08)

Age Categories by Diagnosis

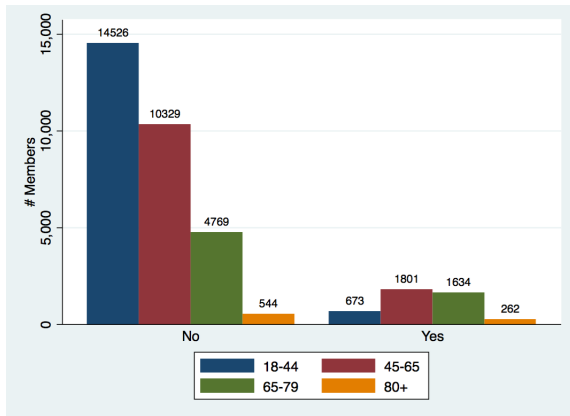
Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition



Odds of Hypertension: Naive

Methods

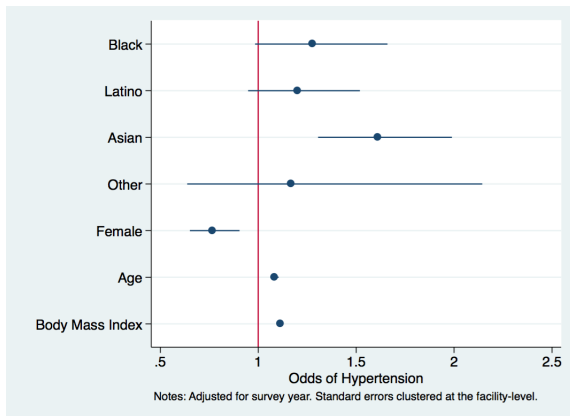
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Members 18-44 only



Odds of Hypertension: SBD

Methods

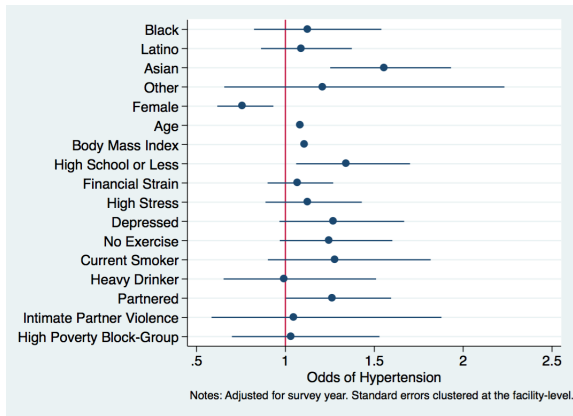
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Members 18-44 only



Odds of Hypertension: SBD Index

Methods

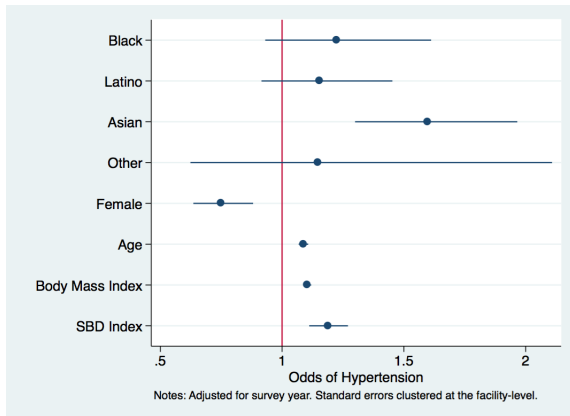
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Members 18-44 only



Area Under the ROC

Methods

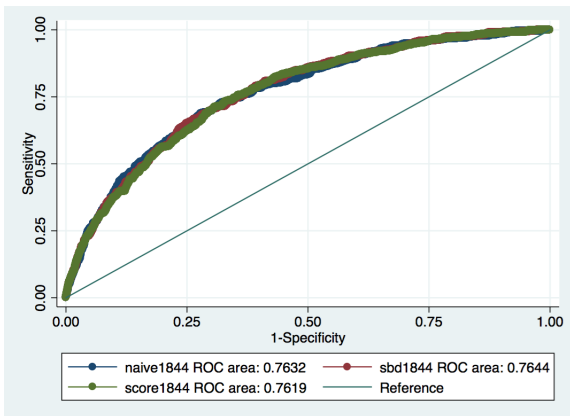
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Test for equality (DeLong 1988) $p=0.1894$



Odds of Hypertension: Naive

Methods

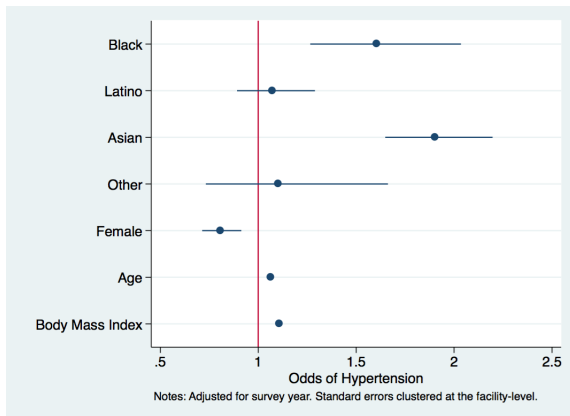
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Members 45-64 only



Odds of Hypertension: SBD

Methods

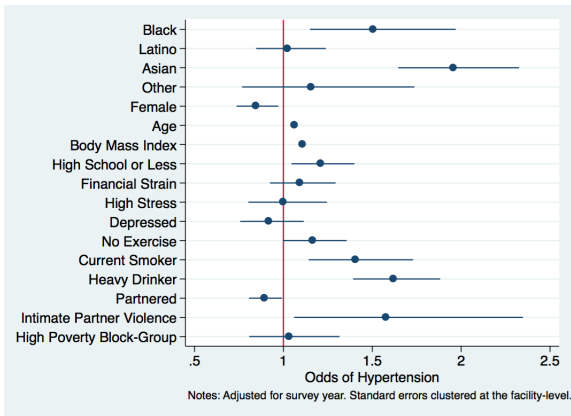
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Members 45-64 only



Odds of Hypertension: SBD Index

Methods

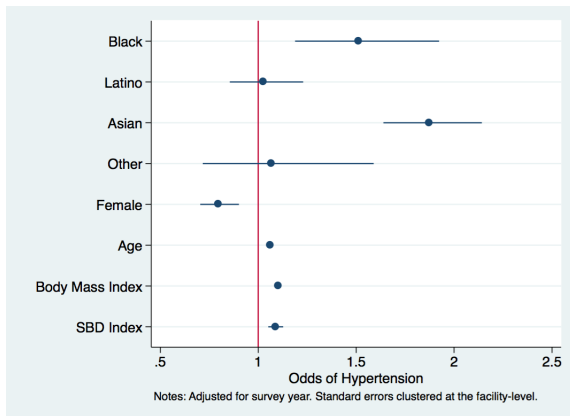
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Members 45-64 only



Area Under the ROC

Methods

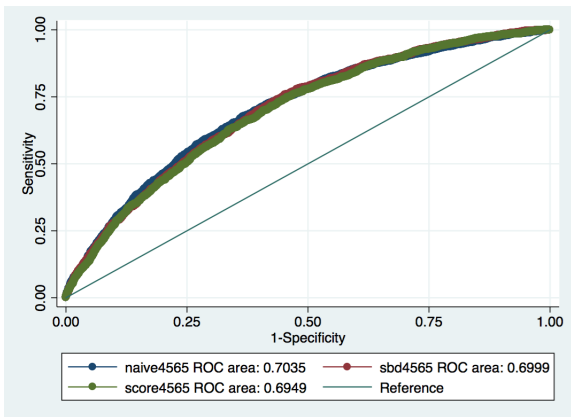
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Test for equality (DeLong 1988) $p < 0.01$



Odds of Hypertension: Naive

Methods

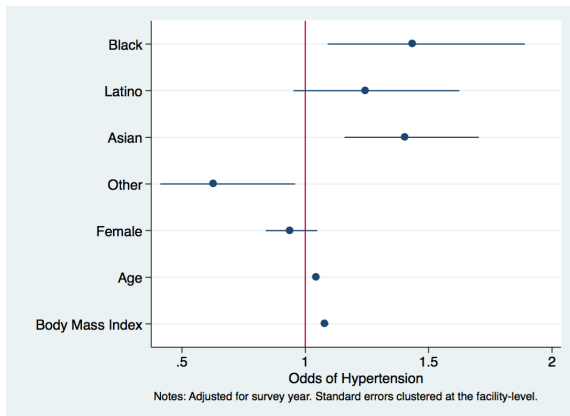
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Members 65-79 only



Odds of Hypertension: SBD

Methods

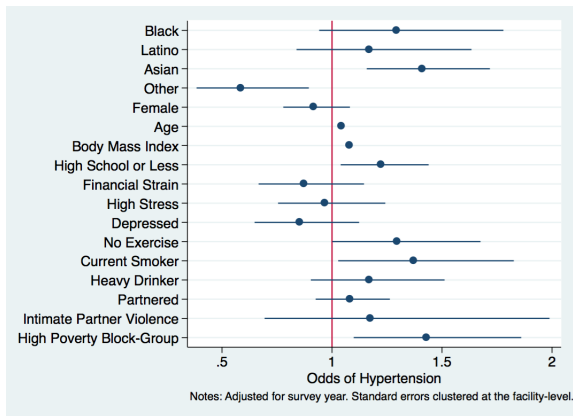
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Members 65-79 only



Odds of Hypertension: SBD Index

Methods

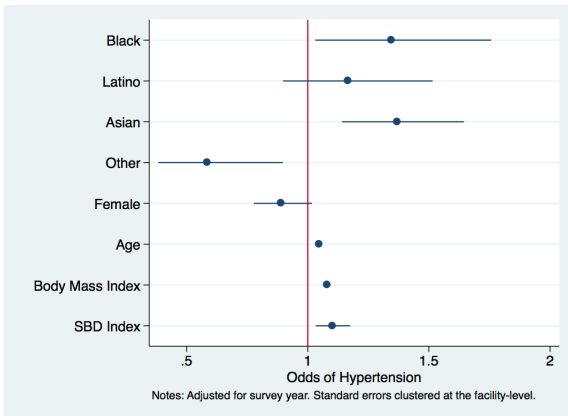
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Members 65-79 only



Area Under the ROC

Methods

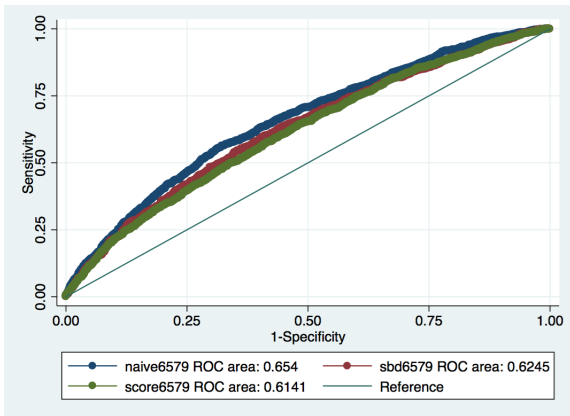
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Test for equality (DeLong 1988) $p < 0.01$



Odds of Hypertension: Naive

Methods

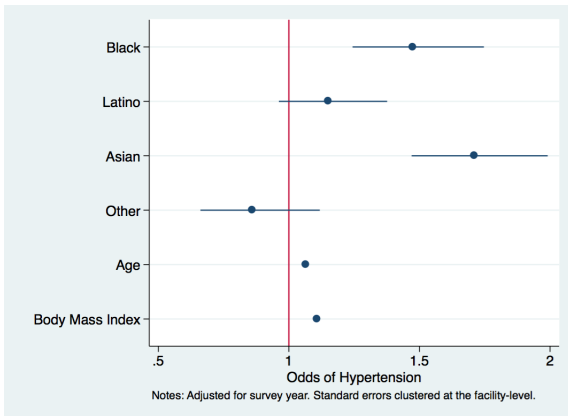
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Females only



Odds of Hypertension: SBD

Methods

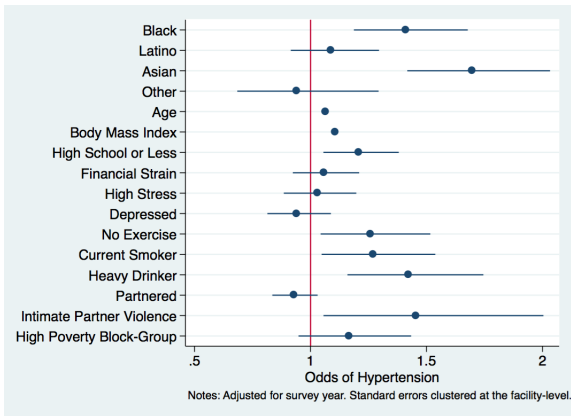
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Females only



Odds of Hypertension: SBD Index

Methods

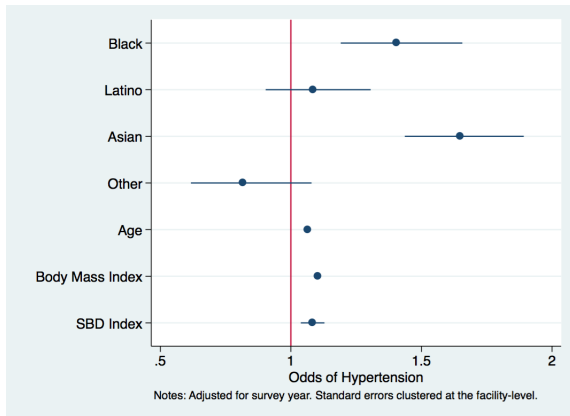
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Females only



Area Under the ROC

Methods

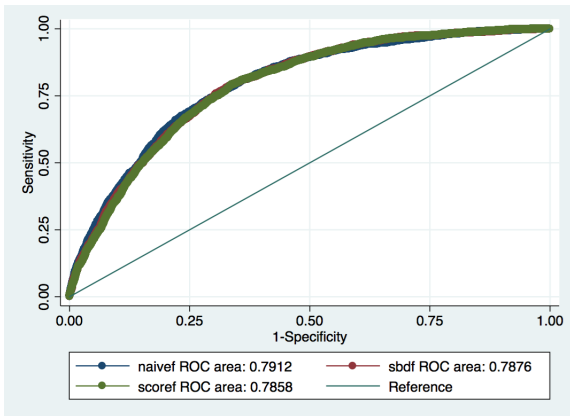
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Test for equality (DeLong 1988) $p < 0.01$



Odds of Hypertension: Naive

Methods

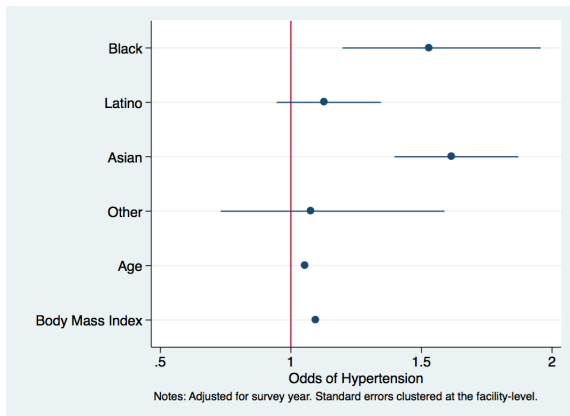
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Males only



Odds of Hypertension: SBD

Methods

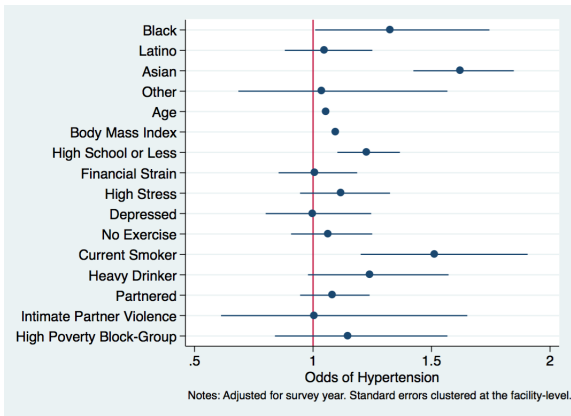
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Males only



Odds of Hypertension: SBD Index

Methods

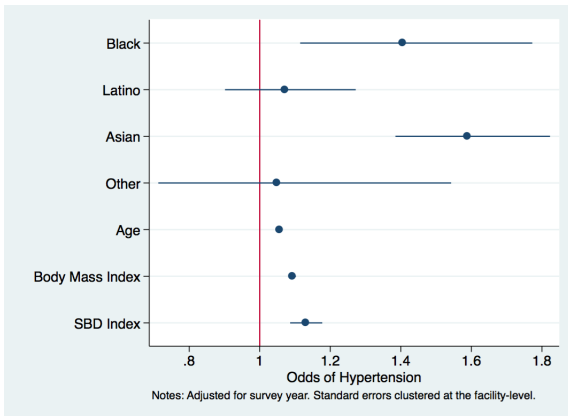
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Males only



Area Under the ROC

Methods

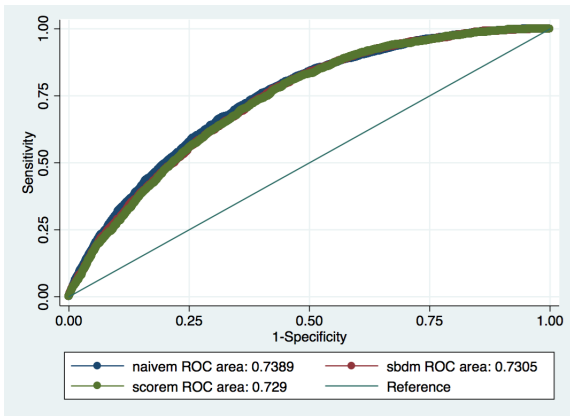
Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Test for equality (DeLong 1988) $p < 0.01$



Kaplan Meier: By Sex

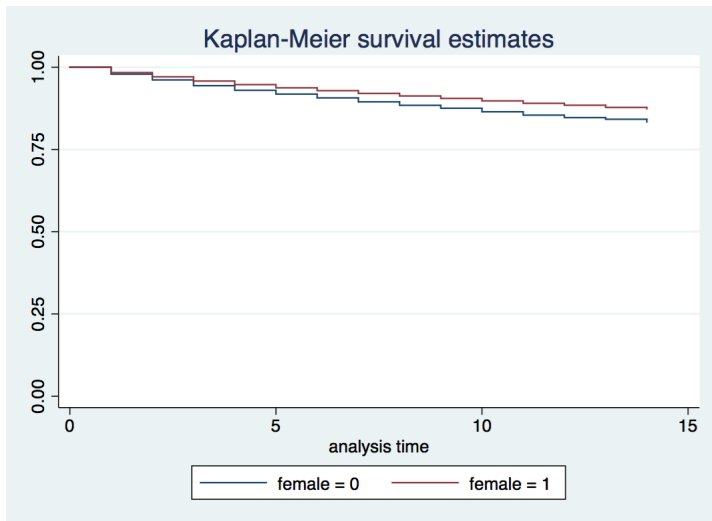
Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition



Kaplan Meier: By Age

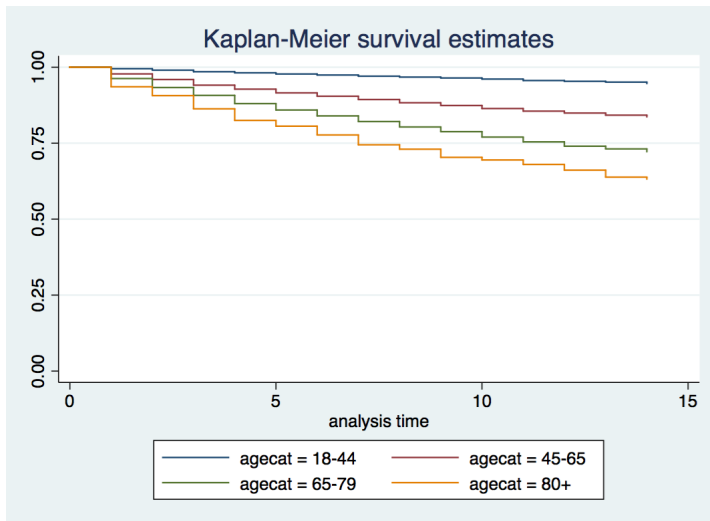
Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition



Kaplan Meier: By Obese

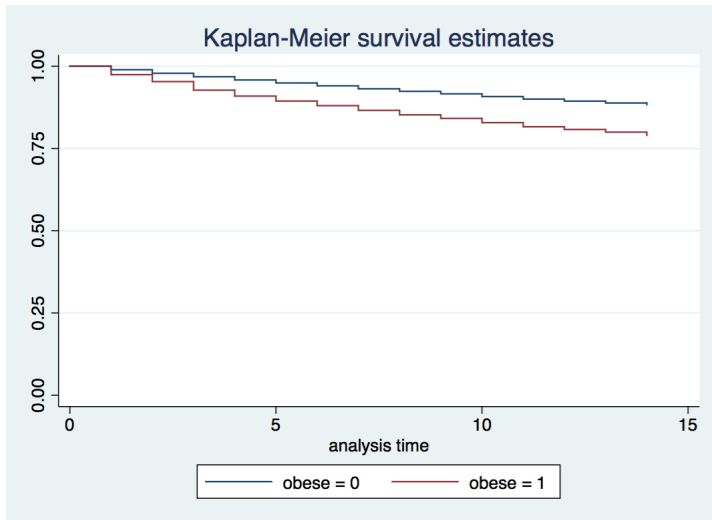
Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition



Hazard Ratios



Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition

	(1) Naive		(2) SBD Variables		(3) Score	
Age	1.052***	(0.000)	1.053***	(0.000)	1.053***	(0.000)
Female	0.867***	(0.000)	0.851***	(0.000)	0.823***	(0.000)
White	1	(.)	1	(.)	1	(.)
Black	1.391***	(0.000)	1.325***	(0.000)	1.335***	(0.000)
Latino	1.138*	(0.021)	1.092	(0.148)	1.116	(0.070)
Asian	1.555***	(0.000)	1.611***	(0.000)	1.606***	(0.000)
Other	0.985	(0.887)	1.010	(0.930)	1.013	(0.909)
Body Mass Index	1.084***	(0.000)	1.084***	(0.000)	1.083***	(0.000)
High School or Less			1.211***	(0.000)		
Financial Strain			1.071	(0.119)		
High Stress			1.089	(0.131)		
Depressed			0.975	(0.654)		
No Exercise			1.125*	(0.014)		
Current Smoker			1.343***	(0.000)		
Heavy Drinker			1.281***	(0.000)		
Partnered			1.018	(0.655)		
Intimate Partner Violence			1.170	(0.206)		
High Poverty Block-Group			1.104	(0.148)		
SBD Index					1.109***	(0.000)
Observations	30985		27929		27929	

Exponentiated coefficients; *p*-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Comparing Models

Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Harrell's C and Somers' D Concordance

- Naive Model: 0.7553 / 0.5015
- SBD Model: 0.7585 / 0.517
- Score Model: 0.7568 / 0.5136

Characteristics by Attrition



Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Table: Differences in Characteristics by Attrition

	No Attrition		Attrition		Difference	
	mean	sd	mean	sd	b	t
Hypertension Dx	0.14	0.34	0.07	0.26	0.06***	(15.21)
Female	0.58	0.49	0.57	0.49	0.01	(1.08)
Age	49.97	16.31	44.33	17.91	5.64***	(20.11)
Body Mass Index	27.11	5.48	26.98	5.59	0.13	(1.22)
Non-Hispanic black	0.05	0.22	0.06	0.23	-0.00	(-1.04)
High School or Less	0.18	0.38	0.20	0.40	-0.02***	(-3.79)
Financial Strain	0.25	0.43	0.33	0.47	-0.08***	(-10.26)
High Stress	0.16	0.37	0.22	0.42	-0.06***	(-9.60)
Depressed	0.12	0.32	0.14	0.35	-0.03***	(-5.41)
No Exercise	0.13	0.34	0.17	0.38	-0.04***	(-6.13)
Current Smoker	0.07	0.26	0.11	0.32	-0.04***	(-8.05)
Heavy Drinker	0.07	0.26	0.09	0.28	-0.02***	(-3.61)
Partnered	0.24	0.43	0.31	0.46	-0.07***	(-9.58)
Intimate Partner Violence	0.02	0.13	0.03	0.16	-0.01***	(-4.15)
High Poverty Block-Group	0.06	0.23	0.08	0.27	-0.02***	(-5.41)
SBD Index	1.28	1.26	1.67	1.42	-0.39***	(-16.93)
Observations	29935		4603		34538	

Odds of Attrition

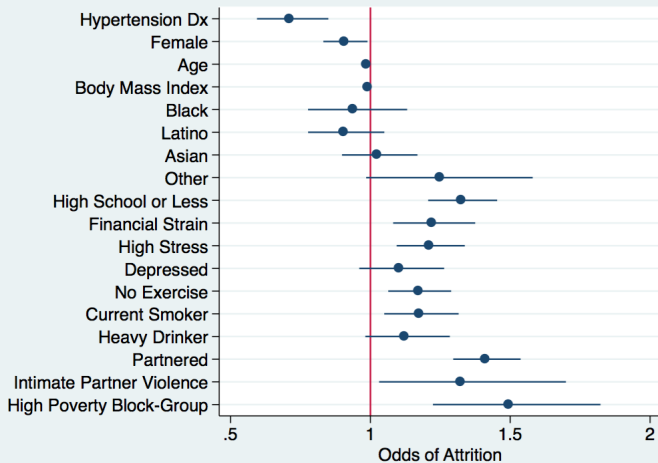
Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition



Notes: Standard errors clustered at the facility-level.

Odds of Attrition

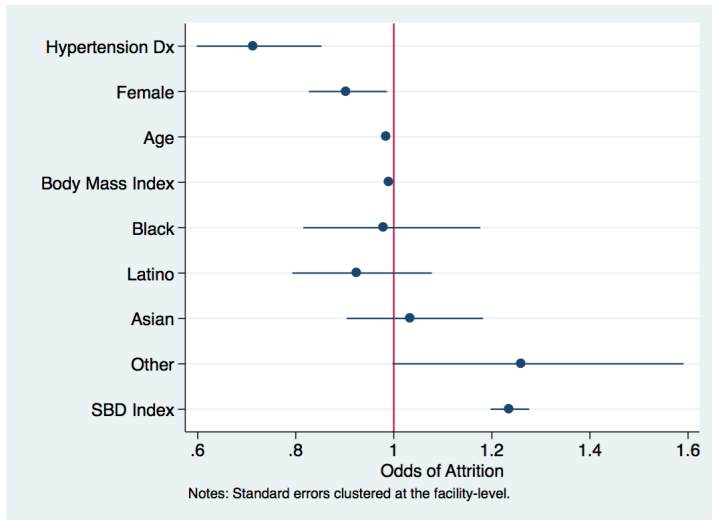
Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition



Odds of Hypertension (weighted)

Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition

	(1) Naïve		(2) SBD Variables		(3) Score	
Hypertension Dx						
White	1	(.)	1	(.)	1	(.)
Black	1.502***	(0.000)	1.414***	(0.000)	1.419***	(0.000)
Latino	1.144*	(0.032)	1.100	(0.166)	1.127	(0.078)
Asian	1.671***	(0.000)	1.731***	(0.000)	1.723***	(0.000)
Other	0.950	(0.659)	0.991	(0.944)	0.995	(0.968)
Female	0.856***	(0.000)	0.853***	(0.000)	0.823***	(0.000)
Age	1.060***	(0.000)	1.061***	(0.000)	1.061***	(0.000)
Body Mass Index	1.103***	(0.000)	1.103***	(0.000)	1.102***	(0.000)
High School or Less			1.234***	(0.000)		
Financial Strain			1.061	(0.229)		
High Stress			1.070	(0.273)		
Depressed			0.965	(0.569)		
No Exercise			1.127*	(0.034)		
Current Smoker			1.349***	(0.000)		
Heavy Drinker			1.339***	(0.000)		
Partnered			1.013	(0.789)		
Intimate Partner Violence			1.186	(0.241)		
High Poverty Block-Group			1.121	(0.152)		
SBD Index					1.111***	(0.000)
Observations	30985		27929		27929	

Exponentiated coefficients; *p*-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Comparing Models

Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Harrells C Concordance

- Naive Model: 0.7753
- SBD Model: 0.7764
- Score Model: 0.7753

Hazard Ratio (weighted)

Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition

	(1) Naive		(2) SBD Variables		(3) Score	
Age	1.052***	(0.000)	1.053***	(0.000)	1.053***	(0.000)
Female	0.868***	(0.000)	0.852***	(0.000)	0.825***	(0.000)
White	1	(.)	1	(.)	1	(.)
Black	1.400***	(0.000)	1.338***	(0.000)	1.347***	(0.000)
Latino	1.145*	(0.015)	1.101	(0.112)	1.125	(0.051)
Asian	1.557***	(0.000)	1.610***	(0.000)	1.607***	(0.000)
Other	0.972	(0.780)	1.001	(0.995)	1.005	(0.967)
Body Mass Index	1.084***	(0.000)	1.084***	(0.000)	1.083***	(0.000)
High School or Less			1.212***	(0.000)		
Financial Strain			1.072	(0.113)		
High Stress			1.091	(0.119)		
Depressed			0.974	(0.639)		
No Exercise			1.124*	(0.016)		
Current Smoker			1.340***	(0.000)		
Heavy Drinker			1.268***	(0.000)		
Partnered			1.019	(0.634)		
Intimate Partner Violence			1.159	(0.241)		
High Poverty Block-Group			1.104	(0.159)		
SBD Index					1.109***	(0.000)
Observations	30985		27929		27929	

Exponentiated coefficients; *p*-values in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Comparing Models

Methods

Descriptives

Logistic
Regression

Survival
Analysis

Attrition

Harrells C Concordance

- Naive Model: 0.7743
- SBD Model: 0.7753
- Score Model: 0.7742