

Health Inequality and Economic Disparities by Race, Ethnicity, and Gender

Nicolò Russo, Rory McGee, Mariacristina De Nardi, Margherita Borella, and Ross Abram

December 5, 2024

Presented by Jake Harmon

- How unequally distributed is health by race, ethnicity, and gender?

Introduction: Objective

- How unequally distributed is health by race, ethnicity, and gender?
- How can these disparities explain differences in key economic outcomes such as:
 - Disability
 - Length of working life
 - Nursing home entry
 - Duration of life spent in poor health
 - Lifespan

Health and Retirement Study (HRS)

- Longitudinal survey of a representative sample of Americans over age 50
- Participants interviewed every 2 years
- Participants provide data on health care, housing, assets, pensions, employment and disability

RAND HRS Longitudinal File combines individual survey data into single 1992-2018 dataset

- Data used is 1996-2018 (key variables not tracked until 1996)
- Looking only at participants under age 100 who identify as non-Hispanic White, non-Hispanic Black, or Hispanic
- Sample consists of 216,166 individual-year observations

1. Construct measure of frailty (fraction of one's possible health deficits)
2. Correct for under-reporting of medical diagnoses for Black and Hispanic individuals
3. Estimate differences in key economic outcomes

Table 1: Health deficits

Deficit	Deficit
<i>ADLs</i>	Difficulty lifting a weight heavier than 10 lbs
Difficulty bathing	Difficulty lifting arms over the shoulders
Difficulty dressing	Difficulty picking up a dime
Difficulty eating	Difficulty pulling/pushing large objects
Difficulty getting in/out of bed	Difficulty sitting for two hours
Difficulty using the toilet	
Difficulty walking across a room	<i>Diagnoses</i>
Difficulty walking one block	Diagnosed with high blood pressure
Difficulty walking several blocks	Diagnosed with diabetes
	Diagnosed with cancer
<i>IADLs</i>	Diagnosed with lung disease
Difficulty grocery shopping	Diagnosed with a heart condition
Difficulty making phone calls	Diagnosed with a stroke
Difficulty managing money	Diagnosed with psychological or psychiatric problems
Difficulty preparing a hot meal	Diagnosed with arthritis
Difficulty taking medication	
Difficulty using a map	<i>Healthcare Utilization</i>
	Has stayed in the hospital in the previous two years
<i>Other Functional Limitations</i>	Has stayed in a nursing home in the previous two years
Difficulty climbing one flight of stairs	
Difficulty climbing several flights of stairs	<i>Addictive Diseases</i>
Difficulty getting up from a chair	Has BMI larger than 30
Difficulty kneeling or crouching	Has ever smoked cigarettes

Notes: Each deficit takes a value of 0 (if the respondent reports not having it) or 1 (if the respondent reports having it).

Methodology II: Correct for Under-Reporting

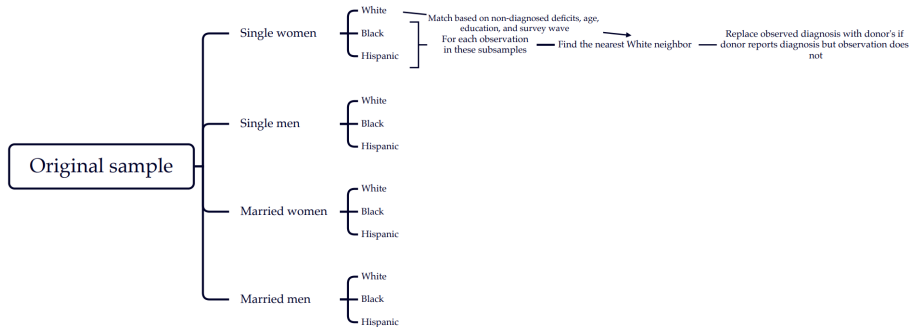
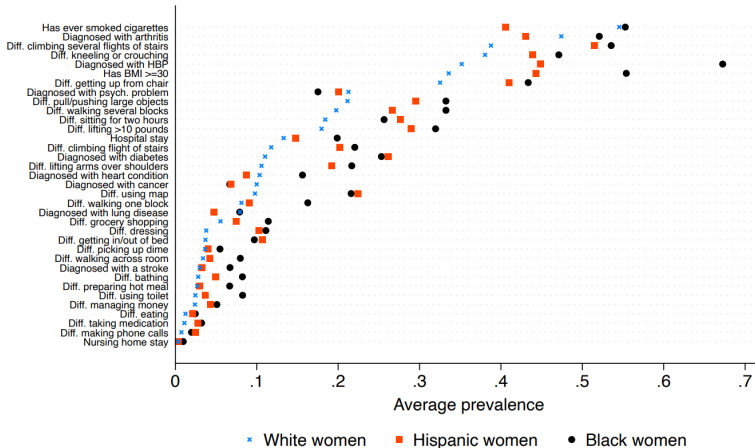


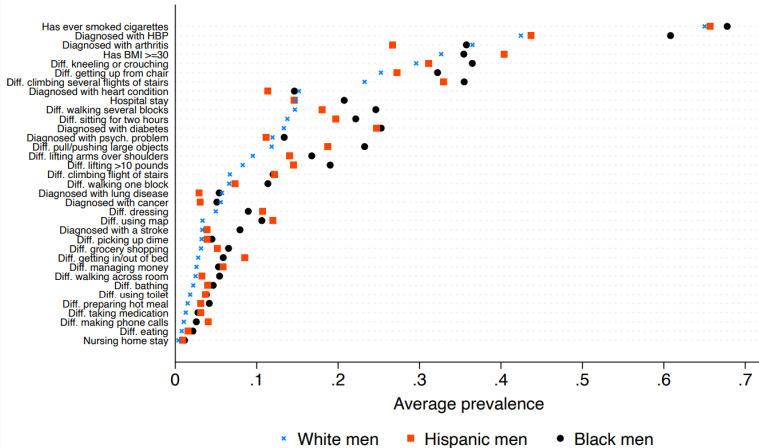
Figure A-1: Summary of our imputation procedure.

Results

Figure 1: Health deficits prevalence. Age 55-59

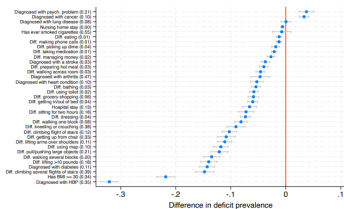


Results

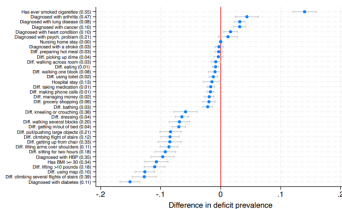


Results

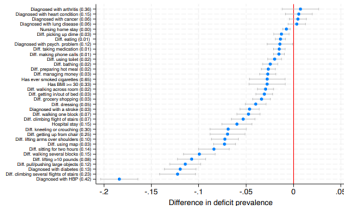
Figure 2: Differences in health deficits prevalence. Age 55-59. Positive values indicate a deficit is more common among White individuals, while negative values show higher prevalence among non-White individuals.



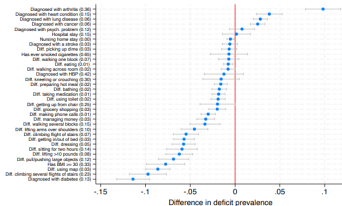
(a) White and Black women



(b) White and Hispanic women



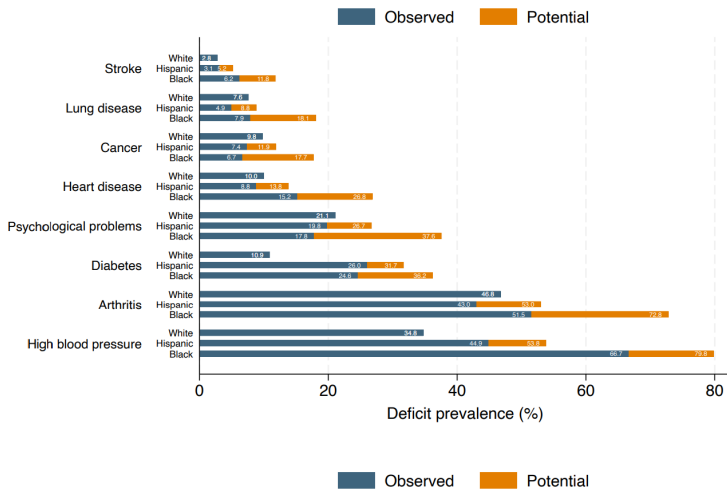
(c) White and Black men



(d) White and Hispanic men

Results

Figure 3: Potential health deficits prevalence. Age 55-59



Results

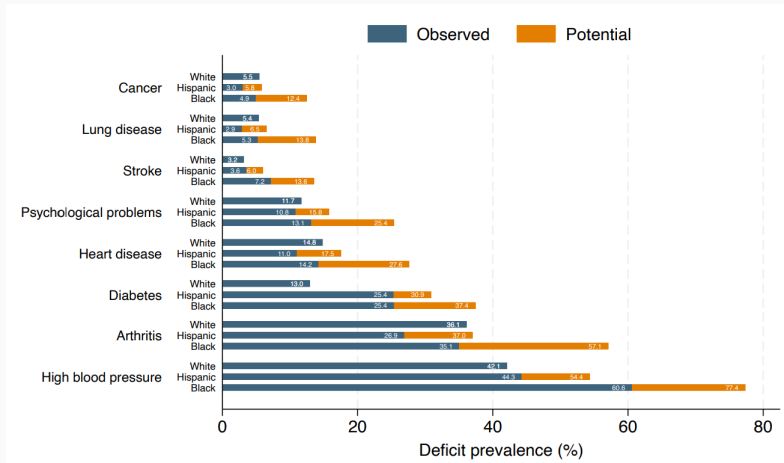
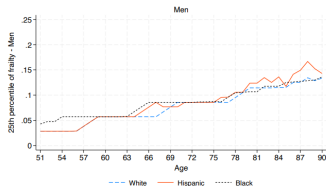


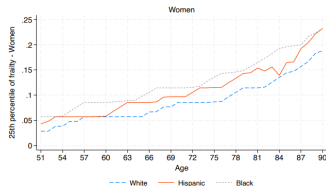
Table 2: Pseudo-R² table

		Women			Men		
		White	Hispanic	Black	White	Hispanic	Black
SDI Recipient Next Wave	Basic Controls	0.048	0.046	0.036	0.045	0.022	0.032
	SRHS	0.212	0.122	0.129	0.186	0.112	0.122
	Frailty	0.244	0.193	0.185	0.245	0.222	0.175
	Frailty and SRHS	0.268	0.202	0.199	0.264	0.241	0.196
SS Benefits Recipient Next Wave	Basic Controls	0.118	0.081	0.083	0.134	0.101	0.120
	SRHS	0.128	0.110	0.102	0.140	0.128	0.126
	Frailty	0.126	0.091	0.097	0.142	0.112	0.139
	Frailty and SRHS	0.132	0.123	0.114	0.147	0.145	0.145
NH Entry Next Wave	Basic Controls	0.241	0.172	0.169	0.220	0.144	0.122
	SRHS	0.285	0.209	0.206	0.266	0.194	0.176
	Frailty	0.315	0.231	0.214	0.303	0.272	0.234
	Frailty and SRHS	0.319	0.250	0.227	0.308	0.291	0.244
Death Next Wave	Basic Controls	0.166	0.157	0.120	0.140	0.157	0.109
	SRHS	0.240	0.194	0.169	0.219	0.212	0.151
	Frailty	0.266	0.221	0.189	0.237	0.244	0.176
	Frailty and SRHS	0.276	0.230	0.201	0.251	0.253	0.182
<i>Percentage change from basic controls</i>							
SDI Recipient Next Wave	SRHS	341%	166%	260%	318%	412%	283%
	Frailty	407%	320%	416%	450%	916%	449%
	Frailty and SRHS	458%	341%	454%	492%	1,005%	514%
<i>Percentage change from basic controls</i>							
SS Benefits Recipient Next Wave	SRHS	9%	37%	23%	5%	27%	5%
	Frailty	7%	13%	17%	6%	11%	16%
	Frailty and SRHS	12%	53%	38%	10%	43%	21%
<i>Percentage change from basic controls</i>							
NH Entry Next Wave	SRHS	18%	21%	22%	21%	35%	44%
	Frailty	31%	34%	27%	38%	89%	92%
	Frailty and SRHS	32%	45%	34%	40%	102%	102%
<i>Percentage change from basic controls</i>							
Death Next Wave	SRHS	45%	24%	41%	57%	35%	39%
	Frailty	60%	41%	57%	69%	55%	62%
	Frailty and SRHS	66%	47%	67%	79%	61%	61%

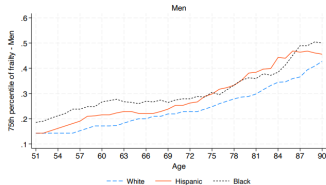
Results



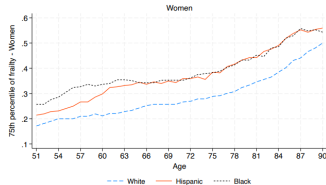
(a) 25th percentile of frailty. Men



(b) 25th percentile of frailty. Women



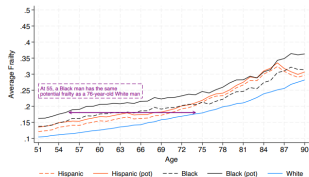
(c) 75th percentile of frailty. Men



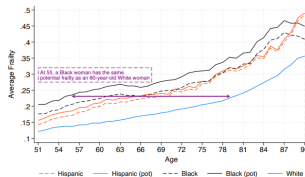
(d) 75th percentile of frailty. Women

Figure 5: 25th (first row) and 75th (second row) frailty percentile by age. Men (left column) and women (right column). Each statistic is smoothed using a three-year moving average.

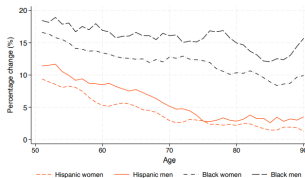
Results



(a) Average frailty. Men



(b) Average frailty. Women



(c) Percentage change

Figure 6: Comparison between observed and potential frailty for men (Panel (a)) and women (Panel (b)) and within-race percentage change between observed and potential frailty (Panel (c)). The averages in Panels (a) and (b) are smoothed using a three-year moving average. The percentage change in Panel (c) is computed using the smooth averages from Panels (a) and (b).

Results

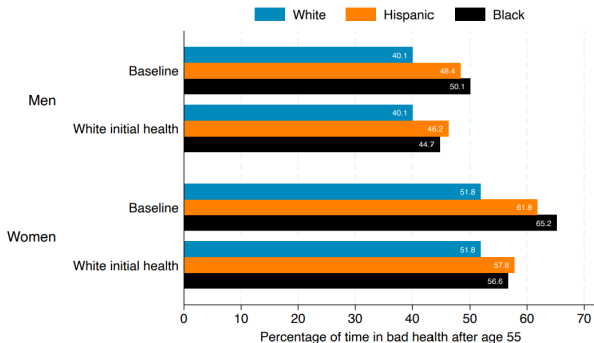


Figure 8: Average fraction of remaining life spent in bad health starting from age 55. This is computed as the fraction of remaining life spent in one of the two lowest health states (“poor” and “fair” health, or frailty quintiles), conditional on remaining alive

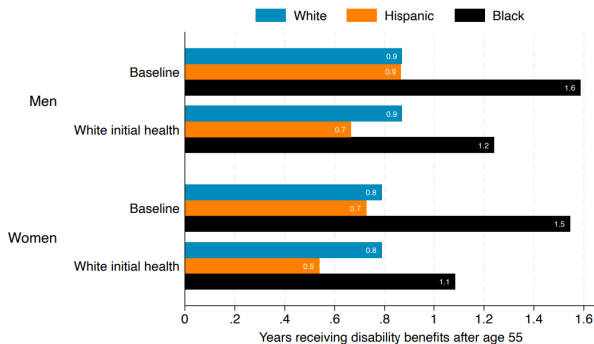


Figure 10: Average number of years receiving disability benefits after age 55

Results

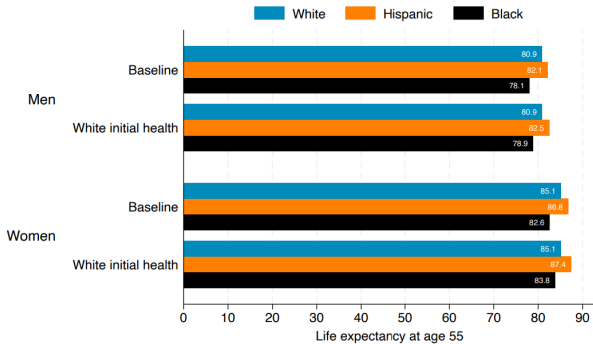


Figure 9: Average life expectancy as of age 55

Results

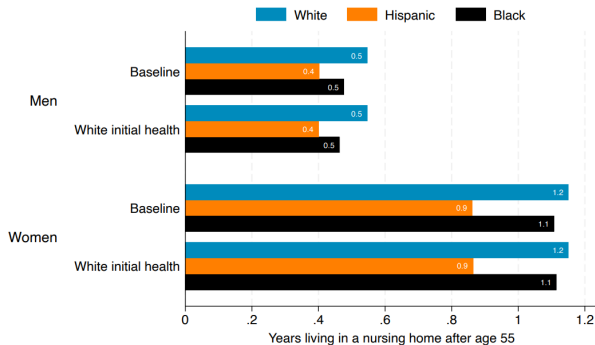


Figure 13: Average number of years in a nursing home after age 55

1. Evidence of substantial health inequality
2. White men and women also have much lower frailty (i.e., better health), on average, than Hispanic and Black ones

"Factors like informal care from extended family may have a greater impact. For instance, Almeida, Molnar, Kawachi, and Subramanian (2009) shows that Hispanic Americans have large family networks and high levels of social support, which may explain why they spend less time in nursing homes than their White and Black counterparts."

Extension: Objective

1. Incorporate additional measures of family structure
2. Incorporate additional measures of income and assets
3. Improve existing predictions of nursing home entry with additional measures
4. Logistic regression v. machine learning

$$Pr(nursinghome_{i,t+1} = 1) = NH(h_{it}, X_{i,t})$$

Where

- h is estimated health transition probabilities
- X is the set of basic controls, self reported health status, and frailty

$$Pr(nursinghome_{i,t+1} = 1) = NH(h_{it}, X_{i,t}, Z_{i,t})$$

Where

- h is estimated health transition probabilities
- X is the set of basic controls, self reported health status, and frailty
- Z is the set of variables measuring family structure and income and assets

Extension: Data

Table 1: Summary Statistics

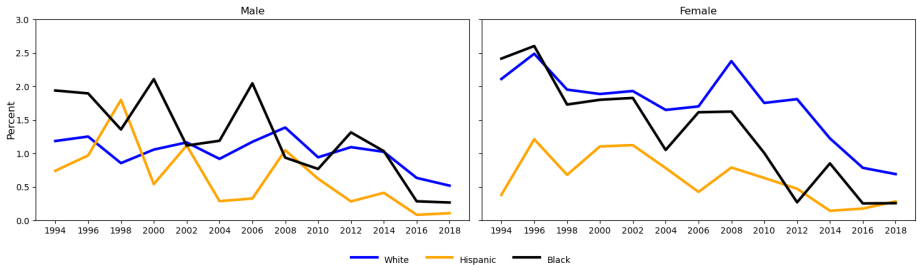
	White		Hispanic		Black	
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev
<i>Family Structure</i>						
Number of Living Children	2.996	<i>1.923</i>	3.820	<i>2.411</i>	3.663	<i>2.509</i>
Number of Living Siblings	2.421	<i>2.083</i>	4.551	<i>3.068</i>	3.859	<i>2.939</i>
<i>Income and Assets</i>						
Total Non-Housing Wealth (thousands)	343.9	<i>1076</i>	69.64	<i>436.7</i>	57.71	<i>316.4</i>
Total Household Income (thousands)	70.20	<i>196.7</i>	38.59	<i>132.1</i>	40.95	<i>55.96</i>
Poverty Status (binary)	5.8%		26.8%		22.5%	
<i>Parents</i>						
Mother's Age (or age at death)	76.50	<i>13.95</i>	73.84	<i>15.70</i>	71.77	<i>15.89</i>
Father's Age (or age at death)	71.93	<i>14.01</i>	72.50	<i>15.16</i>	69.93	<i>14.99</i>
Mother's Education (years)	10.20	<i>3.10</i>	5.57	<i>4.35</i>	9.15	<i>3.48</i>
Father's Education (years)	9.92	<i>3.51</i>	5.86	<i>4.59</i>	8.38	<i>3.69</i>
<i>Other</i>						
Veteran Status (binary)	23.1%		8.8%		14.6%	
Residing in Urban area (binary)	44.9%		56.9%		62.4%	
Residing in Rural area (binary)	32.2%		13.1%		16.9%	

Note: summary statistics computed across male and female respondents and across waves 2 through 14.

- Exclude individuals that lived in nursing home last wave or before
- Oversampling the minority group of nursing home entrants

Extension: Methodology

Figure 1: Percent Nursing Home Entry Next Wave by Race and Gender



Extension: Results

Table 2: Pseudo R-squared

	Women			Men		
	White	Hispanic	Black	White	Hispanic	Black
Basic Controls	0.241	0.172	0.169	0.220	0.144	0.122
SHRS	0.285	0.209	0.206	0.266	0.194	0.176
Frailty	0.315	0.231	0.214	0.303	0.272	0.234
Frailty and SHRS	0.319	0.250	0.227	0.308	0.291	0.244
and Household Characteristics (HC)*	0.319	0.259	0.242	0.310	0.334	0.253
and Income and Assets (I&A)*	0.320	0.246	0.239	0.311	0.309	0.256
and both HC and I&A*	0.323	0.262	0.251	0.313	0.335	0.264
<i>Percent Change from Basic Controls</i>						
SHRS	18%	21%	22%	21%	35%	44%
Frailty	31%	34%	27%	38%	89%	92%
Frailty and SHRS	32%	45%	34%	40%	102%	102%
and Household Characteristics (HC)	32%	50%	43%	41%	132%	108%
and Income and Assets (I&A)	33%	43%	41%	41%	114%	110%
and both HC and I&A	34%	52%	48%	42%	133%	117%

*Average Pseudo R-squared across 4 folds of 75%/25% stratified K-Fold cross validation training/test sets

- L1/L2 Regularization (Elastic Net)
- Random Forest
- Gradient Boosting

Extension: Machine Learning Results

Table 3: Model Comparisons (Pseudo R-squared)

	Women			Men		
	White	Hispanic	Black	White	Hispanic	Black
Logistic Regression	0.323	0.262	0.251	0.313	0.335	0.264
Logistic Regression with L1/L2 Regularization	0.323	0.258	0.251	0.314	0.329	0.265
Random Forest	0.490	0.532	0.469	0.491	0.624	0.511
XGBoost	0.616	0.8919	0.739	0.6393	0.9129	0.8274

Note: Average Pseudo R-squared across 4 folds of 75%/25% stratified K-Fold cross validation training/test sets

Extension: Machine Learning Results

Table 3: Model Comparisons (Pseudo R-squared)

	Women			Men		
	White	Hispanic	Black	White	Hispanic	Black
Logistic Regression	0.323	0.262	0.251	0.313	0.335	0.264
Logistic Regression with L1/L2 Regularization	0.323	0.258	0.251	0.314	0.329	0.265
Random Forest	0.490	0.532	0.469	0.491	0.624	0.511
XGBoost	0.616	0.8919	0.739	0.6393	0.9129	0.8274

Note: Average Pseudo R-squared across 4 folds of 75%/25% stratified K-Fold cross validation training/test sets

"McFadden (1977) argues that values between 0.2 and 0.4 denote an "excellent fit" of the full model."

Extension: Machine Learning Results

$$F_1 \text{ score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Extension: Machine Learning Results

Table 4: Model Comparisons (F1 Score)

	Women			Men		
	White	Hispanic	Black	White	Hispanic	Black
Logistic Regression	0.974	0.992	0.982	0.985	0.992	0.983
Random Forest	0.8869	0.9327	0.9107	0.892	0.9553	0.9321
XGBoost	0.9104	0.9728	0.9422	0.9195	0.9818	0.9645
<i>Percent Change from Logistic Regression</i>						
Random Forest	-9.0%	-5.9%	-7.2%	-9.5%	-3.7%	-5.1%
XGBoost	-6.6%	-1.9%	-4.0%	-6.7%	-1.0%	-1.8%

Note: Average Pseudo R-squared across 4 folds of 75%/25% stratified K-Fold cross validation training/test sets

Extension: Summary

1. Incorporate additional measures of family structure
2. Incorporate additional measures of income and assets
3. Improve existing predictions of nursing home entry with additional measures
4. Logistic regression v. machine learning