Health inequality and economic disparities by race, ethnicity, and gender

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

December 15, 2024

Abstract

How large is health inequality in middle age, and how does it shape subsequent economic disparities by race, ethnicity, and gender? Using the Health and Retirement Study, we document severe health disparities. At age 55, Black men and women exhibit the frailty levels, or the biological age, of White individuals 13 and 20 years older. Hispanic men and women show comparable frailty to White individuals 5 and 6 years older. Equalizing health at age 55 would reduce future disparities in many key economic outcomes by 40-70%. This suggests that targeted earlier health interventions for minorities could significantly narrow economic and quality-of-life inequalities in middle and old age.

*Nicolò Russo: Goethe University Frankfurt; Rory McGee: Western University; Mariacristina De Nardi: University of Minnesota, Federal Reserve Bank of Minneapolis, CEPR, and NBER; Margherita Borella: University of Torino and CeRP-Collegio Carlo Alberto, Italy; Ross Abram: Churchill Asset Management.
We thank Bocar Ba, Courtney Coile, Martín García-Vázquez, Jon Skinner, and Marcos Vera Hernandez for helpful comments. The research reported herein was performed pursuant to grant RDR18000003 from the US Social Security Administration (SSA), funded as part of the Retirement and Disability Research Consortium. Borella gratefully acknowledges funding from the European Union - Next Generation EU, Mission 4 Component 2, within the NRRP (National Recovery and Resilience Plan) project "Age-It - Ageing well in an ageing society" (PE0000015 - CUP H73C22000900006).

1 Introduction

How unequally distributed is health by race, ethnicity, and gender? To what extent can these health disparities explain differences in key economic outcomes such as disability, length of working life, nursing home entry, duration of life spent in poor health, and overall lifespan?

Answering these questions requires identifying a health measure that is comparable across groups and easy to interpret. Moreover, analyzing how health shapes economic outcomes and their inequality requires a rich and comprehensive evaluation of its implications for many key economic outcomes.

Compared to previous work, we make several contributions toward these goals. First, rather than adopting a single health measure, we consider several, including one that accounts for the under-reporting of diagnoses by race, ethnicity, and gender. We then assess their validity by evaluating their predictive power for many key economic outcomes across these demographic groups. Second, we use our preferred health measures to document health inequality by race, ethnicity, and gender. Third, rather than focusing on the effects of health on a single outcome, we examine its impact on multiple economic outcomes, providing a richer perspective on how health shapes both the quantity and quality of life.

Why should we worry about how we measure health by race, ethnicity, and gender? Many datasets contain information on both one's perceived overall health status and on many impairments, diagnoses, and conditions. Previous literature has documented important differences in assessing both one's health and in reporting diagnoses and medical treatment across the groups of people that we are interested in. Hence, we use both sources of information to construct several health measures, which we then evaluate in terms of their ability to predict several key economic outcomes. Specifically, we consider one's self-reported health status (SRHS) and several measures of frailty. SRHS, commonly used in economic studies, relies on individuals rating their health as excellent, very good, good, fair, or poor. In contrast, frailty, originating from the medical literature, is constructed by using health deficits, thereby serving as a measure of biological age. When constructing frailty, the underlying rich health information requires making two main sets of choices: which health deficits to include and how to weigh them. In terms of which health deficits to include, we follow the medical literature and include difficulties with activities of daily living, diagnoses, and several other health outcomes. Regarding weighting, we start from the frailty measure proposed by the medical literature, which weighs all deficits equally. Then, we use principal component analysis (PCA) to determine the weights attributed to the various health deficits to construct a PCA-frailty measure. To address the issue of underreporting and underdiagnosis of medical conditions for non-White people, we also construct a measure of "potential" frailty, which accounts for this underreporting. Our main conclusion from this part is that both SRHS and frailty are highly predictive of economic outcomes but that frailty does somewhat better. Importantly, we also find that PCA-frailty does not outperform frailty in its baseline, equally weighted version. Consequently, we focus on frailty as our primary health measure for the remainder of our analysis.

Second, we measure health inequality by race, ethnicity, and gender using frailty and potential frailty. Our analysis reveals enormous inequality in frailty by race, ethnicity, and gender. At age 55, Black men and women have frailty levels, or a biological age, comparable to White men and women who are 13 and 20 years older, respectively. The corresponding gaps for our measure of potential frailty are even larger, 20 and 25 years. It is worth pointing out that these gaps in biological age are much larger than the differences in life expectancy (which are typically used to assess both health inequality and evaluate its effects), which range between three and four years.

Third, we provide a comprehensive analysis of how health affects many key future economic outcomes by estimating a dynamic system and performing a counterfactual in which we give Black and Hispanic men and women the distribution of health at age 55 of their White counterparts. Our exercise reveals that frailty at age 55 is a powerful determinant of economic inequality later in life. White men and women spend 40% and 52% of their remaining years in poor health, respectively, compared to 50% and 65% for Black men and women and 48% and 62% for Hispanic men and women. Equalizing health at age 55 reduces the health span gap between Black and White individuals—by 54% for men and 64% for women. Moreover, health disparities at midlife substantially contribute to life expectancy differences, where eliminating these disparities reduces the lifespan gap by 29% for Black men and 46% for Black women. Thus, racial and gender disparities in health in middle age generate large differences in both the *quality* and *quantity* of remaining life, as measured by individuals' health and life span. Health inequality also significantly affects other economic outcomes, including disability and retirement duration, with Black individuals over 55 spending twice as long on disability compared to White and Hispanic people. Equalizing health gaps at this age halves this disparity. Additionally, Black individuals receive the shortest duration of retirement benefits post-55, with health disparities at this age accounting for nearly half of this discrepancy with their White counterparts. For reasons discussed later in the paper, we perform these decompositions on frailty rather than potential frailty.

We derive the latter set of results by estimating a rich set of dynamic equations relating health and other observables to our economic outcomes of interest and using its estimated parameters to generate our counterfactuals. This methodology leverages empirical relationships between health and economic outcomes, allowing us to focus directly on the disparities that we study without the added complexity of a fully (and correctly) specified behavioral model. Importantly, our counterfactuals of interest only change initial conditions. Just like any policy functions derived from a structural model are invariant functions of initial conditions and their subsequent histories, so are our estimated laws of motion. Our exercise is thus not subject to the Lucas' critique because of the specific counterfactuals that we study.

Ours can be thought of as a semi-structural approach (e.g., Altonji, Smith Jr, and Vidangos (2013); Arellano, Blundell, and Bonhomme (2017)) in the sense that there is a structural model guiding the empirical approach. This is complementary to specifying and estimating a structural model. While structural models are powerful tools that enable evaluating a wider range of counterfactuals, they require detailed functional form assumptions and the calibration or estimation of many parameters. It is also important to note that whether one uses policy functions coming from a structural model or the flexible relationships between health and economic outcomes that we estimate, the thought experiment is the same: change initial conditions and evaluate how the effects of policy rules change as a result.

Thus, our method contributes to the broader understanding of how health disparities shape economic outcomes. The structural literature, including studies such as French (2005), De Nardi, French, and Jones (2009), and Attanasio, Kitao, and Violante (2010), has provided valuable insights into the interactions between health, labor supply, savings, and retirement. However, this work abstracts from racial and ethnic disparities. By focusing on these gaps, our approach complements the structural literature by providing a flexible and empirically grounded framework to study the economic implications of health inequality.

Previous work evaluating inequality by race, ethnicity, and gender mostly adopts one health measure without evaluating it across groups and studies its effects on a smaller set of outcomes. For instance, Cobb, Thomas, Laster Pirtle, and Darity (2016) uses allostatic load to measure health inequality among White and Black people. Currie and Schwandt (2016) studies mortality by race. Halliday, Mazumder, and Wong (2021) quantifies racial gaps in health mobility. Blundell, Britton, Dias, French, and Zou (2022) and Blundell, Britton, Dias, and French (2023) look at the effect of health on employment. Beck et al. (2014) finds that Black Americans have lower self-reported health status than White ones. Wu et al. (2023) studies frailty inequality by race and ethnicity among older adults in Hawaii and California and finds that average frailty is highest for Black and Latino individuals.

Recent work by Danesh, Kolstad, Spinnewijn, and Parker (2024) uses healthcare claims data from the Netherlands to infer chronic health conditions and construct a chronic disease index with weights determined via a Double-Lasso procedure based on the effect each chronic disease has on mortality at age 70. They use it to study health inequality by income. Our works are complementary in that we have a much broader set of health impairments while they also have data on younger individuals. In addition, they focus on a relatively more homogeneous population of Dutch people, while we are particularly interested in heterogeneity by race, ethnicity, and gender.

The rest of the paper is organized as follows. Section 2 discusses our data and variables construction. Section 3 evaluated the predictive power of frailty and SRHS. Section 4 quantifies health inequality. Section 5 quantifies the effects of removing health inequality on economic inequality. Section 6 concludes.

2 Data

We use data from the Health and Retirement Study (HRS), which began in 1992 and is conducted every two years. The HRS provides data on U.S. residents aged 51 and older, as well as their spouses, and oversamples Black and Hispanic individuals (HRS Staff (2017)). Several studies have documented the high quality of the HRS in recruiting and retaining minority respondents (Ofstedal and Weir (2011) and Schroeder, Weir, and West (2023)).

Because key variables such as difficulties with activities of daily living (ADLs) first appeared in the 1996 survey, we use data from 1996 to 2018. We select respondents younger than age 100 who identify as non-Hispanic White, non-Hispanic Black, or Hispanic.[1] Our sample consists of 216,166 individual-year observations. Appendix A presents more details.

The first step in constructing a frailty index is selecting which health deficits to include. We follow the guidelines in Searle et al. (2008) and select 35 binary deficits for our baseline frailty index. In a robustness exercise, detailed in Section 3.1, we augment our baseline index with more deficits and construct a frailty index composed of 51 deficits. Appendix A.1 reports more details on all health deficits.

1. We follow the 2020 U.S. Census (available at https://www.census.gov/programs-surveys/decennial-census/technical-documentation/questionnaires.2020_Census.html, which categorizes "White" and "Black" as races, and "Hispanic" as an ethnicity. Our data does not allow us to distinguish races further. The HRS race variable takes three values: White, Black, and "other," which includes American Indians, Alaskan Natives, Asians, Native Hawaiians, and Pacific Islanders. In our unselected starting sample, these observations make up between 5 and 10% of the total sample. However, because the groups in the "other" race category are very different from each other, we drop them from our sample.

To facilitate exposition, Table 1 groups deficits consistently with the Katz Index of Independence in Activities of Daily Living (Katz, Downs, Cash, and Grotz (1970) and Katz (1983)), which is a tool used by medical professionals to assess one's ability to perform basic activities independently. These groups comprise activities of daily living (ADLs), difficulties with instrumental activities of daily living (IADLs), and other functional limitations. ADLs refer to basic activities required to take care of oneself and include having difficulty bathing and dressing. IADLs refer to more complex activities that allow people to live independently. We include as IADLs the deficits that appear in the Lawton-Brody Instrumental Activities of Daily Living scale (Lawton and Brody (1969)), which is the most common checklist used by medical professionals to determine one's difficulties with IADLs. We classify as "other functional limitations" all the remaining deficits that refer to functional limitations that do not enter either the Katz Index of Independence in Activities of Daily Living or the Lawton-Brody Instrumental Activities of Daily Living scale. The fourth and fifth grouping of deficits include diagnoses by medical professionals (as reported by the respondent) and indicators of healthcare utilization. Finally, there are addictive diseases, such as obesity (i.e., having a body-mass index (BMI) larger than 30) and smoking. Regarding the latter deficits, we follow the medical literature and classify obesity and smoking as diseases. The American Medical Association (AMA) recognized obesity as a chronic disease in 2013. Many papers in the medical literature (for instance, Bernstein and Toll (2019)) also consider smoking to be a chronic disease.

To augment our baseline frailty index, we include deficits related to chronic pain (Skinner and Atlas (2010) documents the prevalence of pain among HRS respondents, emphasizing its importance across demographic groups), mental health, cognition, and harmful habits like smoking and being a heavy alcohol user.

Table 1: Health deficits

Deficit	Deficit
Baseline frailty	
ADLs	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	***Diagnoses***
Difficulty walking one block	Diagnosed with high blood pressure
Difficulty walking several blocks	Diagnosed with diabetes
	Diagnosed with cancer
IADLs	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	***Healthcare Utilization***
	Has stayed in the hospital in the previous two years
Other Functional Limitations	Has stayed in a nursing home in the previous two years
Difficulty climbing one flight of stairs	
Difficulty climbing several flights of stairs	***Addictive Diseases***
Difficulty getting up from a chair	Has BMI larger than 30
Difficulty kneeling or crouching	Has ever smoked cigarettes
Augmented frailty	
Pain	***Cognition***
Frequently troubled by pain	Gets lost in familiar environment
	Wanders off
Mental health	Cannot be left alone
Felt depressed	Has hallucinations
Felt like everything was an effort	Cannot count backwards from 20
Had restless sleep	
Did not feel happy most of the time	***Harmful habits***
Felt alone	Smokes now
Felt sad	Heavy alcohol use
Could not get going	
Did not enjoy life	

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

2.1 Deficits Prevalence

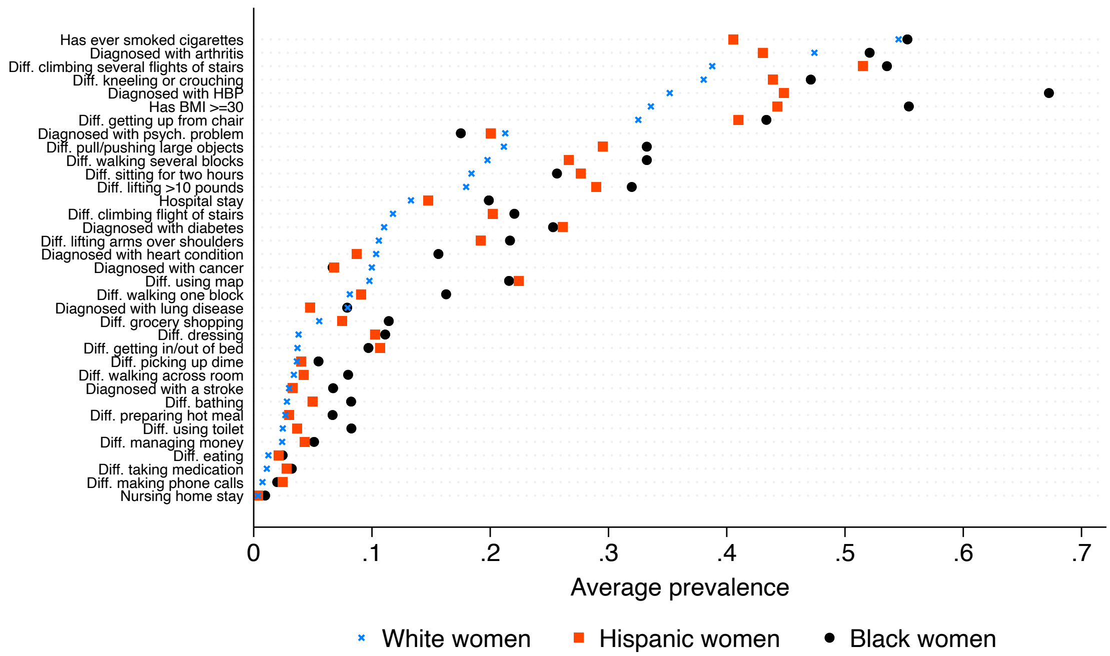
Figure 1 summarizes the prevalence of deficits for the 55-59 age group, for both women and men, for all deficits included in our baseline measure of frailty.[2] It shows that the most prevalent deficit for women varies by race. For White women, it is having ever smoked (54.5%); for Hispanic women, it is having difficulties climbing several flights of stairs (51.5%); and for Black women, it is high blood pressure (67.2%). In contrast, the most common deficit for men in these three groups is high blood pressure, affecting 42.4%, 43.7%, and 60.8% of White, Hispanic, and Black men, respectively. This is consistent with the prevalence of high blood pressure by age and race reported in McWilliams, Meara, Zaslavsky, and Ayanian (2009).

Among other key deficits, obesity and diabetes are more prevalent among Hispanic and Black men and women (as also found by Peek, Cargill, and Huang (2007) and Petersen, Pan, and Blanck (2019)). The share of obese (i.e., with a BMI greater than 30) White women is 33.6%, and those of Hispanic and Black women are 44.3% and 55.4%. Similarly, while 32.7% of White men are obese, 35.4% and 40.4% of Hispanic and Black men are. Diabetes affects 11.0% of White women and 26.1% and 25.3% of Hispanic and Black women. While 13.3% of White men have diabetes, 24.7% and 25.3% of Hispanic and Black men suffer from it. Finally, while 38.8% of White women report having difficulties climbing several flights of stairs, this share rises to 51.5% and 53.5% for Hispanic and Black women. Moreover, 23.3% of White men report having difficulty climbing several flights of stairs, compared to 33.0% and 35.5% for Hispanic and Black men.

Figure 2 reports the differences in health deficit prevalence between White men and women and their Black and Hispanic counterparts. It shows that while most deficits are significantly more prevalent among Black and Hispanic individuals, the medical diagnosis of various conditions is typically less frequent. This may indicate that, as suggested by the

2. We do not report data for our younger group (ages 51 to 54) because, due to the nature of the sampling frame, it is the smallest group and under-represents men.

Figure 1: Health deficits prevalence. Age 55-59



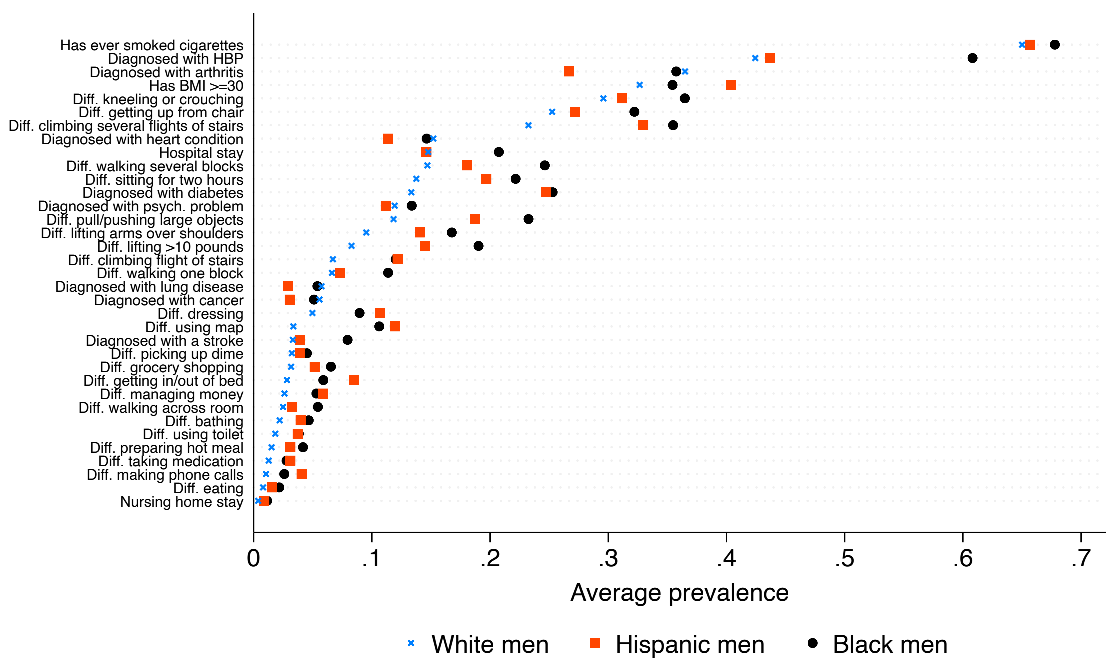
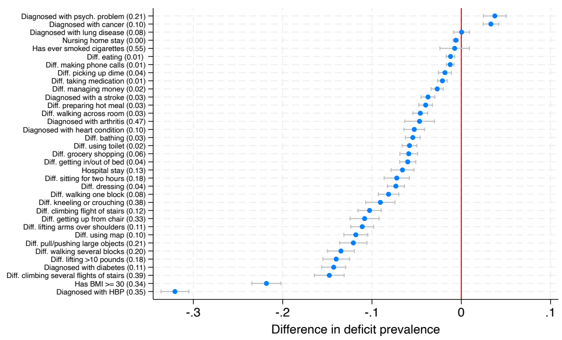
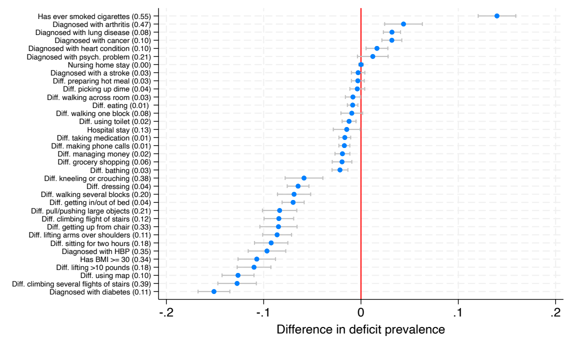


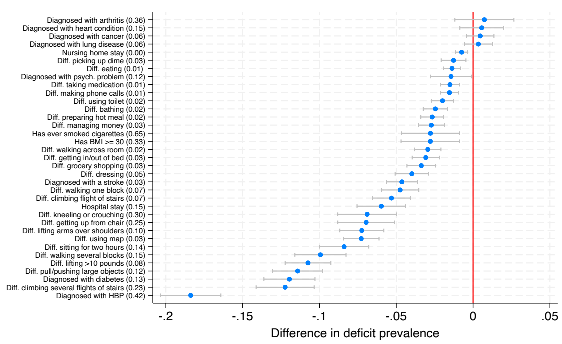
Figure 2: Differences in health deficits prevalence. Age 55-59. Positive values indicate a deficit is more common among White individuals, while negative values show higher prevalence among non-White individuals.



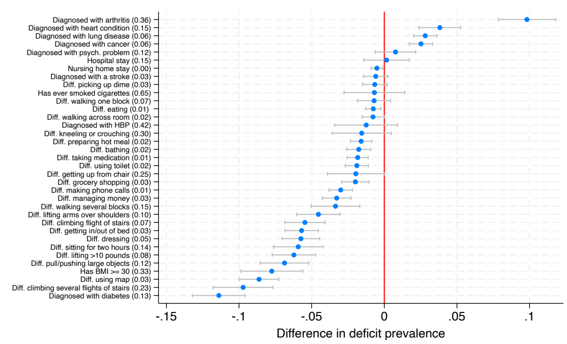
(a) White and Black women



(b) White and Hispanic women



(c) White and Black men



(d) White and Hispanic men

Notes: Dots: Prevalence differences in health deficits. Ticks: 95% confidence intervals. Deficit labels include numbers in parentheses showing the average prevalence for the White group.

medical literature, Black and Hispanic populations are underdiagnosed (see Appendix B for a discussion).

2.2 Constructing Frailty and Potential Frailty

Equally-weighted **frailty** is the ratio between a person's health deficits at a certain age and the total number of deficits considered. To construct our baseline measure of frailty, we use the 35 health deficits that we described, and we weigh them equally. In Section 3.1, we also experiment with weighing deficits using Principal Component Analysis.

Because of the evidence of differential diagnosis by group, we also construct **potential frailty**, which imputes diagnosed conditions for Hispanic and Black individuals. To do this, we identify a group of White individuals with access to either government-provided health insurance or private health insurance plans, for whom we believe under-diagnosis is likely to be limited.[3] For each Black and Hispanic individual, we select their nearest neighbor of the same gender and marital status in the corresponding insured-White subsample based on the 27 non-diagnosed deficits, age, education, and survey wave. Once a White "donor" is assigned to each non-White observation, we replace the observed diagnosed deficits with those of the donor whenever the donor reports a diagnosis that the non-White observation does not. This procedure exploits the underlying biological correlations between the conditions that we observe and environmental factors that contribute to health by additionally matching to those of a similar age, birth cohort, and education. We then use the imputed diagnosed deficits along with the 27 remaining non-imputed deficits to compute potential frailty for Black and Hispanic individuals. Our imputation strategy is similar in spirit to that of Meyer, Mittag, and Goerge (2022). Appendix B provides the details of this procedure and includes an imputation validation exercise.

3. For example, results from the National Health and Nutrition Examination Survey, which uses clinical diagnoses as well as in-person assessments and lab work, show that over 90% of White individuals with hypertension are covered by insurance (Hayes et al. 2022).

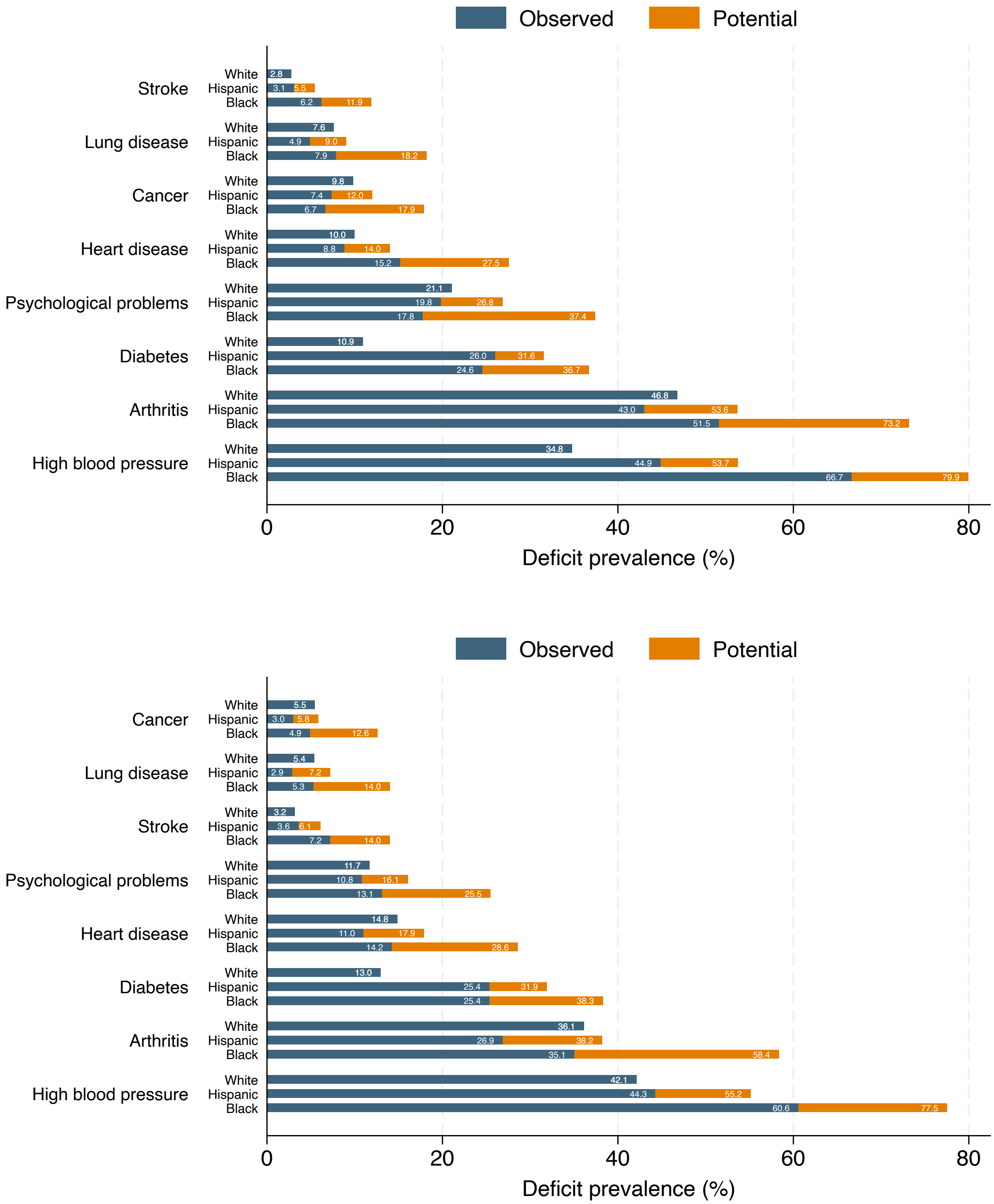
2.3 Potential Deficits Prevalence

Figure 3 compares the prevalence of observed and potential diagnosed deficits for men and women between age 55 and 59.[4] It provides several important findings. First, it confirms that potential diagnosed deficits are substantially more prevalent than observed deficits for Black and Hispanic individuals. Second, it highlights that the most under-diagnosed deficit is lung disease, while the least under-reported deficit is high blood pressure. That is, for Black men, potential lung disease is 164.2% more prevalent than diagnosed lung disease, and potential high blood pressure is 27.9% more prevalent than diagnosed blood pressure. These patterns are consistent with the fact that while high blood pressure can be assessed by a medical professional other than a doctor, diagnosing lung disease requires access to a specialist. Third, the figure indicates that under-diagnosis is more widespread for Black people than for Hispanic people. The largest difference between Black and Hispanic people is in women's cancer, with Black women having a percentage change between potential and observed cancer rates of over 105 percentage points higher than Hispanic ones. The only exception to this is women's high blood pressure. In this case, there is almost no difference between this potential and observed deficit. Finally, the figure suggests that under-reporting is worse for men than for women. This is especially true for lung disease in Hispanic people: in this case, the percentage change is 33.6 percentage points higher for men than for women. Under-reporting of cancer, stroke, and psychological problems is more severe for Black men than for Black women.

Using the National Health and Nutrition Examination Survey, which administers in-person blood work in addition to collecting information on diagnoses and medication, Fang et al. (2023) find under-diagnoses of diabetes in the population and race, ethnicity and gender gaps. Similarly, Hayes et al. (2022) use the same data to study rates of high blood pressure.

4. Appendix B.2 reports results for other age groups and our overall sample in a table format.

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



Notes: The top panel is for women, and the bottom panel is for men. Blue bars denote observed prevalence, while the orange ones denote potential prevalence.

3 How Should We Measure Health?

We now turn to comparing the extent to which frailty and SRHS help predict becoming a disability insurance recipient, starting to receive Social Security retirement benefits, entering a nursing home, and dying. To do this, we estimate logistic regressions for each of these four outcomes. Appendix C provides more details about our empirical strategy.

It is important to note that we use frailty instead of potential frailty in our regressions. This is because, since we estimate our specifications separately by race and ethnicity, potential frailty offers no additional predictive power compared to frailty. Rather, our estimated coefficients on frailty and potential frailty account for any systematic differences in frailty and their correlations with other variables by race and ethnicity (with the interpretation of these coefficients differing accordingly).

Table 2 reports the pseudo-R^2 values for our logistic regressions. For each outcome, the first row of results (labeled "Basic Controls") refers to a regression with our basic controls only. The following rows report the results when adding one of our two measures of health. The last row for each group of outcomes includes both of our measures of health.[5]

Table 2 reveals several interesting facts. First, health is an important determinant of all outcomes for all demographic groups. That is, the pseudo-R^2 jumps up for all outcomes and groups when adding either measure of health. Second, including both SRHS and frailty helps better explain all outcomes for most of our groups and that when only one health indicator is included, frailty outperforms SRHS for most outcomes.

Third, the importance of health varies by outcome and demographic group. Health adds the most predictive power to the basic-controls-only regression for disability insurance recipiency, followed by nursing home entry in the next wave, death, and receiving Social Security benefits. The improvements in explanatory power range from 5% (for SRHS, when predicting becoming a Social Security retirement benefits recipient next wave for White

5. In Appendix C.1, we also quantify the effects of health on the economic outcomes described in this Section.

men) to 1005% (for including both SRHS and frailty, when predicting becoming a disability insurance recipient for Hispanic men). Several papers have examined the effects of health on retirement and found results consistent with ours, including French (2005), Blundell, French, and Tetlow (2016), and French and Jones (2017).

Hence, the answer to our first question is that both SRHS and frailty effectively predict key economic outcomes by race and ethnicity and, in this sense, are reliable measures of health. Combined, they predict these outcomes even more accurately. When considered individually, frailty has an edge over SRHS.

3.1 Alternative Health Deficits and Weights

We now turn to evaluating alternative versions of our frailty index to assess how additional deficits and weighting methods affect its predictive power. We begin with the baseline frailty index, constructed using 35 deficits, and compare it to three variants: PCA-weighted baseline frailty, equally-weighted frailty with 51 deficits, and PCA-weighted frailty with 51 deficits.

To evaluate whether our baseline frailty is a better measure of health than its variants, we compare its predictive power for the outcomes described in the previous section by estimating logistic regressions, changing the frailty measure, and comparing the resulting pseudo-R^2.

For our PCA-frailty, we perform principal component analysis (PCA) and derive weights based on the first principal component of the deficits, following Poterba, Venti, and Wise (2017) and Hosseini, Kopecky, and Zhao (2022). Figure 4 shows that PCA increases the weights of ADL and IADL deficits while assigning lower or even negative weights to behaviors like heavy alcohol use and smoking. Negative weights for heavy alcohol use arise because it is negatively correlated with most deficits (and positively correlated with lung disease, depression, and other behavioral deficits). Importantly, PCA-frailty shows no significant improvement in predictive power over the equally weighted baseline frailty (see Table 3).[6]

6. Appendix C.2 reports the proportion of variance in the data explained by the principal components and the PCA weights in table format.

Table 2: Pseudo-R^2 table

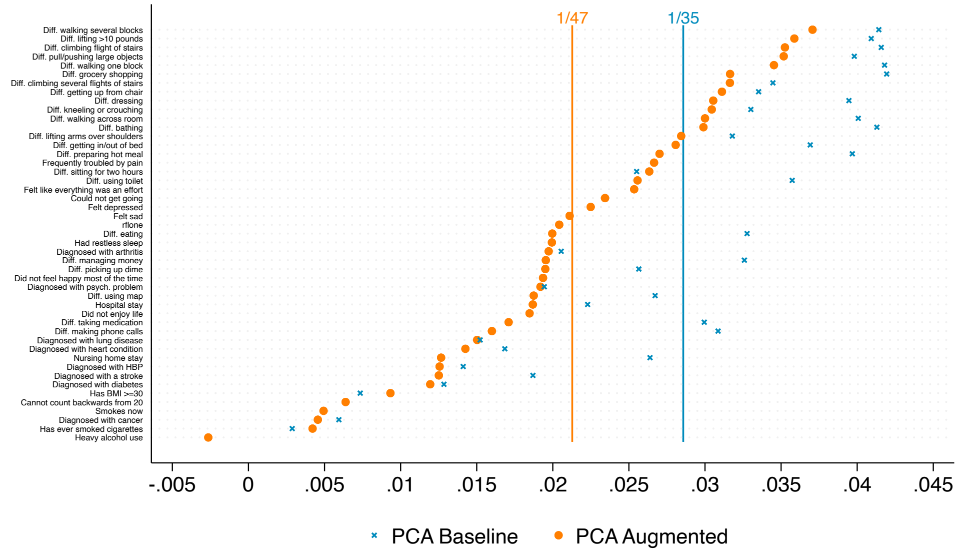
		Women			Men		
		White	Hispanic	Black	White	Hispanic	Black
SDI Recipient Next Wave	Basic Controls	0.048	0.046	0.036	0.045	0.022	0.032
	SRHS	0.212	0.122	0.129	0.186	0.112	0.122
	Frailty	0.244	0.193	0.185	0.245	0.222	0.175
	Frailty and SRHS	0.268	0.202	0.199	0.264	0.241	0.196
SS Benefits Recipient Next 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
		Percentage change from basic controls					
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%
		Percentage change from basic controls					
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%
		Percentage change from basic controls					
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%
		Percentage change from basic controls					
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%

To explore the impact of including more health information, we construct an augmented frailty index with 51 deficits. This version incorporates additional measures related to pain, mental health, cognition, and unhealthy behaviors like smoking and excessive alcohol consumption (See bottom panel of Table 1). The inclusion of these deficits also does not result in higher predictive power for most outcomes compared to the baseline frailty index.

Finally, we apply PCA to the augmented frailty index, assigning weights based on the first principal component of the 51 deficits. Because four cognition-related deficits have minimal variation (such as hallucinations and wandering off), PCA-weighting excludes them and only uses 47 deficits. Figure 4 shows that, similar to the PCA-weighted baseline frailty, this measure places greater weight on ADL and IADL deficits and less on behavioral factors. As shown in Table 3, baseline frailty outperforms PCA-weighted augmented frailty for most outcomes.

Thus, the baseline frailty index remains the most predictive health measure in our sample. Neither the addition of more deficits nor the application of PCA weighting substantially improves predictive power or alters frailty dynamics, consistent with Hosseini, Kopecky, and Zhao (2022). We thus focus on baseline frailty and, for the purposes of our descriptive analysis, also on baseline potential frailty.

Figure 4: Deficit weights



Notes: The blue markers report the weights associated with the PCA-weighted baseline frailty. The solid blue line denotes the weight for the equally weighted baseline frailty, equal to 1/35=0.0286. The orange dots represent the weights associated with the PCA-weighted augmented frailty. The solid orange line marks the weight for the equally weighted augmented frailty, equal to 1/47=0.0213.

Table 3: Pseudo-R^2 for alternative frailty measures

		Women			Men		
		White	Hispanic	Black	White	Hispanic	Black
SDI Recipient Next Wave	Baseline Frailty	0.244	0.192	0.185	0.245	0.222	0.175
	PCA Baseline Frailty	0.247	0.185	0.185	0.245	0.212	0.184
	Augmented Frailty	0.242	0.182	0.184	0.253	0.188	0.176
	PCA Augmented Frailty	0.248	0.178	0.189	0.254	0.201	0.183
SS Benefits Recipient Next Wave	Baseline Frailty	0.126	0.093	0.097	0.142	0.114	0.137
	PCA Baseline Frailty	0.127	0.092	0.097	0.142	0.114	0.137
	Augmented Frailty	0.125	0.095	0.095	0.143	0.117	0.139
	PCA Augmented Frailty	0.126	0.095	0.095	0.144	0.122	0.140
NH Entry Next Wave	Baseline Frailty	0.321	0.233	0.212	0.302	0.272	0.237
	PCA Baseline Frailty	0.320	0.230	0.211	0.304	0.268	0.238
	Augmented Frailty	0.290	0.209	0.181	0.277	0.242	0.175
	PCA Augmented Frailty	0.287	0.204	0.179	0.276	0.236	0.173
Death Next Wave	Baseline Frailty	0.270	0.220	0.192	0.239	0.239	0.174
	PCA Baseline Frailty	0.265	0.216	0.187	0.236	0.233	0.172
	Augmented Frailty	0.204	0.145	0.116	0.201	0.196	0.123
	PCA Augmented Frailty	0.197	0.143	0.109	0.196	0.191	0.119

4 How Large are Health Disparities?

Given that frailty is the single most predictive measure of health and has a quantitative interpretation, we use it to study health inequality. However, because frailty likely understates the health deficits for some of the groups we consider, we also document the inequality revealed by our potential frailty measure.

4.1 How Unequal is Frailty?

Frailty is a crucial indicator of an individual's health and resilience. But does the burden of frailty differ across racial and ethnic groups? To explore this, we turn to Figure 5, where Panels (a) and (b) report average frailty levels for men and women, respectively.

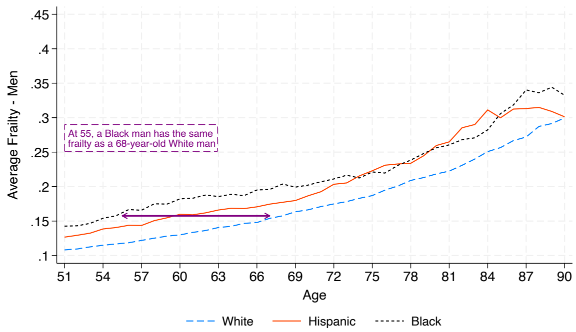
The data suggest a clear pattern: on average, White men and women experience lower levels of frailty compared to Black and Hispanic men and women. For example, a 55-year-old Black man typically exhibits a level of frailty similar to that of a Hispanic man who is 5 years older (age 60) and a White man who is 13 years older (age 68). Similarly, a 55-year-old Black woman tends to show frailty comparable to a Hispanic woman who is 6 years older (age 61) and a White woman who is 20 years older (age 75). These disparities persist throughout life but tend to narrow as individuals age, primarily because sicker individuals, particularly men, tend to have shorter lifespans.

Since frailty is constructed using 35 deficits, we can go from frailty to the number of one's health deficits by multiplying one's frailty by 35. For instance, 55-year-old Black women have, on average, one more health deficit compared to White women of the same age. As Panel (a) of Figure 2 shows, the most prevalent deficits for White women between 55 and 59 are having ever smoked, being diagnosed with arthritis and high blood pressure, and having difficulties kneeling and climbing several flights of stairs. Beyond these five deficits, Black women are also affected by obesity. Moreover, the most common deficits that White and Black women share tend to be more prevalent for Black women. Similarly, 55-year-old

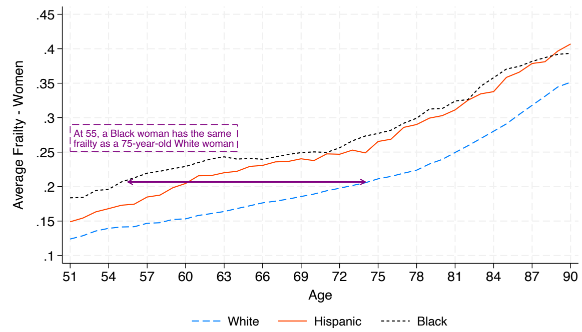
Black men, on average, have over two more health deficits compared to White men of the same age. In particular, as shown in Panel (c) of Figure 2, the four most prevalent deficits for White men between 55 and 59 are having ever smoked, being diagnosed with high blood pressure and arthritis, and obesity. These four deficits are also among the most common for Black men, but Black men in this age group also report having difficulty kneeling and climbing several flights of stairs. Here, too, the four most common deficits that Black and White men share tend to be more prevalent for Black men. These findings align with those of Carey, Miller, and Molitor (2024), who show that Black Americans are unhealthier than their White and Hispanic counterparts.

Panels (c) and (d) illustrate the percentage of individuals without frailty or health deficits by race and ethnicity, with men on the left and women on the right. Panel (c) shows that, up to approximately age 75, White men exhibit the highest proportion of individuals free from health deficits. For instance, at age 55, the share of White men with no health deficits stands at 8.9%, which is one and a half times greater than that of Black men (6.0%) and 0.5 percentage points higher than Hispanic men (8.4%). Beyond age 75, these proportions tend to converge across racial and ethnic lines, partly due to the impact of mortality. These patterns hold true for women as well. For example, at age 55, the share of White women without frailty is 8.1%, more than double that of Black women (2.6%) and 1.2 percentage points higher than Hispanic women (6.9%). Notably, disparities in women's average frailty persist for a longer period, continuing until around age 80.

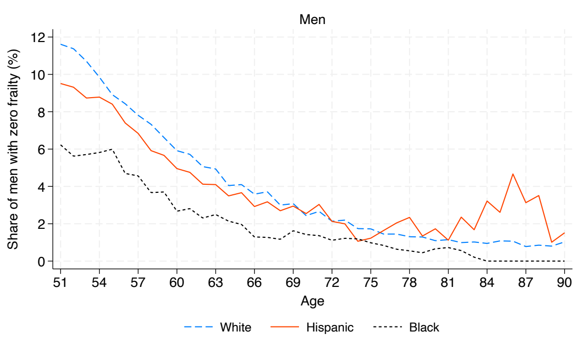
Panels (e) and (f) display the standard deviations of frailty for men and women by race and ethnicity. Before age 70, women tend to exhibit greater variability in frailty compared to men in all demographic groups. Interestingly, the standard deviations of frailty are relatively similar between Black and Hispanic individuals despite the differences in their average frailty levels. This suggests that Black individuals not only have a higher proportion experiencing positive frailty (higher averages) but also exhibit a wider range of frailty levels within their group (higher standard deviation). Additionally, it is worth noting that the standard deviation of frailty tends to decrease with age. This trend can be attributed to two factors: the impact of mortality, as those with higher frailty levels are more likely to pass away, and the inherent construction of frailty, which has an upper limit of one, causing frailty levels to converge as individuals age.



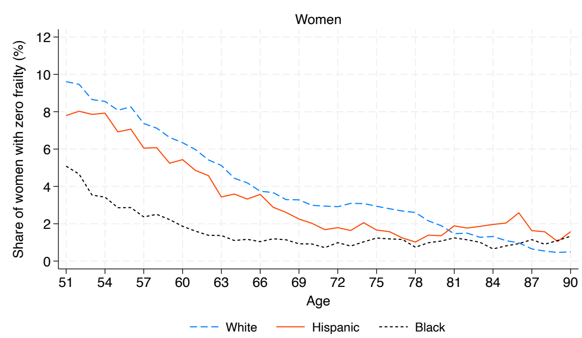
(a) Average frailty. Men



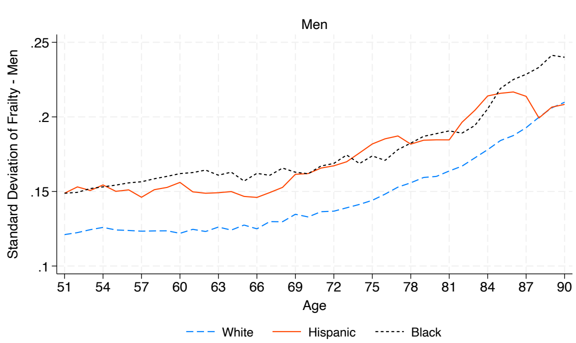
(b) Average frailty. Women



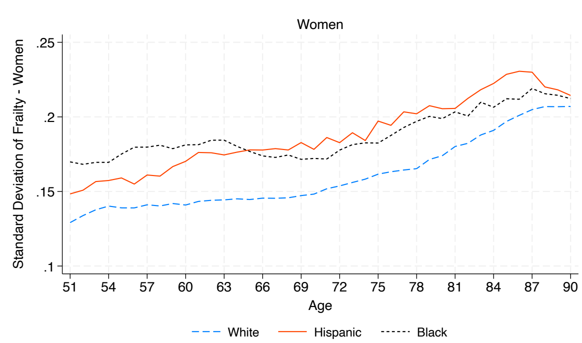
(c) Share with zero frailty. Men



(d) Share with zero frailty. Women



(e) Standard deviation of frailty. Men



(f) Standard deviation of frailty. Women

Figure 5: Average frailty, share with zero frailty, and standard deviation of frailty by age. Men (left) and women (right). Each statistic is smoothed using a three-year moving average.

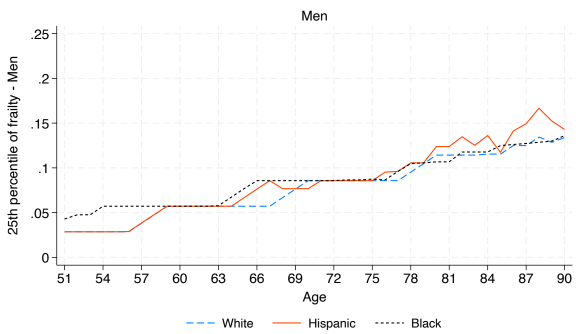
Figure 6 presents data on the 25th and 75th percentiles of frailty, categorized by age, race, and gender. Starting with men, Panels (a) and (c) show that differences in frailty levels among the healthiest individuals (those in the 25th percentile of frailty) are relatively modest across various racial and ethnic groups. At this frailty percentile, 60-year-old men from White, Hispanic, and Black backgrounds all experience fewer than two health deficits. However, as frailty levels increase, these disparities become more pronounced. Notably, Black men in the 75th percentile of frailty exhibit higher levels of frailty compared to their White and Hispanic counterparts at the same percentile. For instance, 60-year-old Black men in the 75th percentile of frailty have 9.3 health deficits, compared to 7.6 deficits for Hispanic men and 6.0 deficits for White men at the same frailty percentile.

Turning to women, Panels (b) and (d) reveal more substantial disparities by race and ethnicity across all percentiles. In general, White women experience fewer deficits. For example, at age 60, White and Hispanic women at the 25th frailty percentile experience approximately two health deficits, whereas Black women face 3.0 health deficits. The contrast is even more pronounced at the 75th frailty percentile, with figures standing at 7.4, 10.4, and 11.8 health deficits for White, Hispanic, and Black women, respectively.

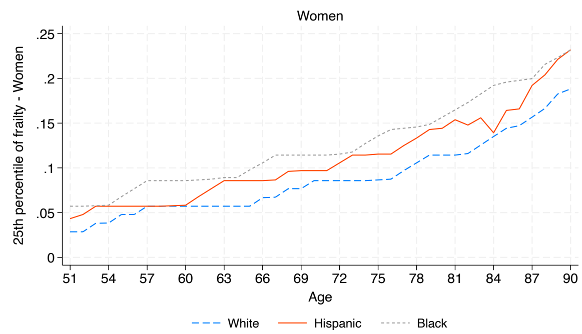
As argued by many others (including Alesina, Ferroni, and Stantcheva (2021)), racial gaps are pervasive. Our results emphasize that these disparities extend beyond educational attainment and direct measures of economic well-being such as wages (as shown, for instance, by Borjas and Katz (2007)) or earnings (as documented by Kondo et al. (2024)) and that they encompass many facets of health. Health, in turn, is not only important per se but also affects many other economic outcomes.

4.2 How Unequal is Potential Frailty?

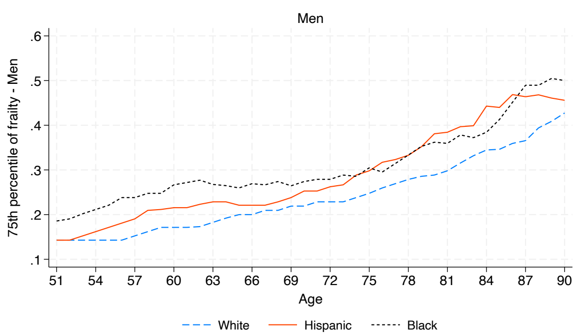
In Section 2.3, we highlight how correcting for under-reporting of diagnosed conditions affects diagnosis prevalence in the Black and Hispanic populations. Here, we compare average frailty



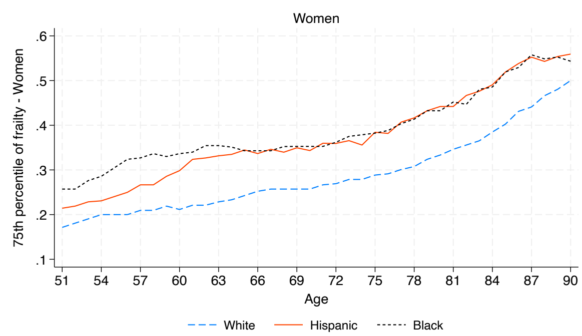
(a) 25th percentile of frailty. Men



(b) 25th percentile of frailty. Women



(c) 75th percentile of frailty. Men



(d) 75th percentile of frailty. Women

Figure 6: 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.

and potential frailty by race, gender, and ethnicity to gauge the extent to which potential frailty amplifies measured health inequality.

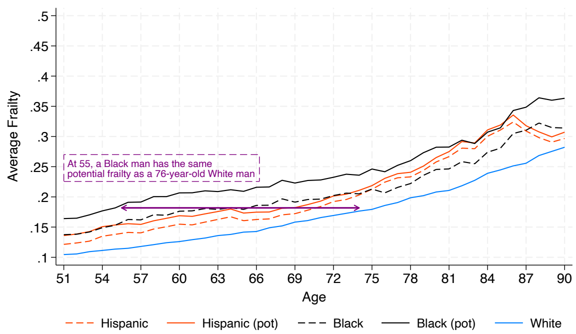
Panels (a) and (b) in Figure 7 show that potential frailty is consistently higher than frailty for Black and Hispanic men and women. Specifically, between the ages of 51 and 90, average potential frailty exceeds average frailty by 15.8%, 12.1%, 6.0%, and 4.2% for Black men, Black women, Hispanic men, and Hispanic women, respectively.

Consequently, potential frailty increases differences in biological age, defined as the age at which a non-White person has the same frailty as a White person. For example, a 55-year-old Black woman has the same frailty as a 75-year-old Hispanic woman (a 20-year gap) and the same potential frailty as an 80-year-old White woman (a 25-year gap). Similarly, a

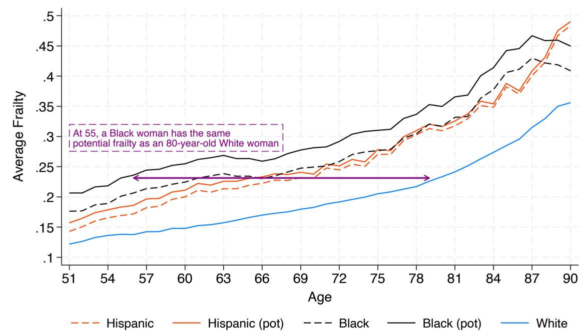
55-year-old Black man has the same frailty as a 69-year-old White man (a 14-year gap) and the same potential frailty as a 76-year-old White man (a 21-year gap). Therefore, compared to frailty, potential frailty increases biological age by 5 and 7 years for Black women and men, respectively. We observe a similar pattern for Hispanic people. For instance, a 55-year-old Hispanic woman has the same frailty as a 66-year-old White woman (an 11-year gap) and the same potential frailty as a 68-year-old White woman (a 13-year gap). Likewise, a 55-year-old Hispanic man has the same frailty as a 65-year-old White man (a 10-year gap) and the same potential frailty as a 68-year-old White man (a 13-year gap). Thus, compared to frailty, potential frailty increases the gaps in biological age by 2 and 3 years for Hispanic women and men, respectively, compared to their White counterparts.[7]

Next, Panel (c) shows that the differences between average frailty and potential frailty are greater for Black individuals than for Hispanic individuals. The percentage change between observed and potential frailty is particularly high for Black men of all ages and lowest for Hispanic women. For example, at age 55, the percentage change for Black men (18.0%) is almost double that of Hispanic men of the same age (9.9%). This panel also highlights that these gaps decrease with age, a trend more pronounced among Hispanic individuals. Specifically, the percentage changes between frailty and potential frailty at age 51 are 9.4% for Hispanic women and 11.4% for Hispanic men. By age 90, these differences decrease to 1.3% and 3.5%, respectively. This trend may reflect that older individuals in these groups are more likely to receive diagnoses or that those more likely to receive diagnoses are more likely to live longer (or both).

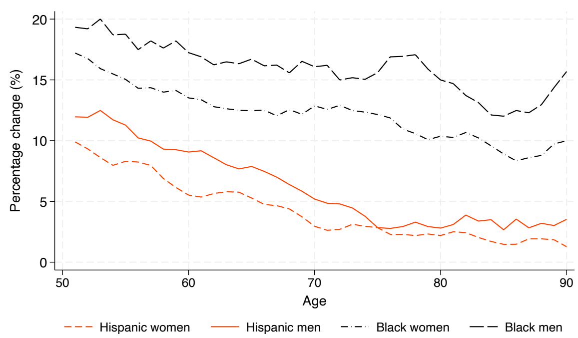
7. The biological age numbers presented here differ slightly from those in Section 4.1 because they are computed using the sub-sample of observations with 35 observed deficits.



(a) Average frailty. Men



(b) Average frailty. Women



(c) Percentage change

Figure 7: 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).

5 Quantifying the Effects of Health Inequality

Next, we measure the extent to which initial health differences at age 55 affect life expectancy and the duration of disability, retirement, and nursing home residency. We do so by estimating a statistical model that captures the dynamic evolution of health, mortality, and our economic outcomes of interest. We start by estimating how health and these outcomes change over time. Next, we simulate these outcomes to create simulated histories. Then, we simulate these outcomes by assigning Black and Hispanic men and women the same initial health distribution at age 55 as White men and women. Appendix D provides additional details. For tractability and ease of interpretation, in this part of the paper, we discretize frailty into five quintiles and label each category as excellent, very good, good, fair, and poor health, which are also the possible responses for self-reported health.

Despite the downward bias in measures of racial inequality introduced by under-diagnosis, we focus on frailty rather than potential frailty here because our imputation procedure is ill-suited to dynamic analysis at the individual level. Our nearest neighbor imputation approach, which compares the health deficits of two individuals at any point in time, is flexible enough to address bias in cross-sectional analysis but does not ensure dynamically consistent imputation at the individual level. As a result, using potential frailty in individual time series would reduce the persistence of health over time. Given that we model the evolution of health flexibly by race and gender, we consider the mismeasurement of frailty a smaller issue than the introduction of artificial volatility.[8]

Just like a structural model makes assumptions about causality and timing, so does our dynamic system. Figure 8 describes how we restrict the dynamic and contemporaneous

8. In principle, our imputation procedure could be adapted to use observed histories of deficits rather than the point-in-time information we currently rely on. However, our unbalanced panel makes this less than ideal. For example, if the ideal donor at a given age is a White individual with a shorter lifespan than the Black or Hispanic individual we are interested in, problems arise due to left-censored health histories from the HRS sampling criteria. Avoiding donors with different lifespans would force the algorithm to select donors with different health deficits, reducing the quality of matches. Alternatively, assigning deficits to deceased individuals, consistent with typical frailty modeling, would assume they accumulate all deficits, altering matches and introducing spurious diagnoses. For instance, if a White donor never had lung disease while alive, imputing this diagnosis posthumously would introduce significant biases into the dynamic analysis.

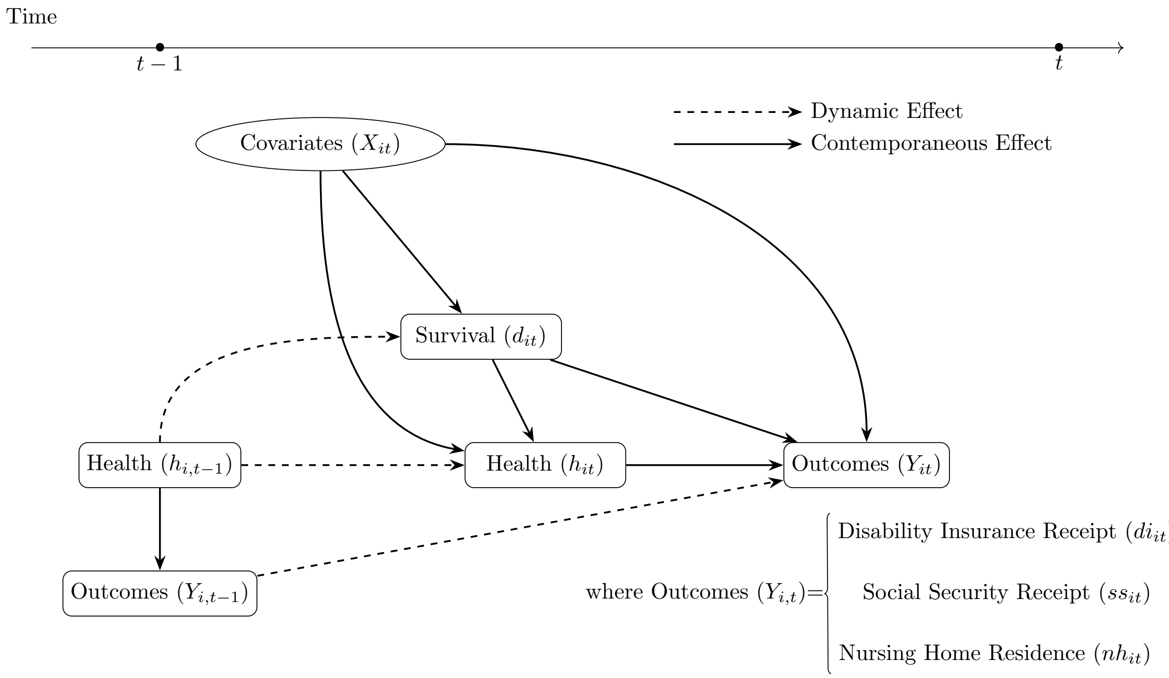


Figure 8: A Dynamic Model of Survival, Health Evolution and Economic Outcomes

relationships between the outcomes we model. For example, last period's health directly affects survival and, conditional on surviving, the probability of transitioning to better or worse health today. However, previous health has no direct effect on disability insurance or Social Security benefit receipt, or nursing home residency. We assume that only *current* health directly affects these outcomes.[9] Nevertheless, our model generates rich correlations between previous health and these outcomes through three indirect channels: (1) the dynamic effect on health today, (2) the impact on previous outcomes and their dynamic effects, and (3) common covariates such as race, gender, and education over time.

Our specification allows race and gender to directly affect the probability of survival, health transitions, and economic outcomes. Additionally, we allow them to have differential effects by current health. These features capture two important forces that may generate

9. Additionally, we allow last period's disability insurance receipt, Social Security receipt, and nursing home residency to have direct effects. These effects are modeled differently depending on the outcome, for example, to capture that Social Security receipt is an absorbing state. We provide full details in Appendix D.

inequality. First, they capture (potentially optimal) differences in the choices of individuals, such as those with longer expected lifespans retiring later. Second, they capture structural barriers that might lead to different outcomes across groups even if agents make the same choices. For example, the leniency of disability insurance screening may differ by gender (Low and Pistaferri 2019) and race, or certain groups may be systematically less likely to find gainful employment even when searching for work.

We also allow transitions to depend on education, marital status and whether an individual is covered by health insurance. These factors may have direct effects on health transitions that mediate the effects of race. In addition, they allow us to capture broad differences in the occupations of individuals that affect future health and the necessary health capital to continue working as well as the non-wage amenity value of employment. Our results (see Appendix D) suggest this is important for the timing of social security claiming and disability insurance receipt.

Our statistical approach yields a flexible model that incorporates many factors and allows for rich dynamic relationships between our outcomes of interest. Although we do not specify a full model of inter-temporal decision-making, we view our estimated transitions as both capturing the evolution of exogenous state variables and approximating the decision rules that relate choices to the state variables and that arise in a structural model based on lifetime utility maximization. As we do not estimate a fully specified structural model and recover the "deep" parameters governing preferences, we assume that those factors remain fixed when simulating our counterfactual. In fact, our counterfactuals only change agents' initial conditions. This change does not entail re-optimization, and hence, our approximations to their endogenous decision rules still hold. Thus, our results are consistent with a large class of structural models where forward-looking agents endogenously make health investments, choose whether to apply for disability, claim social security, and enter a nursing home.

5.1 Marginal Effects

How important is health compared to other observables? This section briefly discusses the marginal effects computed from our dynamic system. Appendix E reports the corresponding tables.

Current health has the largest impact on **future health**. For example, compared to someone in "excellent" health, someone in poor health is 75.6 percentage points less likely to be in "excellent" health within two years and 90.9 percentage points more likely to remain in "poor" health. The second-largest factor affecting future health is being Black: all else equal, Black individuals are 1.3 percentage points less likely than White individuals to be in "excellent" health and 0.2 percentage points more likely to be in "poor" health. Not being legally married has significant effects on all possible future health realizations. For instance, being single reduces the likelihood of "excellent" health by 0.7 percentage points and raises the likelihood of "poor" health by 0.5 percentage points. Being one year older decreases the probability of "excellent" or "very good" health (by 0.2 and 0.05 percentage points, respectively) and increases the probability of "fair" or "poor" health (by 0.05 and 0.2 percentage points). Education, instead, has positive effects, with each additional year raising the likelihood of "excellent" or "very good" health (by 0.3 and 0.01 percentage points) and lowering the likelihood of worse health realizations. Hispanic individuals are 0.3 percentage points less likely to be in "excellent" health compared to White individuals, and men are 0.2 percentage points less likely than women to be in "poor" health. Interestingly, health insurance has no significant effects on future health.

Current health is the observable with the largest impact on the likelihood of dying. Compared to someone in "excellent" health, the increase in the **probability of death** ranges from 1.1 percentage points for someone in "very good" health to 13.7 percentage points for someone in "poor" health. The second largest contributor to the probability of death is gender, with being a man increasing the probability of dying next wave by 3.8 percentage points. Gender is followed by being Hispanic, which decreases one's death probability by

1.5 percentage points, while being single rather than married increases it by 1.2 percentage points. One additional year of age increases the probability of death by 0.3 percentage points. Interestingly, being Black, more educated, and having health insurance coverage have no effects.

Current health is also the most important determinant of disability benefits recipiency. Compared to someone in "excellent" health, the increase in the **probability of receiving disability benefits** ranges from 2.2 percentage points for someone in "very good" health to 14.5 percentage points for someone in "poor" health. Next, gender is the second-largest determinant of disability recipiency, with men being 1.7 percentage points more likely to receive benefits than women. The third most important contributor is being Hispanic, which lowers it by one percentage point. Compared to married people, single people are 0.7 percentage points more likely to receive disability benefits. Having health insurance, being Black, and being older has a small positive effect, ranging from 0.6 to 0.04 percentage points. Finally, one additional year of education has almost no effect on this probability.

Age is the most important determinant of the **probability of receiving retirement benefits**. Being one year older increases this probability by 7.4 percentage points. The second most important determinant is having health insurance coverage, which lowers the probability of receiving retirement benefits by 5.8 percentage points. The third most important contributor is being Hispanic, which lowers this probability by 4.3 percentage points. Being between one to two years from the full retirement age increases the probability of receiving retirement benefits by an additional 3.3 percentage points. In turn, compared to someone in "excellent" health, being in good health increases the probability of receiving retirement benefits by 2.9 percentage points. Being a man, Black, and having an additional year of education lower the probability of receiving retirement benefits by 2.5, 2.1, and 1.6 percentage points, respectively.

Having lived in a nursing home in the past one or two years is the largest determinant of the **probability of living in a nursing home**, increasing this probability by 5.9 percentage points. The second largest contributor is being in "poor" rather than "excellent" health, which increases the probability of living in a nursing home by 4.5 percentage points. Third, being single increases the probability by 0.9 percentage points. In turn, being Hispanic reduces this probability by 0.8 percentage points, while being Black reduces it by 0.4 percentage points. Finally, being a man and being one year older increase the probability of living in a nursing home by 0.3 and 0.1 percentage points, respectively.

5.2 Does Inequality in Health at Age 55 Affect Future Outcomes?

We now turn to examining the extent to which the worse health at age 55 of Black and Hispanic individuals explains the gap in their later outcomes compared with those of White individuals.

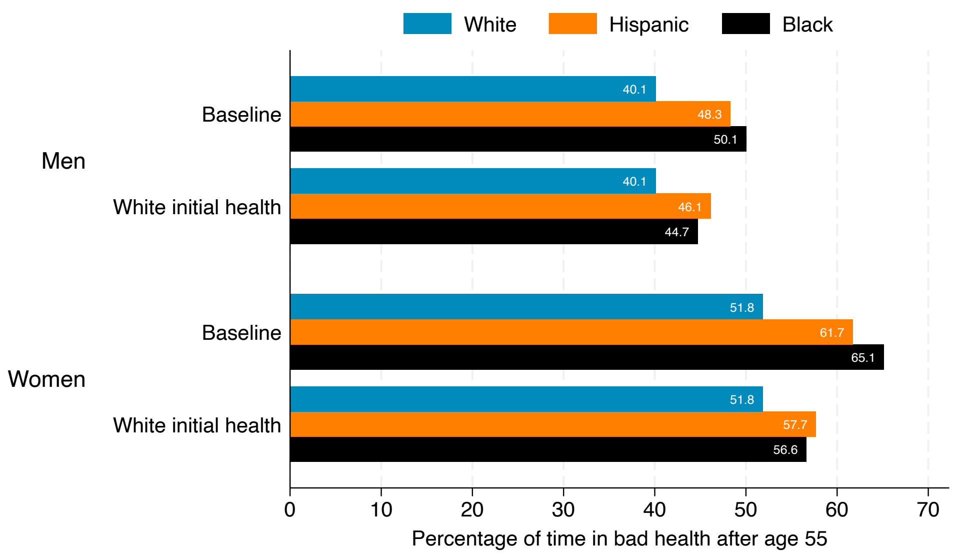


Figure 9: Average fraction of remaining life spent in bad health starting from age 55. This is computed as the fraction of remaining life spent in one of the two lowest health states ("poor" and "fair" health, or frailty quintiles), conditional on remaining alive

Figure 9 shows the average fraction of one's remaining life spent in bad health ("poor" and "fair" health states). The "Baseline" line reveals that women spend more of their remaining lives in bad health than men (40.1% for White men and 51.8% for White women) and that

Hispanic men and women spend 8.2 and 9.9 percentage points more time in bad health, respectively, than their White counterparts. For Black men and women, these figures are 10.0 and 13.34 percentage points higher, respectively.

Next, we perform a counterfactual simulation in which we assign Black and Hispanic individuals the initial health of White individuals at age 55. The effects are substantial and highlight that frailty at age 55 explains a large portion of the disparities in time spent in bad health. Specifically, for Hispanic individuals compared with White individuals, initial health accounts for 26.8% of the gap for men and 40% of the gap for women. For Black individuals, it accounts for 54% of the gap for men, and for 64% for women. To the extent that health proxies an individual's quality of life, this highlights large disparities in the quality of remaining life by race.

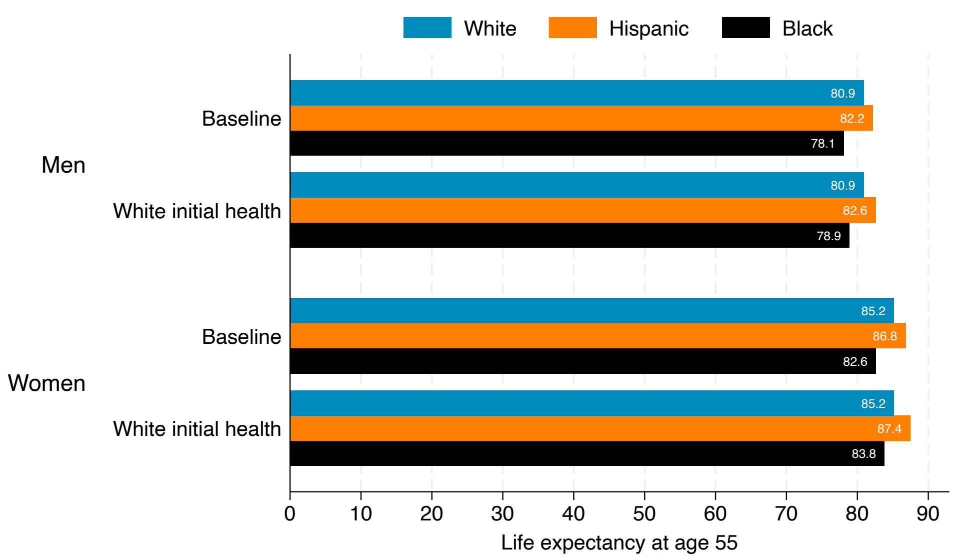


Figure 10: Average life expectancy as of age 55

Figure 10 reports simulated life expectancy at age 55. Hispanic men and women have the longest life expectancy, while Black individuals have the shortest. This result aligns with life expectancy at birth findings by Costa (2015). Figure 10 also shows that women of all races and ethnicities have a higher life expectancy than men, which is consistent with the results, among others, of Goldin and Lleras-Muney (2019). The observation that Hispanic individuals are in worse health but live longer is known as the "Hispanic health paradox," a phenomenon documented in the medical literature by Fernandez, García-Pérez, and Orozco-Aleman (2023), Cortes-Bergoderi et al. 2013, and Markides and Coreil (1986). Equalizing initial health increases the life expectancy of both Hispanic and Black people and would close the gap between Black and White people by 28.6% for men and 46% for women. It is worthwhile noticing that gaps in life expectancy between White and Black people remain despite the decrease in mortality for Black people documented between 1990 and 2010 by Currie and Schwandt (2016). Interestingly, Meara, Richards, and Cutler (2008) documents that the decrease in mortality among Black people is concentrated among the most educated.

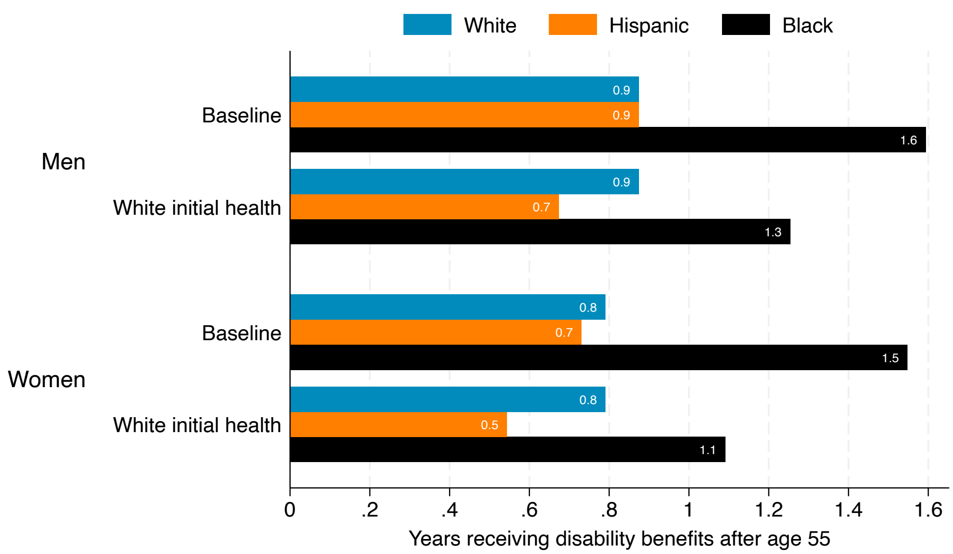


Figure 11: Average number of years receiving disability benefits after age 55

Figure 11 reports the years spent receiving disability benefits after age 55. In the baseline, Black men and women spend the most years receiving disability benefits, while Hispanic and White people spend similar amounts of time. Specifically, Black men and women spend almost twice as long (1.6 years) receiving disability benefits than White and Hispanic individuals. Equalizing initial health at age 55 would close 43% of the gap between Black and White men and 57% of the gap between Black and White women.

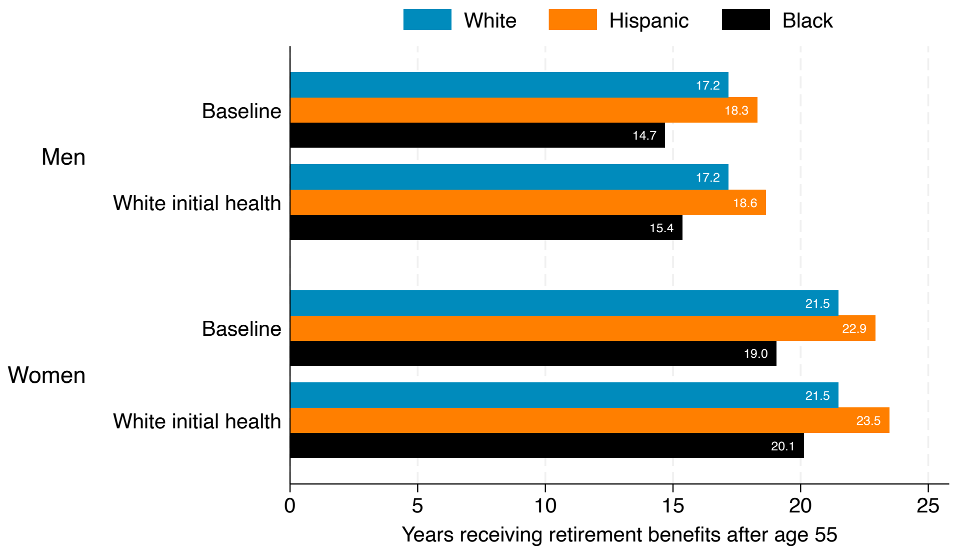


Figure 12: Average number of years receiving Social Security retirement benefits after age 55

Figure 12 shows that in our simulations, Hispanic individuals spend the longest time claiming Social Security benefits, while Black individuals the shortest. Specifically, Hispanic men and women receive retirement benefits for 1.1 and 1.4 years longer than White men and women, respectively. In contrast, Black men and women receive retirement benefits for 2.5 years less than their White counterparts. Equalizing initial health would significantly reduce the inequality in the length of receipt of retirement benefits between Black and White individuals. The gap between Black and White men would decrease by 28%, while the one between Black and White women would decrease by 44%.

Figure 13 displays the number of working years after age 55, defined as years not receiving Social Security or disability benefits. Hispanic individuals work between 1.2 and 3.6 months (0.1 and 0.3 years) longer than White individuals and over one year longer than Black individuals. Equalizing initial health to that of White individuals increases the number of working years. This increase ranges from about two months (0.2 years) for Hispanic men to almost four months (0.3 years) for Black women. Notably, the effects of equalizing initial health are comparable to or larger than many Social Security reforms. For example, French

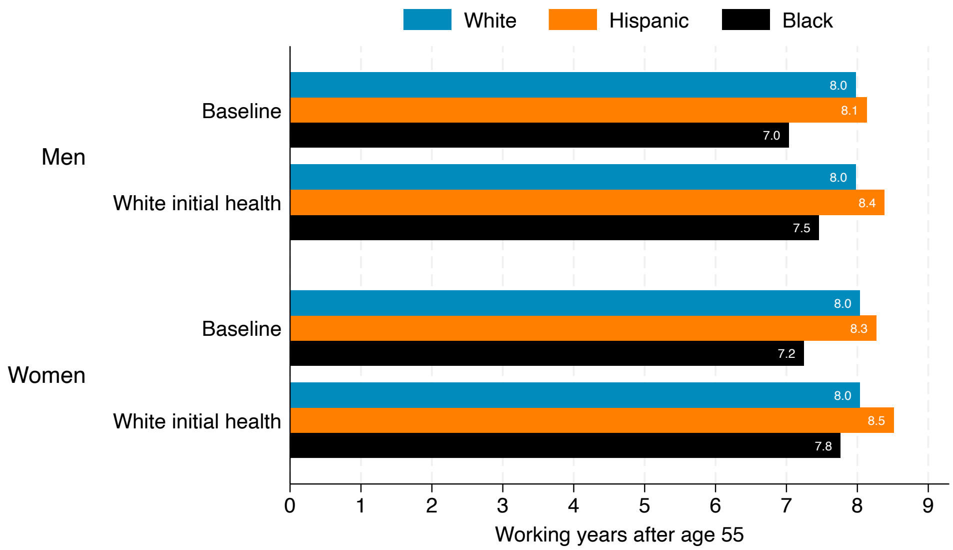


Figure 13: Average number of working years after age 55. Working years are defined as years not receiving Social Security or disability benefits.

(2005) finds that reducing Social Security benefits by 20% leads to an increase of 0.23 working years for men. This is close to the effect of equalizing the health of Hispanic men to that of White men but is less than half the effect of equalizing health for Black men. Overall, health inequality at age 55 explains about half of the differences between Black and White individuals. Our results are consistent with those of Blundell, Britton, Dias, French, and Zou (2022), who show that racial differences in health are a major determinant of differences in employment across races. Moreover, our results suggest that the worse labor market outcomes experienced by Black individuals, such as higher unemployment rates and lower labor force participation described by Boulware and Kuttner (2024), may also be due to differences in health.

Figure 14 shows the number of years spent in a nursing home after age 55. In our baseline simulations, White men and women spend the most time in a nursing home, while Hispanic individuals spend the least, despite having worse health and a longer life expectancy. Specifically, White men and women spend 0.6 and 1.2 years in a nursing home, respectively, while Hispanic men and women spend 0.4 and 0.9 years, respectively. Consistent with

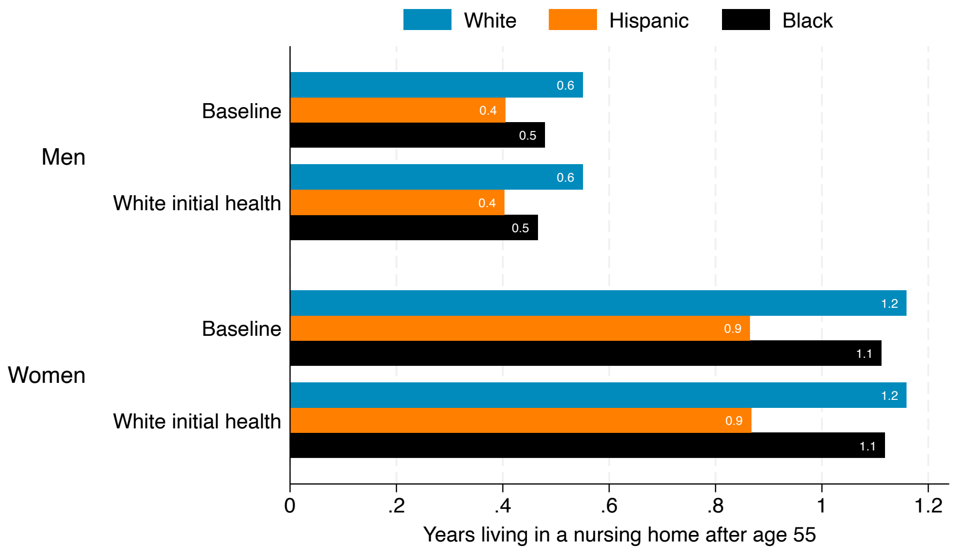


Figure 14: Average number of years in a nursing home after age 55

women's longer life expectancy, women of all races and ethnicities spend more years in a nursing home than men. Equalizing initial health does not change the time spent in a nursing home for any group. This is likely because people typically enter a nursing home at around age 84 (Lam et al. (2023)), and, by then, health at age 55 is no longer an important determinant of nursing home residency. 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.

Overall, our simulation results show that assigning 55-year-old non-White people the frailty of their White counterparts vastly reduces gaps in our outcomes of interest. Moreover, Andrews and Logan (2010) shows that racial health gaps are an important determinant of gaps in educational attainment. Therefore, if policies to reduce health gaps were available, they could also reduce gaps in other important economic outcomes. An example of such a policy is the Moving to Opportunity program in the USA, which offered people the opportunity to move to lower-poverty neighborhoods and, as shown by Sanbonmatsu et al. (2012), resulted in beneficial effects on mental (through lower depression and reduced levels of psychological distress) and physical health (thanks to lower obesity rates).

6 Conclusions and Directions for Future Research

Our paper tackles three questions: first, how to best measure health by race, ethnicity, and gender; second, how health is distributed among these groups of Americans; and third, how health inequality affects inequality in key economic outcomes.

We answer the first question by constructing several measures of health and evaluating their ability to predict many key economic outcomes. Our main conclusion from this part is that the baseline version of frailty proposed by the medical literature, which weighs all deficits equally, outperforms self-reported health status and measures of frailty which use more sophisticated weighting schemes, such as PCA.

Answering our second question reveals substantial health inequality by race, ethnicity, and gender. For instance, at age 55, the fraction of completely healthy women (with zero frailty) is 8.1% for White ones, 6.9% for Hispanic ones, and 2.6% for Black ones. Moreover, at age 55, Black men and women have frailty levels, or a biological age, comparable to White men and women who are 13 and 20 years older, respectively. The corresponding gaps for our measure of potential frailty are even larger: that is, 20 and 25 years.

In the last part of our paper, we address our third question and show that health inequality in middle age is a crucial determinant of economic inequality. For instance, assigning 55-year-old Black people the frailty of their White counterparts would halve the Black-White life expectancy gap. Health inequality at 55 is also a crucial determinant of the overall time spent in bad health, and removing racial disparities at 55 vastly reduces the gaps between White and non-White people. Similarly, eliminating health inequality also reduces the gaps in time spent claiming disability and retirement benefits.

Our findings underscore the importance of understanding health formation before age 55. Big contributors to health formation are environmental factors (such as pollution), access to healthcare, and key behaviors such as diet, exercise, and healthy habits (not smoking and drinking excessively). As of now, the literature has not provided a definitive answer on the relative importance of these factors and the time in life when they are most productive.

References

Alesina, Alberto, Matteo F Ferroni, and Stefanie Stantcheva. 2021. *Perceptions of Racial Gaps, their Causes, and Ways to Reduce Them.* Working Paper, Working Paper Series 29245. National Bureau of Economic Research, September.

Almeida, Joanna, Beth E. Molnar, Ichiro Kawachi, and S.V. Subramanian. 2009. "Ethnicity and nativity status as determinants of perceived social support: Testing the concept of familism." *Social Science & Medicine* 68 (10): 1852–1858.

Alsan, Marcella, Owen Garrick, and Grant Graziani. 2019. "Does Diversity Matter for Health? Experimental Evidence from Oakland." *American Economic Review* 109, no. 12 (December): 4071–4111.

Alsan, Marcella, and Marianne Wanamaker. 2017. "Tuskegee and the Health of Black Men." *The Quarterly Journal of Economics* 133, no. 1 (August): 407–455.

Altonji, Joseph G, Anthony A Smith Jr, and Ivan Vidangos. 2013. "Modeling earnings dynamics." *Econometrica* 81 (4): 1395–1454.

Andrews, Rodney J., and Trevon D. Logan. 2010. "Family Health, Children's Own Health, and Test Score Gaps." *American Economic Review: Papers & Proceedings* 100, no. 2 (May): 195–99.

Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme. 2017. "Earnings and consumption dynamics: a nonlinear panel data framework." *Econometrica* 85 (3): 693–734.

Attanasio, Orazio, Sagiri Kitao, and Giovanni L Violante. 2010. "Financing Medicare: A general equilibrium analysis." In *Demography and the Economy,* 333–366. University of Chicago Press.

Barnett, Michael L, Ellen Meara, Terri Lewinson, Brianna Hardy, Deanna Chyn, Moraa Onsando, Haiden A Huskamp, Ateev Mehrotra, and Nancy E Morden. 2023. "Racial inequality in receipt of medications for opioid use disorder." *New England Journal of Medicine* 388 (19): 1779–1789.

Beck, Audrey N, Brian K Finch, Shih-Fan Lin, Robert A Hummer, and Ryan K Masters. 2014. "Racial disparities in self-rated health: Trends, explanatory factors, and the changing role of socio-demographics." *Social Science & Medicine* 104:163–177.

Bernstein, Steven L, and Benjamin A Toll. 2019. "Ask about smoking, not quitting: a chronic disease approach to assessing and treating tobacco use." *Addiction Science & Clinical Practice* 14 (1): 1–4.

Blundell, R., E. French, and G. Tetlow. 2016. "Chapter 8 - Retirement Incentives and Labor Supply," edited by John Piggott and Alan Woodland, 1:457–566. Handbook of the Economics of Population Aging. North-Holland.

Blundell, Richard, Jack Britton, Monica Costa Dias, and Eric French. 2023. "The impact of health on labor supply near retirement." *Journal of Human Resources* 58 (1): 282–334.

Blundell, Richard, Jack Britton, Monica Costa Dias, Eric French, and Weijian Zou. 2022. *The Dynamic Effects of Health on the Employment of Older Workers: Impacts by Gender, Country, and Race.* MRDRC Working Paper MRDRC WP 2022-451. Ann Arbor, MI: University of Michigan Retirement and Disability Research Center (MRDRC). https://mrdrc.isr.umich.edu/publications/papers/pdf/wp451.pdf.

Borjas, George J, and Lawrence F Katz. 2007. "The evolution of the Mexican-born workforce in the United States." In *Mexican immigration to the United States,* 13–56. University of Chicago Press.

Boulware, Karl David, and Kenneth N. Kuttner. 2024. "What Explains Black Employment Dynamics?" *AEA Papers and Proceedings* 114:191–95.

Boulware, L. Ebony, Lisa A. Cooper, Lloyd E. Ratner, Thomas A. LaVeist, and Neil R. Powe. 2003. "Race and Trust in the Health Care System." *Public Health Reports* 118 (4): 358–365.

Carey, Colleen, Nolan Miller, and David Molitor. 2024. *Racial and Ethnic Disparities in SSDI Entry and Health.* Working Paper, Center Paper Series NB23-04. National Bureau of Economic Research, July.

Cobb, Ryon J., Courtney S. Thomas, Whitney N. Laster Pirtle, and William A. Darity. 2016. "Self-identified race, socially assigned skin tone, and adult physiological dysregulation: Assessing multiple dimensions of "race" in health disparities research." *SSM - Population Health* 2:595–602. ISSN: 2352-8273. https://doi.org/https://doi.org/10.1016/j.ssmph.2016.06.007. https://www.sciencedirect.com/science/article/pii/S2352827316300362.

Cook, Benjamin Lê, and Willard G Manning. 2009. "Measuring racial/ethnic disparities across the distribution of health care expenditures." *Health services research* 44 (5p1): 1603–1621.

Cortes-Bergoderi, Mery, Kashish Goel, Mohammad Hassan Murad, Thomas Allison, Virend K Somers, Patricia J. Erwin, Ondrej Sochor, and Francisco Lopez-Jimenez. 2013. "Cardiovascular mortality in Hispanics compared to non-Hispanic whites: A systematic review and meta-analysis of the Hispanic paradox." *European Journal of Internal Medicine* 24 (8): 791–799.

Costa, Dora L. 2015. "Health and the Economy in the United States from 1750 to the Present." *Journal of Economic Literature* 53 (3): 503–70.

Currie, Janet, and Hannes Schwandt. 2016. "Mortality Inequality: The Good News from a County-Level Approach." *Journal of Economic Perspectives* 30, no. 2 (May): 29–52. https://doi.org/10.1257/jep.30.2.29.

Danesh, Kaveh, Jonathan T Kolstad, Johannes Spinnewijn, and William D Parker. 2024. *The Chronic Disease Index: Analyzing Health Inequalities Over the Lifecycle.* Working Paper, Working Paper Series 32577. National Bureau of Economic Research, June.

Darden, Michael E, and Mario Macis. 2024. *Trust and Health Care-Seeking Behavior.* Working Paper, Working Paper Series 32028. National Bureau of Economic Research, January. https://doi.org/10.3386/w32028.

De Nardi, Mariacristina, Eric French, and John Bailey Jones. 2009. "Life Expectancy and Old Age Savings." *American Economic Review* 99, no. 2 (May): 110–15. https://doi.org/10.1257/aer.99.2.110. https://www.aeaweb.org/articles?id=10.1257/aer.99.2.110.

Dieleman, Joseph L., Carina Chen, Sawyer W. Crosby, Angela Liu, Darrah McCracken, Ian A. Pollock, Maitreyi Sahu, et al. 2021. "US Health Care Spending by Race and Ethnicity, 2002-2016." *JAMA* 326, no. 7 (August): 649–659. ISSN: 0098-7484.

Du Bois, William Edward Burghardt. 1899. *The Philadelphia negro: A social study.* Shocken Books; New York.

Eli, Shari J., Trevon D. Logan, and Boriana Miloucheva. 2023. "The Enduring Effects of Racial Discrimination on Income and Health." *Journal of Economic Literature* 61, no. 3 (September): 924–40.

Fang, Ling, Huafang Sheng, Yingying Tan, and Qi Zhang. 2023. "Prevalence of diabetes in the USA from the perspective of demographic characteristics, physical indicators and living habits based on NHANES 2009-2018." *Frontiers in Endocrinology* 14:1088882.

Fernandez, José, Mónica García-Pérez, and Sandra Orozco-Aleman. 2023. "Unraveling the Hispanic Health Paradox." *Journal of Economic Perspectives* 37, no. 1 (February): 145–68. https://doi.org/10.1257/jep.37.1.145.

French, Eric. 2005. “The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour.” *The Review of Economic Studies* 72, no. 2 (April): 395–427. ISSN: 0034-6527. https://doi.org/10.1111/j.1467-937X.2005.00337.x.

French, Eric, and John Bailey Jones. 2017. “Health, Health Insurance, and Retirement: A Survey.” *Annual Review of Economics* 9 (1): 383–409. https://doi.org/10.1146/annurev-economics-063016-103616.

Geiger, H Jack. 2003. “Racial and ethnic disparities in diagnosis and treatment: a review of the evidence and a consideration of causes.” *Unequal treatment: Confronting racial and ethnic disparities in health care* 417:1–38.

Goldin, Claudia, and Adriana Lleras-Muney. 2019. “XX>XY?: The changing female advantage in life expectancy.” *Journal of Health Economics* 67:102224.

Halliday, Timothy, Bhashkar Mazumder, and Ashley Wong. 2021. “Intergenerational mobility in self-reported health status in the US.” *Journal of Public Economics* 193:104307. ISSN: 0047-2727. https://doi.org/10.1016/j.jpubeco.2020.104307.

Hayes, Donald K, Sandra L Jackson, Yanfeng Li, Gregory Wozniak, Stavros Tsipas, Yuling Hong, Angela M Thompson-Paul, Hilary K Wall, Cathleen Gillespie, Brent M Egan, et al. 2022. “Blood pressure control among non-Hispanic Black adults is lower than non-Hispanic white adults despite similar treatment with antihypertensive medication: NHANES 2013–2018.” *American journal of hypertension* 35 (6): 514–525.

Hill, Latoya, Samantha Artiga, and Sweta Haldar. 2022. “Key facts on health and health care by race and ethnicity.” *Kaiser Family Foundation* 26.

Hosseini, Roozbeh, Karen Kopecky, and Kai Zhao. 2022. “The evolution of health over the life cycle.” *Review of Economic Dynamics* 45:237–263. ISSN: 1094-2025. https://doi.org/10.1016/j.red.2021.07.001.

HRS Staff. 2017. "Sample Sizes and Response Rates" (April). https://hrs.isr.umich.edu/sites/default/files/biblio/ResponseRates_2017.pdf.

Katz, S. 1983. "Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living." *Journal of the American Geriatrics Society* 31, no. 12 (December): 721–727.

Katz, Sidney, Thomas D. Downs, Helen R. Cash, and Robert C. Grotz. 1970. "Progress in Development of the Index of ADL." *The Gerontologist* 10, no. 1_Part_1 (March): 20–30. ISSN: 0016-9013. https://doi.org/10.1093/geront/10.1_Part_1.20.

Kim, Eun Ji, Taekyu Kim, Joseph Conigliaro, Jane M Liebschutz, Michael K Paasche-Orlow, and Amresh D Hanchate. 2018. "Racial and ethnic disparities in diagnosis of chronic medical conditions in the USA." *Journal of general internal medicine* 33:1116–1123.

Kondo, Illenin O, Kevin Rinz, Natalie Gubbay, Brandon Hawkins, John L Voorheis, and Abigail K Wozniak. 2024. *Granular Income Inequality and Mobility using IDDA: Exploring Patterns across Race and Ethnicity.* Working Paper, Working Paper Series 32709. National Bureau of Economic Research, July.

Lam, Kenneth, Irena Cenzer, Cari R Levy, Dan D. Matlock, Alexander K. Smith, and Kenneth E. Covinsky. 2023. "The Natural History of Disability and Caregiving Before and After Long-Term Care Entry." *JAMA Internal Medicine* 183, no. 12 (December): 1295–1303.

Lawton, M Powell, and Elaine M Brody. 1969. "Assessment of older people: self-maintaining and instrumental activities of daily living." *The gerontologist* 9 (3_Part_1): 179–186.

Lin, Pei-Jung, Allan T Daly, Natalia Olchanski, Joshua T Cohen, Peter J Neumann, Jessica D Faul, Howard M Fillit, and Karen M Freund. 2021. "Dementia diagnosis disparities by race and ethnicity." *Medical care* 59 (8): 679.

Low, Hamish, and Luigi Pistaferri. 2019. *Disability Insurance: Error Rates and Gender Differences.* Working Paper, Working Paper Series 26513. National Bureau of Economic Research, November. https://doi.org/10.3386/w26513.

Markides, Kyriakos S, and Jeannine Coreil. 1986. "The health of Hispanics in the southwestern United States: an epidemiologic paradox." *Public health reports* 101 (3): 253.

McFadden, Daniel. 1977. "Quantitative Methods for Analyzing Travel Behaviour of Individuals: Some Recent Developments." *Cowles Foundation Discussion Papers,* no. 707.

McWilliams, J Michael, Ellen Meara, Alan M Zaslavsky, and John Z Ayanian. 2009. "Differences in control of cardiovascular disease and diabetes by race, ethnicity, and education: US trends from 1999 to 2006 and effects of medicare coverage." *Annals of internal medicine* 150 (8): 505–515.

Meara, Ellen R., Seth Richards, and David M. Cutler. 2008. "The Gap Gets Bigger: Changes In Mortality And Life Expectancy, By Education, 1981–2000." PMID: 18332489, *Health Affairs* 27 (2): 350–360. https://doi.org/10.1377/hlthaff.27.2.350.

Meyer, Bruce D., Nikolas Mittag, and Robert M. Goerge. 2022. "Errors in Survey Reporting and Imputation and Their Effects on Estimates of Food Stamp Program Participation." *Journal of Human Resources* 57 (5): 1605–1644.

Morden, Nancy E, Deanna Chyn, Andrew Wood, and Ellen Meara. 2021. "Racial inequality in prescription opioid receipt—role of individual health systems." *New England Journal of Medicine* 385 (4): 342–351.

Ofstedal, Mary B., and David R. Weir. 2011. "Recruitment and Retention of Minority Participants in the Health and Retirement Study." *The Gerontologist* 51, no. suppl_1 (June): S8–S20.

Peek, Monica E, Algernon Cargill, and Elbert S Huang. 2007. "Diabetes health disparities." *Medical care research and review* 64 (5_suppl): 101S–156S.

Petersen, Ruth, Liping Pan, and Heidi M Blanck. 2019. "Peer reviewed: Racial and ethnic disparities in adult obesity in the United States: CDC's tracking to inform state and local action." *Preventing chronic disease* 16. https://doi.org/10.5888/pcd16.180579.

Poterba, James M., Steven F. Venti, and David A. Wise. 2017. "The asset cost of poor health." *The Journal of the Economics of Ageing* 9:172–184.

Rockwood, Kenneth, Xiaowei Song, and Arnold Mitnitski. 2011. "Changes in relative fitness and frailty across the adult lifespan: evidence from the Canadian National Population Health Survey." *Cmaj* 183 (8): E487–E494.

Sanbonmatsu, Lisa, Nicholas A. Potter, Emma Adam, Greg J. Duncan, Lawrence F. Katz, Ronald C. Kessler, Jens Ludwig, et al. 2012. "The Long-Term Effects of Moving to Opportunity on Adult Health and Economic Self-Sufficiency." *Cityscape* 14 (2): 109–136. ISSN: 1936007X.

Schroeder, Heather, David Weir, and Brady West. 2023. *Representativeness of the oversamples of Blacks and Hispanics in the Health and Retirement Study: A geographic analysis of 2020.* Technical report. University of Michigan.

Searle, Samuel D, Arnold Mitnitski, Evelyne A Gahbauer, Thomas M Gill, and Kenneth Rockwood. 2008. "A standard procedure for creating a frailty index." *BMC geriatrics* 8 (1): 1–10. https://doi.org/10.1186/1471-2318-8-24.

Skinner, Jonathan, and Steven J. Atlas. 2010. "6. Education and the Prevalence of Pain." In *Research Findings in the Economics of Aging,* edited by David A. Wise, 145–166. Chicago: University of Chicago Press. ISBN: 9780226903088. https://doi.org/doi:10.7208/9780226903088-008. https://doi.org/10.7208/9780226903088-008.

Spalter-Roth, Roberta M, Terri Ann Lowenthal, and Mercedes Rubio. 2005. *Race, ethnicity, and the health of Americans.* ASA.

Williams, David R., and Michelle Sternthal. 2010. "Understanding Racial-ethnic Disparities in Health: Sociological Contributions." *Journal of Health and Social Behavior* 51 (1_suppl): S15–S27.

Wu, Anna H, V Wendy Setiawan, Daniel O Stram, Eileen M Crimmins, Chiu-Chen Tseng, Unhee Lim, Song-Yi Park, Kami K White, Iona Cheng, Christopher A Haiman, et al. 2023. "Racial, ethnic, and socioeconomic differences in a deficit accumulation frailty index in the Multiethnic Cohort Study." *The Journals of Gerontology: Series A* 78 (7): 1246–1257.

APPENDICES FOR ONLINE PUBLICATION

A The Data

We use the RAND HRS Longitudinal File 2018 (V2), which covers the years between 1992 and 2018. Table A-1 describes our sample selection. Our initial sample consists of 264,620 observations for all 14 waves in the HRS. Because we do not observe key health variables until wave 3, we drop observations before the third wave. Then, we restrict our attention to respondents aged 51 to 100. This leaves us with a sample of 222,552 observations. Finally, we drop all observations that report a race or ethnicity other than White, Black, or Hispanic. Our final sample consists of 216,166 individual-year observations.

Table A-1: Sample Selection

Sample	Selected out	Selected in
Initial Sample		264,620
Waves 3 - 14	32,294	232,326
Age between 51 and 100	9,774	222,552
White, Black, and Hispanic Responders	6,386	216,166

Table A-2 shows our sample breakdown by race, ethnicity, and gender in 5-year age bins. It shows that the majority of respondents for each age are White women. This happens because, at younger ages, respondents' younger wives tend to be more numerous, and at older ages because men tend to die faster. The last row of the table also shows that Black and Hispanic respondents tend to be younger than their White counterparts by 5 and 7 years, respectively.

A.1 Candidate Deficit Variables and Their Inclusion

Tables A-3 and A-4 list the 118 health deficits present in the RAND HRS data set, grouped by category, and specify those we do not include in our baseline measure of frailty, as well as the reason for it. The first column shows the name of the variable in the dataset. The

Table A-2: Sample Composition by 5-year age bins

	White		Hispanic		Black		
	Men	Women	Men	Women	Men	Women	All
Age 51-54	4,620	7,231	1,292	1,907	1,524	2,698	19,272
	0.24	0.38	0.07	0.10	0.08	0.14	1.00
Age 55-59	10,572	13,098	2,463	3,111	3,096	4,796	37,136
	0.28	0.35	0.07	0.08	0.08	0.13	1.00
Age 60-64	11,068	13,494	2,092	2,738	2,796	4,426	36,614
	0.30	0.37	0.06	0.07	0.08	0.12	1.00
Age 65-69	10,576	12,731	1,510	1,948	2,157	3,298	32,220
	0.33	0.40	0.05	0.06	0.07	0.10	1.00
Age 70-74	10,195	12,566	1,174	1,438	1,656	2,514	29,543
	0.35	0.43	0.04	0.05	0.06	0.09	1.00
Age 75-79	8,908	11,421	928	1,196	1,304	2,115	25,872
	0.34	0.44	0.04	0.05	0.05	0.08	1.00
Age 80-84	6,136	8,851	515	796	818	1,460	18,576
	0.33	0.48	0.03	0.04	0.04	0.08	1.00
Age 85-89	3,360	5,644	222	467	400	848	10,941
	0.31	0.52	0.02	0.04	0.04	0.08	1.00
Age 90-94	1,226	2,626	95	217	139	388	4,691
	0.26	0.56	0.02	0.05	0.03	0.08	1.00
Age 95-100	232	795	22	69	31	152	1,301
	0.18	0.61	0.02	0.05	0.02	0.12	1.00
Total	66,893	88,457	10,313	13,887	13,921	22,695	216,166
	0.31	0.41	0.05	0.06	0.06	0.10	1
Individuals	11,361	13,994	2,119	2,628	2,953	4,291	37,346
Average birth year	1937	1936	1943	1943	1942	1942	1938

Notes: The first row denotes the number of observations, while the second one displays their share in that age bin. The last two rows display the number of individuals and the average birth year for each demographic group. The last column shows the total by row.

second letter `w` in each variable name is a placeholder for the corresponding HRS wave. For instance, `r3shlt` denotes the self-reported health status variable in the third wave of the HRS. The second column provides a brief description of the variable, while the third column indicates the range of values each variable can take. The fourth column summarizes our reason for elimination when we eliminate that variable.

To establish whether a health deficit should be included in our frailty index, we evaluate candidate deficits along the following dimensions

1. Whether they meet the five criteria outlined in Searle, Mitnitski, Gahbauer, Gill, and Rockwood (2008):

   (a) The candidate deficit must be related to health status.

   (b) The prevalence of the candidate deficit must generally increase with age.

   (c) The candidate deficit must not saturate too early.

(d) The total set of deficits must cover a range of systems in the body.

(e) If used for comparisons over time, the set of deficits used to construct the frailty index must remain the same.

2. Whether the question related to the deficits has been asked to everyone in every wave.

3. Whether the share of missing values makes the candidate deficit unusable.

Incomplete Variables. Forty-three variables are either not asked consistently between waves 3 and 14 or only asked to a subsample of respondents. We highlight them in yellow in Tables A-3 and A-4. Incomplete variables include several cognition-related deficits, which are only asked about in proxy interviews. We include four of them in our augmented frailty. Namely, we include `rwalone`, `rwhaluc`, `rwhaluc`, and `rwlost`. We do not include the other incomplete cognition variables because several of them are not binary (like `rwdlrc` and `rwser7`) or they are only asked of respondents older than 65 (like `rwcact` and `rwpres`).

Substantial Missing Values Variables. Twenty variables have too many missing values to be usable (between 7% and 45%). A common rule of thumb in the medical and gerontology literature is not to use deficits with more than 5% of missing values when constructing frailty (see Rockwood, Song, and Mitnitski (2011)). Among the twenty variables with an excessive number of missing values, nine are related to depression. We include eight of them in our augmented measure of frailty. In particular, we include `rwdepres`, `rweffort`,`rwsleepr`, `1-rwhappy`, `rwflone`, `rwfsad`, `rwgoing`, and `1-rwenlife`. We do not include the summary mental health score `rwcesd` because it is constructed using fewer variables than the standard CESD score used by clinicians and because there is no clear threshold to establish a risk for clinical depression. We also include the cognition-related variable `rwbwc20` in our augmented frailty index. In particular, we recode this variable so that it takes the value 1 (has the deficit) for everyone who gets the backward count wrong and 0 (does not have the deficit) for everyone who gets the backward count right. We highlight the variables we exclude because of missing values in red in Tables A-3 and A-4.

Vague Variables. Seven variables are vague in the sense that the related questions lack the necessary information to establish whether these variables denote a health deficit. These variables are highlighted in blue in Tables A-3 and A-4. The variable `rwdrugs` reports whether the respondent regularly takes their prescribed medication. However, it does not report (1) The type of medication, (2) Whether the respondent has been prescribed any medication. Without this information, we cannot verify that this variable meets the criteria of Searle, Mitnitski, Gahbauer, Gill, and Rockwood (2008), and thus, it should not be used to construct a frailty index. Similarly, `rwoutpt` does not report the type of outpatient surgery undergone by the respondent, and `rwspcfac` does not specify which type of special facility (such as adult care centers, social work centers, rehabilitation facilities, and meals for the elderly or disabled) the respondent used. The variable `rwdentst` reports whether the respondent has seen a dentist in the previous two years. This variable includes routine checkups and cleaning, so it does not necessarily indicate worse health. Similarly, `rwdoctor` asks whether the respondent reports any doctor visit in the reference period. Doctor visits include annual physical exams and preventive screenings, which are not an indicator of worse health. The variable `rwjoga` reports any difficulty jogging one mile, which might be more related to one's athleticism rather than their overall health status. Finally, `rwhomcar` reports a wide range of home care services. These include, for instance, wound care for pressure sores or a surgical wound, patient and caregiver education, intravenous or nutrition therapy, injections, and monitoring serious illness and unstable health status. Therefore, it is unclear whether this variable meets the criteria of Searle, Mitnitski, Gahbauer, Gill, and Rockwood (2008).

Preventive Care Variables. Six variables refer to preventive care, which is not necessarily a signal of better or worse health. Therefore, these should not be considered deficits. They are highlighted in gray in Tables A-3 and A-4.

Unnecessary Variables. The variables reporting height (`rwheight`) and weight (`rwweight`) are unnecessary because we have a variable reporting BMI. They are highlighted in orange in Tables A-3 and A-4.

Additional Criteria and our frailty definition. In addition, we do not include unhealthy behaviors, that is, the variable related to current smoking (`rwsmoken`) and the three variables related to alcohol consumption (`rwdrink`, `rwdrinkd`, `rwdrinkn`) in our baseline frailty index. However, we include these variables in our augmented frailty. In particular, we include `rwsmoken` and combine `rwdrinkd` and `rwdrinkn` to create a deficit we label "heavy alcohol use."[10] We also include a deficit related to whether respondents are frequently troubled by pain in the augmented frailty index. This deficit is not available in the RAND HRS but is available in the raw HRS data. We exclude self-reported health status from both our frailty indices. Finally, we use BMI as a deficit by creating a binary variable equal to 1 when BMI is greater than 30 (the threshold for obesity). The variables we eliminate in this step are highlighted in purple in Tables A-3 and A-4. Our resulting baseline frailty index is made up of 35 deficits, while our augmented frailty index is made up of 51 deficits, which are summarized in Table A-5.

10. The National Institute on Alcohol Abuse and Alcoholism (NIAAA) defines "heavy alcohol use" as consuming more than five drinks a day or 15 drinks a week for men and more than four drinks a day or eight drinks a week for women. We use the variables `rWdrinkd` (number of days a week a respondent drinks) and `rWdrinkn` (number of drinks when the respondent drinks) to construct the deficit. In particular, we construct the average number of drinks per week by multiplying `rWdrinkd` and `rWdrinkn`. Then, we set the deficit equal to 1 (deficit) if the respondent is a man and drinks more than 15 drinks or if the respondent is a woman and drinks more than eight drinks per week.

Table A-3: Candidate deficits by category

Variable Name	Description	Values	Reason for elimination
	ADLs and physical limitations		
rwarmsa	Any difficulty reaching arms above shoulder level	binary	
rwbatha	Any difficulty bathing	binary	
rwbeda	Any difficulty getting in and out of bed	binary	
rwchaira	Any difficulty getting up from a chair after sitting for long periods	binary	
rwclim1a	Any difficulty climbing one flight of stairs without resting	binary	
rwclimsa	Any difficulty climbing several flights of stairs without resting	binary	
rwdimea	Any difficulty picking up a dime from the table	binary	
rwdressa	Any difficulty getting dressed	binary	
rweata	Any difficulty eating	binary	
rwlifta	Any difficulty lifting or carrying weights over 10 pounds	binary	
rwpusha	Any difficulty pushing or pulling large objects	binary	
rwsita	Any difficulty sitting for about two hours	binary	
rwstoopa	Any difficulty stooping, kneeling, or crouching	binary	
rwtoilta	Any difficulty using the toilet	binary	
rwwalk1a	Any difficulty walking one block	binary	
rwwalkra	Any diffulty walking across a room	binary	
rwwalksa	Any difficulty walking several blocks	binary	
rwjoga	Any difficulty jogging one mile	binary	vague question
	Alcohol and Smoking		
rwsmokev	Ever smoked	binary	
rwdrink	Ever drinks any alcohol	binary	additional
rwdrinkd	Number of days a week they drink	continuous	additional
rwdrinkn	How many drinks when they drink	continuous	additional
rwsmoken	Smoke now	binary	additional
	Cognition		
rwalone	Can be left alone for an hour or so	binary	incomplete
rwvocab	Vocabulary score	1-10 scale	incomplete
rwcact	Correctly name cactus	binary	incomplete
rwscis	Correctly name scissors	binary	incomplete
rwpres	Correctly name the president	binary	incomplete
rwvp	Correctly name the vice-president	binary	incomplete
rwhaluc	Ever sees or hears things that are not really there	binary	incomplete
rwwander	Ever wanders off and does not return on his or her own	binary	incomplete
rwlost	Gets lost in familiar environment	binary	incomplete
rwbwc20	Backwards count from 20	0-2 scale	missing values
rwcogtot	Summary score for word recall and mental status together	continuous	missing values
rwdw	Correct date - day of the week	binary	missing values
rwdy	Correct date - day	binary	missing values
rwmo	Correct date - month	binary	missing values
rwyr	Correct date - year	binary	missing values
rwmstot	Summary score for mental status	continuous	missing values
rwser7	Serial 7s test	continuous	missing values
rwtr20	Summary score for total word recall	continuous	missing values
rwdlrc	Delayed word recall	continuous	missing values
rwimrc	Immediate word recall	continuous	missing values
	Depression		
rwcesd	CESD score	continuous	missing values
rwdepres	Felt depressed much of the time in the week before the interview	binary	missing values
rweffort	Felt like everything is an effort much of the time in the week before the interview	binary	missing values
rwenlife	Enjoyed life much of the time in the week before the interview	binary	missing values
rwflone	Felt lonely much of the time in the week before the interview	binary	missing values
rwfsad	Felt sad much of the time in the week before the interview	binary	missing values
rwgoing	Could not get going much of the time in the week before the interview	binary	missing values
rwhappy	Was happy much of the time in the week before the interview	binary	missing values
rwsleepr	Sleept was restless much of the time in the week before the interview	binary	missing values

Notes: First column: name of the variable in the dataset. Second column: description of the variable. Third column: range of values each variable can take. Fourth column: reason for elimination.

Table A-4: Candidate deficits by category

Variable Name	Description	Values	Reason for elimination
	Diagnoses		
rwarthre	Arthritis or rheumatisms	binary	
rwcancre	Cancer or a malignant tumor of any kind except skin cancer	binary	
rwdiabe	Diabetes or high blood sugar	binary	
rwhearte	Heart attack, coronary heart disease, angina, congestive heart failure, or other heart problem	binary	
rwhibpe	High blood pressure	binary	
rwlunge	Chronic lung disease except asthma such as chronic bronchitis or emphysema	binary	
rwpsyche	Emotional, nervous, or psychiatric problems	binary	
rwstroke	Stroke	binary	
rwalzhee	Ever reported Alzheimer	binary	incomplete
rwmemrye	Ever reported memory-related disease	binary	incomplete
rwdemene	Ever reported dementia	binary	incomplete
rwsleepe	Sleep disorders	binary	incomplete
	Healthcare Utilization		
rwhosp	Hospital stay in the previous 2 years	binary	
rwnrshom	Nursing home stay in the previous 2 years	binary	
rwdentst	Dental visits in the previous 2 years	binary	vague question
rwdoctor	Doctor visit in the previous 2 years	binary	vague question
rwdrugs	Regular use of prescription drugs in the previous 2 years	binary	vague question
rwhomcar	Home health care in the previous 2 years	binary	vague question
rwoutpt	Outpatient surgery in the previous 2 years	binary	vague question
rwspcfac	Use of special facilities or services in the previous 2 years	binary	vague question
	IADLs		
rwmapa	Any difficulty using a map	binary	
rwmealsa	Any difficulty preparing meals	binary	
rwmedsa	Any difficulty taking medications	binary	
rwmoneya	Any difficulty managing money	binary	
rwphonea	Any difficulty using the phone	binary	
rwshopa	Any difficulty shopping for groceries	binary	
rwcalca	Any difficulty using a calculator	binary	incomplete
	Physical Measures		
rwbalful	Full tandem stand	continuous	incomplete
rwbalfulc	Whether made compensatory movements during full-tandem stand	binary	incomplete
rwbalfult	Held a full-tandem stand the max time applicable	binary	incomplete
rwbalsbs	Duration of side-by-side tandem	continuous	incomplete
rwbalsbsc	Whether made compensatory movements during side-by-side stand	binary	incomplete
rwbalsemi	Semi-tandem stand	continuous	incomplete
rwbalsemic	Shether made compensatory movements during semi-tandem stand	binary	incomplete
rwbpdia	Diastolic blood pressure	continuous	incomplete
rwbppos	Position during BP measure	1-3 scale	incomplete
rwbppuls	Pulse	continuous	incomplete
rwbpys	Systolic blood pressure	continuous	incomplete
rwgrp	Hand grip test	continuous	incomplete
rwgrpdom	Dominant heand	binary	incomplete
rwgrpl	Hand grip test - left hand	continuous	incomplete
rwgrppos	Position during hand grip test	1-3 scale	incomplete
rwgrpr	Hand grip test - right hand	continuous	incomplete
rwpmbmi	Measured BMI	continuous	incomplete
rwpmhght	Measured height in centimeters	continuous	incomplete
rwpmwaist	Measured waist	continuous	incomplete
rwpmwght	Measured weight in kilograms	continuous	incomplete
rwpuff	Breathing test	continuous	incomplete
rwpuffpos	Position during breathing test	1-3 scale	incomplete
rwtimwlk	Timed walk test time	continuous	incomplete
rwtimwlka	Timed walk test - walking aid used	binary	incomplete
	Preventive Care		
rwbreast	Monthly self-checks for breast lumps	binary	preventive
rwcholst	Blood test for cholesterol	binary	preventive
rwflusht	Flu shot	binary	preventive
rwmammog	Mammogram	binary	preventive
rwpapsm	Pap smear	binary	preventive
rwprost	Check for prostate cancer	binary	preventive
	Other self-reported measures		
rwbmi	Self-reported BMI	continuous	
rwshlt	Self-reported health status	1-5 scale	additional
rwweight	Self-reported weight in kilograms	continuous	unnecessary
rwheight	Self-reported height in meters	continuous	unnecessary
rwback	Back problems	binary	incomplete
rwvgactx	Frequency of vigorous physical activity	1-5 scale	incomplete
rwvigact	Whether performs vigourous physical activity more than 3 times a week	binary	incomplete
rwltactx	Frequency of light physical activity	1-5 scale	incomplete
rwmdactx	Frequency of moderate physical activity	1-5 scale	incomplete

Notes: First column: name of the variable in the dataset. Second column: description of the variable. Third column: range of values each variable can take. Fourth column: reason for elimination.

Table A-5: Deficits included in our frailty indices

Deficit	Deficit
Baseline frailty	
ADLs	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	***Diagnoses***
Difficulty walking one block	Diagnosed with high blood pressure
Difficulty walking several blocks	Diagnosed with diabetes
	Diagnosed with cancer
IADLs	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	***Healthcare Utilization***
	Has stayed in the hospital in the previous two years
Other Functional Limitations	Has stayed in a nursing home in the previous two years
Difficulty climbing one flight of stairs	
Difficulty climbing several flights of stairs	***Addictive Diseases***
Difficulty getting up from a chair	Has BMI larger than 30
Difficulty kneeling or crouching	Has ever smoked cigarettes
Augmented frailty	
Pain	***Cognition***
Frequently troubled by pain	Gets lost in familiar environment
	Wanders off
Mental health	Cannot be left alone
Felt depressed	Has hallucinations
Felt like everything was an effort	Cannot count backwards from 20
Had restless sleep	
Did not feel happy most of the time	***Harmful habits***
Felt alone	Smokes now
Felt sad	Heavy alcohol use
Could not get going	
Did not enjoy life	

A.2 Frailty Computation

When computing frailty, we allow for at most four missing deficits by observation and rescale the index accordingly. Table A-6 shows that doing so allows us to compute frailty for 99% of observations in our sample. We select this cutoff as it trades off the additional variability at the individual level introduced by including too few deficits with the reduction in variability due to maintaining a large sample. To construct potential frailty, we focus on the subsample of observations with 35 observed deficits, which consists of about 83% of our original sample.

Table A-6 reports the distribution of non-missing deficits in our sample. It shows that we observe a minimum of 12 deficits and that about 83% of observations report non-missing values for all 35 deficits we consider.

Table A-6: Distribution of non-missing deficits

	Frequency	Percentage	Cumulative Percentage
12	9	0.00	0.00
14	1	0.00	0.00
17	3	0.00	0.01
18	9	0.00	0.01
19	7	0.00	0.01
20	8	0.00	0.02
21	16	0.01	0.02
22	14	0.01	0.03
23	19	0.01	0.04
24	27	0.01	0.05
25	34	0.02	0.07
26	50	0.02	0.09
27	91	0.04	0.13
28	140	0.07	0.20
29	247	0.12	0.32
30	478	0.22	0.54
31	1,033	0.48	1.02
32	2,495	1.17	2.19
33	6,593	3.08	5.27
34	25,449	11.91	17.18
35	177020	82.82	100.00

B Correcting for Systematic Under-Diagnosis

The literature warns us that, for many reasons, diagnoses are differentially reported by race, ethnicity, and gender. Indeed, healthcare spending is higher for White people (Cook and

Manning (2009) and Dieleman et al. (2021)), the fraction of uninsured people is significantly higher for Hispanic and Black people than White people (Hill, Artiga, and Haldar (2022)), and non-White Americans have lower trust in the healthcare system (See Alsan and Wanamaker (2017), Boulware et al. (2003), and Darden and Macis (2024)). There is also evidence of underutilization of healthcare by minorities. Alsan, Garrick, and Graziani (2019) shows that the lack of diversity of healthcare professionals contributes to the underutilization of healthcare by minorities. Moreover, racial disparities in diagnosis and treatment are pervasive and have been present since the American Civil War (Eli, Logan, and Miloucheva (2023)). Black and Hispanic women are less likely to be seen for breast cancer screenings and are more likely to be seen for the first time when the cancer is too advanced and to undergo less aggressive treatment (Geiger (2003)). Furthermore, Kim et al. (2018) finds that Black and Hispanic people are more likely to be under-diagnosed with diabetes, even when controlling for differences in healthcare utilization. They also find that Black people are twice as likely to have undiagnosed kidney disease than White ones. Moreover, Lin et al. (2021) argues that Black and Hispanic people are more likely to have a missed or delayed diagnosis of dementia. Morden et al. (2021) finds that Black and Hispanic Americans are less likely to receive opioid analgesics than White ones despite no evidence of racial differences in pain perception or preferences for pain management. Barnett et al. (2023) also document that Black and Hispanic patients receive fewer medications to treat opioid use disorder overdose.

Spalter-Roth, Lowenthal, and Rubio (2005) reviews the sociology literature on racial health inequality and argues that systemic racism, together with socioeconomic inequalities and adverse conditions in segregated neighborhoods, is an important driver of health inequality by race and ethnicity. A long-standing interest in racial health inequality in sociology dates back to the seminal contribution of Du Bois (1899) (see Williams and Sternthal 2010 for a review). An early example of social epidemiology, Du Bois (1899) documented that Black men had worse health than Black women and that the gender differences in health were larger for Black people than for White people in Philadelphia's 7th Ward.

We also find (Section 2.1) that the majority of deficits are significantly less prevalent in the White subsample. The exceptions to this, however, are those deficits that relate to receiving a formal diagnosis for a medical condition. A potential concern with a deficit-based measure of health, such as frailty, is that differences in the reporting of deficits are driven by differences in reporting behavior or access to medical services instead of differences in the underlying latent health of individuals.

Our goal is to address potential differences in reporting due to differential patterns of diagnosis *conditional on true health*. To this end, we focus on the eight deficits that measure formal diagnoses and build an imputation procedure to construct hypothetical deficits at the individual level that are not subject to this concern.[11] Using these alternative individual deficits, we can then construct an alternative measure of frailty for each individual in our sample.

Let the vector $\mathbf{D}_{i,t}$ denote the observed deficits for an individual i in our sample in wave t. We can write this vector as

$$\mathbf{D}_{i,t} = (\mathbf{D}_{i,t}^{U}, \mathbf{D}_{i,t}^{B}), \tag{A1}$$

where $\mathbf{D}_{i,t}^{U}$ is the sub-vector of deficits that do not require a formal diagnosis and $\mathbf{D}_{i,t}^{B}$ is the sub-vector of deficits that do require a formal diagnosis. We assume there is no differential reporting of the twenty-seven deficits that do not require a formal diagnosis. Thus, they are unbiased reports, and we denote this sub-vector with superscript U. The remaining eight deficits may be subject to under-reporting bias, and we denote the sub-vector with superscript B. While we assume the sign of this bias, our procedure makes no assumptions on the magnitude of under-reporting for formal diagnoses. Instead, it allows us to infer this directly from the data.

The key assumption we make to impute hypothetical deficits is that the reported formal diagnoses for the insured White households are not contaminated by under-diagnosis

11. The eight "diagnosed deficits" are being diagnosed with high blood pressure, diabetes, cancer, lung disease, heart condition, a stroke, psychological or psychiatric problems, and arthritis.

and are unbiased. Alternatively, under the assumption that they are only *less* biased, our imputation procedure can be interpreted as assigning the under-diagnosis bias of insured White individuals to non-White individuals. We begin our imputation procedure by partitioning the data by gender and marital status. Then, given a gender-marital status pair (e.g., single women), we identify all non-White individuals. For each of these individuals in the partition, we identify their nearest neighbor in the insured-White sub-sample using the vector $\mathbf{D}^U_{i,t}$ as well as their age in years, education, and the survey wave which allows us to assign an insured White donor to each non-White household.[12] Under the assumption of under-reporting, we then construct a vector of imputed formal diagnoses $\widehat{\mathbf{D}}^B_{i,t}$ by replacing an individual's observed diagnoses with their donor's whenever their donor reports a diagnosis and the original non-White individual does not. Note that the deficits of White individuals are not changed by this imputation procedure. Figure A-1 provides a graphical summary of our imputation procedure.

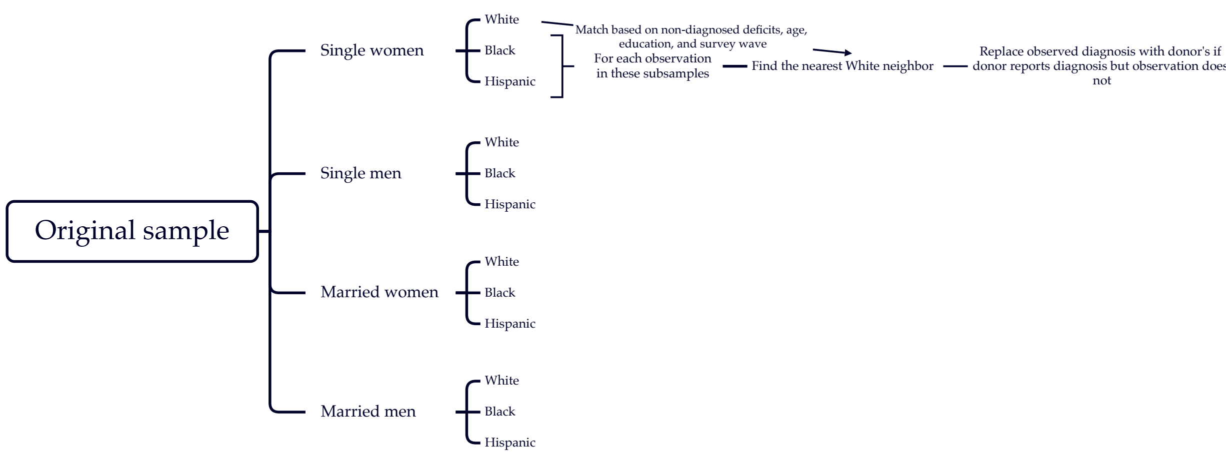


Figure A-1: Summary of our imputation procedure.

There are a number of advantages to our non-parametric imputation procedure. First, by imputing the entire vector of formal diagnoses, we use a multivariate imputation compared to an alternative item-wise imputation procedure. Thus, we are able to capture any arbitrary correlation between the reported formal diagnoses either due to biological factors or medical

12. We implement this using the `teffects nnmatch` command in Stata.

practices. For instance, those with a stroke diagnosis are more likely to have their blood pressure monitored. Second, our imputation allows for flexible correlation between specific health deficits and formal diagnoses, for example, functional limitations and arthritis. In a parametric model, this is only possible by introducing a large number of interaction terms which capture the effects of different combinations of health deficits or by imposing restrictions a priori. Our non-parametric approach captures this in a tractable way. Third, as we additionally match on age, our imputation procedure respects both the average deterioration of health as individuals age and the survivorship bias because potential White donors for non-White individuals at older ages will be healthier than those who are deceased. Fourth, although the one-time calculation of nearest neighbors is computationally intensive, we view this approach as intuitive and transparent.

Implicitly, we assume that health deficits in the non-White and insured-White populations encode the same information about the true latent health. While we make this assumption throughout this paper, our ability to correctly impute formal diagnoses will be hampered if, for example, the association between a cancer diagnosis and the answer to the question "Have you ever smoked" is different in the sample of insured-White and non-White individuals because the intensity or duration of smoking differed even conditional on having ever smoked. However, we believe this approach is preferable to imposing strict parametric assumptions. Finally, we only perform our imputation procedure on the sub-sample of individuals who have complete responses for all thirty-five deficits that we use in calculating frailty.

B.1 Imputation Validation

It is not possible to directly assess the goodness-of-fit of our imputed diagnoses because we do not have "true" diagnoses for our non-White sub-sample. However, because we assume that formal diagnoses for White individuals do not suffer from under-reporting, we can evaluate the predictive accuracy for White individuals. To do this, we duplicate our sample of White individuals and compute the nearest neighbor in our original donor pool. We can then compare their observed diagnoses with the diagnoses of their assigned donor. Note that, in this procedure, an individual's duplicated observation can be their own donor, which is consistent with the spirit of the validation exercise.

Table A-7: Imputation Correct Classification Rates

		Correct Classification Rate		
Diagnosis	White Prevalence	Overall	Has Condition	No Condition
High blood pressure	0.504	0.830	0.803	0.856
Diabetes or high blood sugar	0.172	0.936	0.782	0.968
Cancer	0.144	0.932	0.745	0.964
Chronic lung disease	0.091	0.975	0.859	0.987
Heart Condition	0.231	0.920	0.815	0.952
Stroke	0.073	0.977	0.854	0.987
Psych. problems	0.150	0.955	0.800	0.982
Arthritis	0.548	0.843	0.838	0.848

Table A-7 reports the overall prevalence of each of the diagnosed conditions in the White sub-sample as well as correct classification rates pooling across gender and marital status. This table shows that our imputation procedure has a high level of accuracy, above 80%, across all of the deficits that require formal diagnoses and an accuracy above 90% for three-quarters of the deficits. Furthermore, the conditional classification rates reveal that we achieve a high rate of accuracy irrespective of whether the individual has or does not have the deficit. Reassuringly, the conditional classification rate is higher for those who do not have the diagnoses. Thus, our procedure is conservative in the sense that it produces more false negatives than false positives.

B.2 Prevalence of potential deficits

Table A-8: Prevalence of potential diagnosed deficits for men and women aged 55 to 59

	Women			Men		
	Baseline	Potential	Pct. change	Baseline	Potential	Pct. change
High blood pressure						
White	0.348	0.348	-	0.421	0.421	-
Hispanic	0.449	0.538	19.9%	0.443	0.544	22.7%
Black	0.667	0.798	19.8%	0.606	0.774	27.7%
Diabetes						
White	0.109	0.109	-	0.130	0.130	-
Hispanic	0.260	0.317	21.8%	0.254	0.309	21.8%
Black	0.246	0.362	47.4%	0.254	0.374	47.5%
Cancer						
White	0.098	0.098	-	0.055	0.055	-
Hispanic	0.074	0.119	61.2%	0.030	0.058	91.9%
Black	0.067	0.177	165.5%	0.049	0.124	152.1%
Lung disease						
White	0.076	0.076	-	0.054	0.054	-
Hispanic	0.049	0.088	78.7%	0.029	0.065	126.4%
Black	0.079	0.181	129.9%	0.053	0.138	161.5%
Heart condition						
White	0.100	0.100	-	0.148	0.148	-
Hispanic	0.088	0.138	56.5%	0.110	0.175	58.9%
Black	0.152	0.268	76.7%	0.142	0.276	94.2%
Stroke						
White	0.028	0.028	-	0.032	0.032	-
Hispanic	0.031	0.052	66.7%	0.036	0.060	67.1%
Black	0.062	0.118	90.1%	0.072	0.136	87.9%
Psychological problems						
White	0.211	0.211	-	0.117	0.117	-
Hispanic	0.198	0.267	34.9%	0.108	0.158	45.7%
Black	0.178	0.376	111.5%	0.131	0.254	93.3%
Arthritis						
White	0.468	0.468	-	0.361	0.361	-
Hispanic	0.430	0.530	23.2%	0.269	0.370	37.7%
Black	0.515	0.728	41.3%	0.351	0.571	62.9%

Notes: The "Baseline" column reports the prevalence of observed deficits. The "Potential" column reports the prevalence of potential deficits. The "Pct. change" column displays the percentage change between the potential and observed prevalence.

C Details on our Empirical Strategy for Evaluating Health Measures

We start our empirical analysis by dividing our sample into six demographic groups: White, Black, and Hispanic men and women, and for each outcome, we select the appropriate age range to examine. That is, we include respondents of all ages (that is, between 51 and 100) for the outcomes of entering a nursing home, and dying. Instead, we restrict our attention to a narrower age range for receiving Social Security retirement benefits and disability insurance. In particular, we focus on respondents between the ages of 60 and 75 for receiving Social Security retirement benefits to account for the fact that one cannot claim Social Security benefits before age 62 and that few people retire after age 75. Moreover, because disability insurance converts into retirement benefits, once the recipients reach their full retirement age, we focus on respondents between age 51 and full retirement age for the disability insurance recipiency outcome. Appendix C.1.5 reports more details on the rules regarding disability insurance and the full retirement age.

Table A-9 describes our outcome variables and the values they take. Table A-10 summarizes the age ranges and regressors for each outcome.

Table A-9: Outcome variables

Variable	Description	Values
SDI Recipient Next Wave	In wave t, this variable tells us if the respondent will receive SDI in wave t+1	0 if does not receive SDI in t+1, and did not in t 1 if receives SDI in t+1, but did not in t missing if received SDI in t
Receiving Social Security Benefits Next Wave	In wave t, this variable tells us if the respondent will claim SS benefits in t+1 (ages 60 and older)	0 if no income from SS in t+1 and none in t 1 if positive income from SS in t+1 and none in t missing if claiming SS benefits in t
Nursing Home Entry Next Wave	In wave t, this variable tells us if the respondent will enter a nursing home in wave t+1	0 if does not live in a NH in t+1 and did not in t 1 if lives in a NH in t+1 but did not in t 1 if dies in a NH in t+1 but did not live in it in t missing if lived in a NH in t
Death Next Wave	In wave t, this variable tells us if the respondent will die in wave t+1	0 if alive in t+1 1 if dead in t+1 missing if dead in t

All of our specifications include some "basic" regressors: age (either as a third-order polynomial or age dummies), a second-order polynomial in years of education, and cohort

Table A-10: Age range and regressors other than health and basic regressors

Variable	Age Range	Regressors Other than Health and Basic
SDI Recipient Next Wave	51-FRA	3-order poly in age
Receiving SS Benefits Next Wave	60-75	Age dummies + FRA dummy
Nursing Home Entry Next Wave	51-100	3-order poly in age
Death Next Wave	51-100	3-order poly in age

Notes: Basic regressors include age, years of education, and cohort and marital status dummies. We also interact health with age, age squared, age cubed, and years of education. Age is rescaled as actual age minus 50. To ensure convergence of our logistic regressions, we drop the interactions of SRHS, age squared, and age cubed for SDI recipiency for Hispanic women and Nursing Home Entry for Hispanic men.

and marital status dummies. In some specifications, we then include one of our two health measures and its interactions with age, age squared, age cubed, and years of education. Finally, we include both measures of health and their interactions with age and education. To capture the age discontinuities provided by the Social Security system, we also add a dummy equal to 1 if the respondent is one or two years younger than his or her full retirement age.[13]

To evaluate which health measure is the most predictive one, we compute the McFadden's pseudo-R^2 (or pseudo-R^2) for each regression. It is given by one minus the ratio of the full-model log-likelihood and the intercept-only log-likelihood, that is

$$\text{Pseudo-R}^2 = 1 - \frac{LL(\text{Full Model})}{LL(\text{Intercept-Only Model})}.$$

Therefore, it is not a measure of the proportion of the variance of the dependent variable explained by the model (as in the case of the R^2 in an OLS regression). Instead, it measures the relative improvement in model fit when adding regressors to the intercept-only model. The pseudo-R^2 varies between 0 and 1, and higher values denote a better fit of the full model.

13. Potential frailty has no additional predictive power compared to our baseline frailty measure because we estimate our baseline specification separately by race. Consequently, our estimated coefficients already account for systematic racial differences in frailty and their correlation with frailty (although the interpretation of the coefficient differs). Therefore, we do not show the results for the predictive power of potential frailty.

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

C.1 Quantifiying the Effects on Economic Outcomes

What is the effect of health on economic outcomes, and does it vary by race and ethnicity? To answer this question, we use our estimated logistic regressions for each outcome to compute the average marginal effects and predicted probabilities by frailty, race, ethnicity, and gender.

Next, we report the average marginal effects in table format, computed as the average over the marginal effect for each observation in our sample, leaving all explanatory variables beyond the one of interest at their observed values. We also display graphs in which we compute the effect of frailty on a certain outcome by group. We do this by assigning that frailty value to all observations while leaving all other regressors at their observed values and report the average predicted probability by demographic group.[14] Our graphs report the marginal effect of frailty as a function of the average frailty associated with having between 1 and 19 health deficits. Over 95% of our sample reports at most 19 deficits.

C.1.1 Receiving Disability Insurance Benefits

Table A-11 reports the average marginal effects related to becoming an SDI recipient in the next wave. It shows that higher frailty has a statistically significant effect on the probability of receiving SDI. That is, one additional health deficit increases the probability of receiving disability benefits by 0.6 and 0.4 percentage points for men and women, respectively. Age, instead, does not have a significant effect and thus does not play an important role in driving the recipiency of disability benefits given the other variables that we condition on.

An additional year of education reduces the probability of receiving SDI, and more so for men (0.2 percentage points) than women (0.07 percentage points). Being a Hispanic

14. As discussed in Section 3, our regressions already account for systematic racial differences in frailty by interacting each regressor with race. Therefore, the marginal effects of baseline and potential frailty are the same, and we do not show the results for potential frailty here.

person rather than a White one also reduces this probability, and more so for men (0.8 percentage points) than women (0.5 percentage points). In contrast, being single increases the probability of receiving disability benefits: the probability of becoming an SDI recipient next wave for single men and women is 0.6 percentage points higher than that of married men and women, on average.

Table A-11: Receiving SDI next wave

	Men		Women	
Frailty	0.00563***	(0.000217)	0.00421***	(0.000148)
Black	0.00592**	(0.00285)	0.00470**	(0.00237)
Hispanic	-0.00803***	(0.00287)	-0.00449*	(0.00260)
Age	-0.0000449	(0.000407)	-0.000139	(0.000287)
Years of Education	-0.00162***	(0.000359)	-0.000661**	(0.000314)
Born 1950-1968	0.00218	(0.00217)	0.00137	(0.00165)
Partnered	-0.00161	(0.00343)	0.0112***	(0.00402)
Single	0.00572**	(0.00241)	0.00578***	(0.00169)

Notes: Marginal effects resulting from logistic regressions.

Figure A-2 displays the predicted probability of receiving SDI benefits next wave by the frailty associated with having between 1 and 19 health deficits. As one might expect, more unhealthy men and women are more likely to receive SDI. Looking at men (left panel) more in detail highlights that, for levels of frailty between 0.03 and 0.26, Black men are more likely to receive SDI benefits, but there are no significant differences at higher levels of frailty. Looking at women (right panel) shows that Black and White women tend to have a higher probability of being on disability compared to Hispanic women, especially for frailty higher than 0.43 (15 deficits).

C.1.2 Receiving Social Security Benefits

Table A-12 shows the marginal effects on the probability of becoming a Social Security benefits recipient next wave. Starting from frailty, having worse health (i.e., higher frailty) increases the probability of retiring for men but not for women. More specifically, one additional health deficit increases the probability of retiring by 0.4 percentage points for

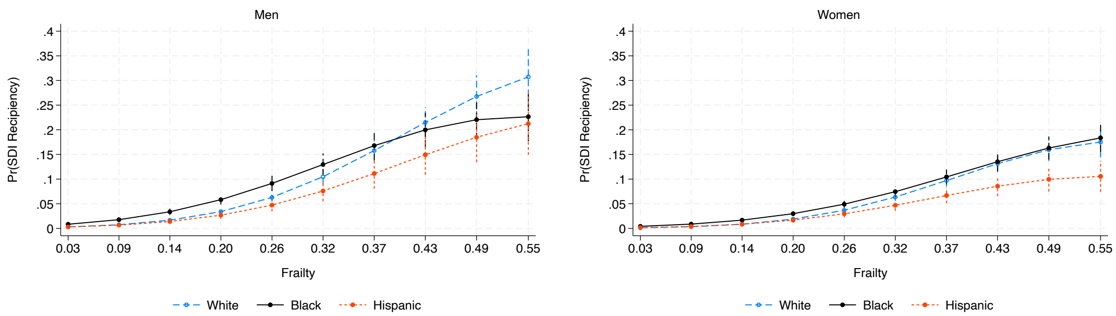


Figure A-2: Predicted probabilities of becoming an SDI recipient next wave by frailty. Men (left panel) and women (right panel). The frailty values reported in the horizontal axis correspond to 1 to 19 conditions. The vertical lines mark the 95% confidence interval.

men (left column), on average. The point estimate for women, instead, is much smaller and not statistically significant. Years of education reduce the probability of retiring for both men and women, with the effect being larger for women (2.5 percentage points) than for men (1.9 percentage points).

Marital status has a particularly large negative effect on women: the probability of retiring for partnered and single women is 5.9 and 6.0 percentage points lower than that of married women, respectively. For both men and women, being Hispanic and being born between 1950 and 1958 significantly reduces the probability of retiring.

Table A-12: Receiving Social Security benefits next wave

	Men		Women	
Frailty	0.00438***	(0.00144)	-0.00113	(0.00106)
Black	-0.0103	(0.0131)	-0.0406***	(0.0111)
Hispanic	-0.0534***	(0.0157)	-0.0477***	(0.0153)
Years of Education	-0.0192***	(0.00156)	-0.0246***	(0.00146)
FRA Dummy	0.0225	(0.0163)	0.0626***	(0.0167)
Born 1950-1968	-0.125***	(0.0104)	-0.0900***	(0.00961)
Partnered	-0.00767	(0.0207)	-0.0593***	(0.0218)
Single	0.0129	(0.0112)	-0.0595***	(0.00837)

Notes: Marginal effects resulting from logistic regressions. FRA dummy = full retirement age dummy.

Figure A-3 displays the predicted probabilities of retiring next wave by the frailty associated with having between 1 and 19 health deficits. Consistent with the marginal effect we computed in Table A-12, the left panel shows that, for men, higher frailty tends to increase the probability of retirement. However, this happens over some of the range of frailty, but not all of it, and its pattern depends on race and ethnicity. That is, the probability of retiring increases in frailty up to 0.37 for Hispanic men, 0.26 for White men, and 0.14 for Black men. Looking at the levels highlights that, at lower levels of frailty, the probability of retiring is significantly lower for Hispanic men.

The right panel shows that, for White and Hispanic women, the probability of retiring is quite flat in frailty, especially considering the large confidence intervals. For Black women, the probability of retiring increases up to a frailty of 0.26 and decreases afterward. There are no significant differences in the levels of the probability of retiring by frailty between Black and Hispanic women, while White women have a significantly higher probability of retiring for both low and high levels of frailty.

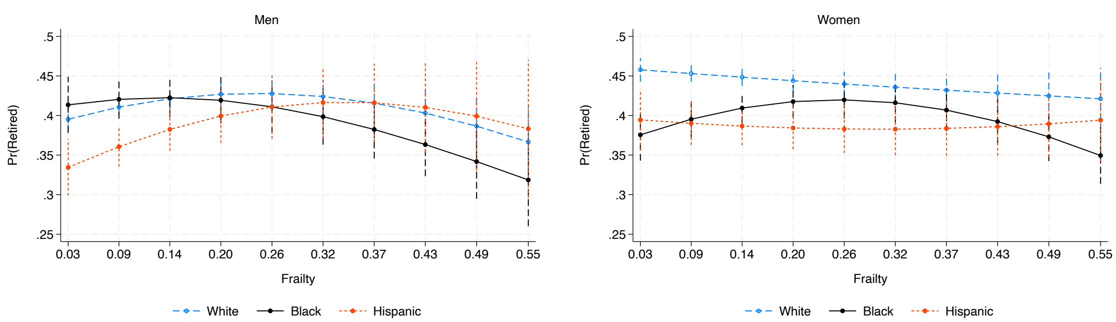


Figure A-3: Predicted probabilities of becoming a Social Security benefits recipient next wave by frailty. Men (left panel) and women (right panel). The frailty values reported in the horizontal axis correspond to 1 to 19 conditions. The vertical lines mark the 95% confidence intervals.

C.1.3 Nursing Home Entry

Table A-13 reports the marginal effects associated with nursing home entry next wave. Higher frailty significantly increases the probability of entering a nursing home: the probability of entering a nursing home increases by 0.3 percentage points for both men and women when they experience one more deficit. Interestingly here, and unlike for disability recipience, age does have an independent effect on the probability of nursing home entry even conditional on frailty. Being a year older increases this probability by about 0.2 percentage points for both men and women. Being single also increases it, especially for men, while being a Hispanic man or woman and a Black woman decreases it. In contrast, education turns out to have an insignificant effect.

Table A-13: Entering a nursing home next wave

	Men		Women	
Frailty	0.00315***	(0.000102)	0.00302***	(0.0000871)
Black	-0.00231	(0.00179)	-0.0100***	(0.00135)
Hispanic	-0.0122***	(0.00195)	-0.0139***	(0.00216)
Age	0.00212***	(0.0000959)	0.00238***	(0.0000866)
Years of Education	-0.0000721	(0.000168)	0.0000356	(0.000173)
Born 1930-1949	-0.00280*	(0.00154)	-0.00554***	(0.00149)
Born 1950-1968	-0.00254	(0.00479)	-0.00750*	(0.00416)
Partnered	0.00290	(0.00326)	0.00482	(0.00444)
Single	0.0125***	(0.00133)	0.00692***	(0.00107)

Notes: Marginal effects resulting from logistic regressions.

Figure A-4 displays the predicted probabilities of entering a nursing home next wave by the frailty associated with having between 1 and 19 health deficits. For men and women of all races and ethnicities, higher frailty leads to a higher probability of entering a nursing home. In particular, the left panel of Figure A-4 shows that White men have the highest probability of entering a nursing home at all frailty levels. This difference, however, is only statistically different from that of Hispanic men, who are the least likely to end up in a nursing home for every level of frailty. This is particularly noticeable for the unhealthiest men. Indeed, White men with 19 health deficits have an 11.6% chance of entering a nursing home next wave, while Black and Hispanic men with the same number of deficits have a probability of entering a nursing home of 9.0% and 5.1%, respectively.

In contrast, the right panel shows that the probability of entering a nursing home is significantly higher for White women than for their Black and Hispanic counterparts. In this case, the predicted probabilities significantly differ by race and ethnicity at almost all frailty levels. Similarly to what we observed for men, White women are the most likely to enter a nursing home, while Hispanic women are the least likely. This is particularly noticeable for the unhealthiest women. Indeed, White women with 19 health deficits have a 10.5% chance of entering a nursing home next wave, while Black and Hispanic women with the same number of deficits have a probability of entering a nursing home of 4.1% and 2.6%, respectively.

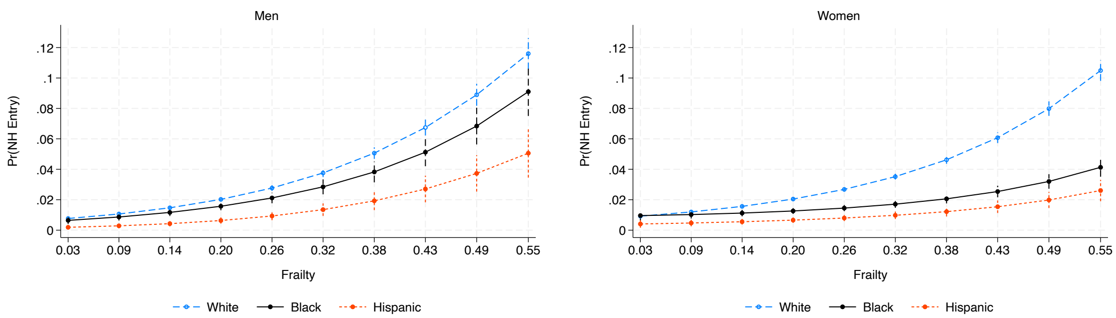


Figure A-4: Predicted probabilities of entering a nursing home next wave by frailty. Men (left panel) and women (right panel). The frailty values reported in the horizontal axis correspond to 1 to 19 conditions.

C.1.4 Death

Table A-14 reports the marginal effects associated with dying next wave. Here, too, frailty has a large effect. Increasing one's frailty by one deficit raises the probability of death by 0.8 and 0.6 percentage points for men and women, respectively. Interestingly, here age also has an independent effect, even conditioning on frailty. One more year of age raises the probability of death by 0.3 percentage points for men and by 0.2 percentage points for women. Being single, rather than married, also increases the probability of death, and more so for men (by 0.1 percentage points) than for women (0.07 percentage points).

Hence, for both men and women, being older, being single, and being more unhealthy increase the probability of death, while being born between 1930 and 1968 and being Hispanic lowers it.

Table A-14: Death next wave

	Men		Women	
Frailty	0.00796***	(0.000143)	0.00588***	(0.0000962)
Black	0.0000404	(0.00279)	-0.00512***	(0.00186)
Hispanic	-0.0120***	(0.00370)	-0.0109***	(0.00303)
Age	0.00330***	(0.000129)	0.00244***	(0.000102)
Years of Education	-0.000611**	(0.000259)	-0.0000203	(0.000228)
Born 1930-1949	-0.0151***	(0.00251)	-0.0103***	(0.00205)
Born 1950-1968	-0.0287***	(0.00436)	-0.0196***	(0.00363)
Partnered	0.0129***	(0.00492)	0.00122	(0.00490)
Single	0.0138***	(0.00195)	0.00675***	(0.00143)

Notes: Marginal effects resulting from logistic regressions.

Figure A-5 presents the predicted probabilities of dying next wave by the average frailty associated with having between 1 and 19 health deficits. For all men and women, higher frailty leads to a higher probability of death. The right panel shows that White men are significantly more likely to die than their Black and Hispanic counterparts for all frailty levels greater than 0.26 (which corresponds to having 9 health deficits). In particular, the most unhealthy White men are more than twice as likely to die as their Hispanic counterparts. Indeed, at a frailty level of 0.55, White men have a 26.7% probability of death, while Black and Hispanic men have a probability of 17.8% and 13.4%, respectively. The right panel displays similar dynamics for women's death probability. Here, for all frailty levels larger than 0.32, White women are the most likely to die, and Hispanic women are the least likely. In particular, the most unhealthy White women are more than twice as likely to die as their Hispanic counterparts. This is signaled by the fact that, at a frailty level of 0.55, the probability of death for White women is 17.5%, while the one for Black and Hispanic women is 10.5% and 7.6%, respectively.

C.1.5 Disability Insurance and Full Retirement Age

The Social Security Administration runs the Disability Insurance program for workers, their spouses, and dependents to provide insurance against health shocks that limit (partially or

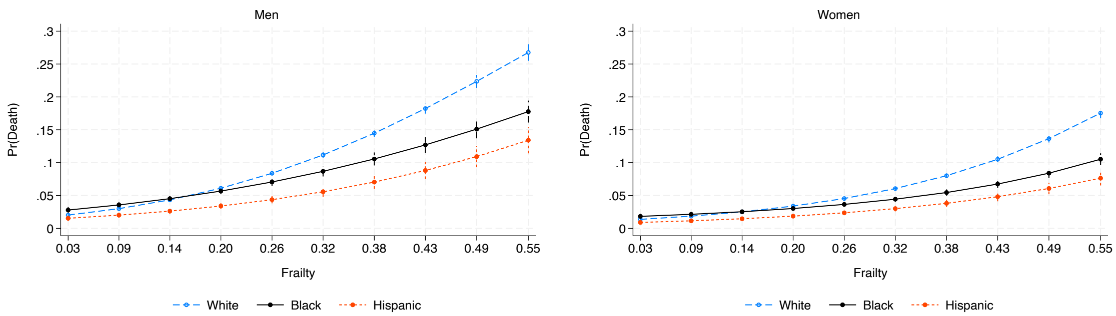


Figure A-5: Predicted probabilities of dying next wave by frailty. Men (left panel) and women (right panel). The frailty values reported in the horizontal axis correspond to 1 to 19 conditions.

entirely) people's ability to work. There are several rules surrounding Disability Insurance eligibility. First, workers must prove a sufficient work history. Second, their condition must meet the Social Security Administration's definition of a disability and last at least a year or result in death. Finally, applicants must be younger than their full retirement age.

The full retirement age depends on a person's year of birth. Table A-15 describes the evolution of the full retirement age as a function of the year of birth.[15]. In our empirical analysis described in Sections 3 and 5, we use a dummy for the Full Retirement Age when estimating logit regressions for the outcome "Receiving Social Security retirement benefits next wave". We construct this dummy using the ages in Table A-15 and setting it equal to 1 if the respondent is between 12 and 24 months younger than their corresponding full retirement age.

C.2 Principal Component Analysis details

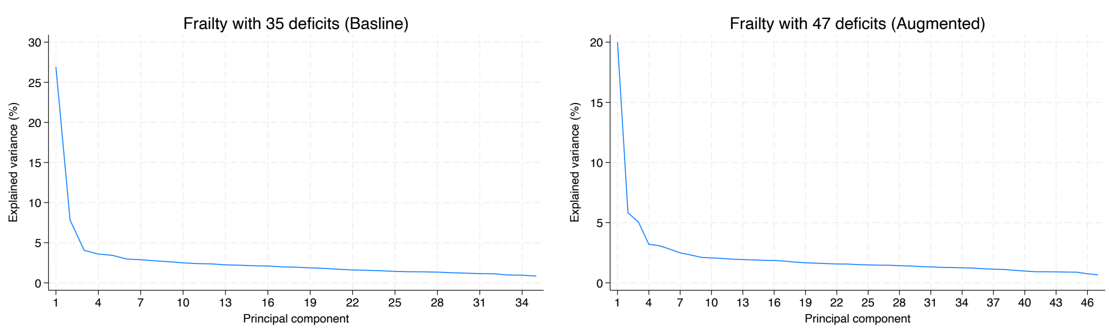
Figure A-6 displays the proportion of the variance explained by principal components for the baseline and the augmented frailty index. In both cases, the first principal component captures over 20% of the overall variance. Table A-16 reports the normalized weights resulting from PCA.

15. This table comes from https://www.ssa.gov/pressoffice/IncRetAge.html

Table A-15: Full retirement age

Year of birth	Full retirement age
1937 or earlier	65
1938	65 and 2 months
1939	65 and 4 months
1940	65 and 6 months
1941	65 and 8 months
1942	65 and 10 months
1943-1954	66
1955	66 and 2 months
1956	66 and 4 months
1957	66 and 6 months
1958	66 and 8 months
1959	66 and 10 months
1960 and later	67

Figure A-6: Proportion of the variance explained by principal components



Notes: These figures report the proportion of the variance in the data explained by the principal components.

Table A-16: PCA weights

(a) Baseline

	PCA weight
Diff. grocery shopping	0.0419
Diff. walking one block	0.0418
Diff. climbing flight of stairs	0.0416
Diff. walking several blocks	0.0414
Diff. bathing	0.0413
Diff. lifting >10 pounds	0.0409
Diff. walking across room	0.0401
Diff. pull/pushing large objects	0.0398
Diff. preparing hot meal	0.0397
Diff. dressing	0.0395
Diff. getting in/out of bed	0.0369
Diff. using toilet	0.0357
Diff. climbing several flights of stairs	0.0345
Diff. getting up from chair	0.0335
Diff. kneeling or crouching	0.0330
Diff. eating	0.0328
Diff. managing money	0.0326
Diff. lifting arms over shoulders	0.0318
Diff. making phone calls	0.0309
Diff. taking medication	0.0300
Diff. using map	0.0267
Nursing home stay	0.0264
Diff. picking up dime	0.0257
Diff. sitting for two hours	0.0255
Hospital stay	0.0223
Diagnosed with arthritis	0.0206
Diagnosed with psych. problem	0.0194
Diagnosed with a stroke	0.0187
Diagnosed with heart condition	0.0168
Diagnosed with lung disease	0.0152
Diagnosed with HBP	0.0141
Diagnosed with diabetes	0.0128
Has BMI $\geq$ 30	0.0073
Diagnosed with cancer	0.0059
Has ever smoked cigarettes	0.0029

(b) Augmented

	PCA weight
Diff. walking several blocks	.0370641
Diff. lifting >10 pounds	.03587
Diff. climbing flight of stairs	.0352521
Diff. pull/pushing large objects	.0351709
Diff. walking one block	.0345307
Diff. grocery shopping	.0316487
Diff. climbing several flights of stairs	.0316381
Diff. getting up from chair	.031113
Diff. dressing	.0305405
Diff. kneeling or crouching	.0304433
Diff. walking across room	.0299998
Diff. bathing	.0298966
Diff. lifting arms over shoulders	.0284349
Diff. getting in/out of bed	.0280726
Diff. preparing hot meal	.0270129
Troubled by pain	.026653
Diff. sitting for two hours	.0263375
Diff. using toilet	.0255662
Felt everything was an effort	.0253467
Could not get going	.0234303
Felt depressed	.0224874
Felt sad	.0211038
Felt alone	.0204219
Diff. eating	.0199634
Had restless sleep	.0199395
Diagnosed with arthritis	.0197268
Diff. managing money	.0195408
Diff. picking up dime	.019508
Did not feel happy	.0193579
Diagnosed with psych. problem	.0191895
Diff. using map	.0187435
Hospital stay	.0186874
Did not enjoy life	.0184713
Diff. taking medication	.017097
Diff. making phone calls	.0159966
Diagnosed with lung disease	.0150271
Diagnosed with heart condition	.0142573
Nursing home stay	.0126592
Diagnosed with HBP	.0125591
Diagnosed with a stroke	.012511
Diagnosed with diabetes	.0119435
Has BMI $\geq$ 30	.0093313
Backward count from 20	.0063847
Smoke now	.004935
Diagnosed with cancer	.0045589
Has ever smoked cigarettes	.0042137
Heavy alcohol use	-.0026377

D The Implementation of our Micro-Simulation Model

To evaluate to what extent health affects how long people spend in a given state, good health, being alive, and so on, we next turn to redefining the variables we study and a simulation exercise. Relative to our prediction exercise in Section 3, the focus of this analysis is the cumulative duration spent in a specific state. For this reason, we use outcome variables that are defined by the current state rather than predicting only the probability of entering a state. Thus, we account for flows both in and out, as well as the probability of remaining.[16] Table A-17 describes our outcome variables and the values they take.

Table A-17: Outcome variables

Variable	Description	Values
Health Next Wave	In wave t, this variable tells us the respondent's discretized health status in wave t+1	1 through 5 (quintile)
Death Next Wave	In wave t, this variable tells us if the respondent will die in wave t+1	0 if alive in t+1 1 if dead in t+1 missing if dead in t
SDI Recipient in Current Wave	In wave t, this variable tells us if the respondent receives SDI in wave t (less than the full retirement age)	0 if does not receive SDI in t 1 if receives SDI in t
Begin Receiving Social Security Benefits in Current Wave	In wave t, this variable tells us if the respondent claims SS benefits in t (ages 60 to 75, not previously claiming in t-1)	0 if no income from SS in t 1 if positive income from SS in t
Being in a Nursing Home in Current Wave	In wave t, this variable tells us if the respondent lives in a NH in wave t	0 if does not live in a NH in t 1 if lives in a NH in t

For the simulation exercise, we flexibly model non-linear health transitions and their impact on our outcomes of interest. We start by estimating a Markov process for frailty, which we discretize in five levels for tractability. While we use the cutoff points of frailty quintiles to determine in which category an individual is, we label each category as excellent, very good, good, fair, and poor health, just like the responses to self-reported health.

16. The two exceptions to this are death and receiving social security benefits because, as we describe below, both are best modeled as absorbing states.

We estimate the health transition probabilities of those who survive next period as

$$Prob(h_{i,t+1} = j) = H(h_{it}, X_{it}), \quad j = \{\text{Excellent, Very good, Good, Fair, Poor}\}, \tag{A2}$$

where X is a set of covariates that includes cohort dummies, race dummies, the interactions of race and discretized frailty, gender dummies and their interactions with discretized frailty, health insurance coverage dummies and their interactions with discretized frailty, a second-order polynomial in age and its interactions with gender, marital status dummies, a second-order polynomial in years of education, and the interaction between years of education and age.

Next, we model the probability of dying by the next wave as

$$\Pr(d_{i,t+1} = 1) = D(h_{it}, X_{it}). \tag{A3}$$

We model the probability of receiving disability benefits as

$$\Pr(di_{it} = 1) = \begin{cases} DI(h_{it}, di_{i,t-1}, X_{it}), & \text{if } age_{it} < FRA_i, \\ 0, & \text{if } age_{it} \geq FRA_i, \end{cases} \tag{A4}$$

where we take into account that disability benefits convert into retirement benefits upon reaching full retirement age (FRA).

We model the probability of receiving Social Security retirement benefits as

$$\Pr(ss_{it} = 1) = \begin{cases} 0 & \text{if } age_{it} \leq 60, \\ SS(h_{it}, X_{it}, t), & \text{if } 60 \leq age_{it} \leq 75 \text{ and } ss_{i,t-1} = 0, \\ 1, & \text{if } age_{it} > 75 \text{ or } ss_{i,t-1} = 1. \end{cases} \tag{A5}$$

Here, the set of controls, X_{it}, also includes a dummy for full retirement age which we describe in subsection C.1.5.

We model the probability of living in a nursing home as

$$\Pr(nh_{it} = 1) = NH(h_{it}, nh_{i,t-1}, X_{it}). \tag{A6}$$

We estimate the health transition probabilities in Equation A2 with an ordered logistic regression and use logistic regressions to estimate the probabilities in Equations A3-A6. We then simulate histories of health, disability and retirement benefits recipiency, nursing home stays, and death.[17] We quantify the effects of removing health inequality by assigning everyone the initial frailty (at age 55) of White people on our realized simulation histories.

Given a sample of initial conditions, we can construct simulated histories using the estimated health transitions and outcome probabilities in Equations A2-A6. To operationalize this, we select the first observation for individuals between the ages of 53 and 57 to produce our initial conditions and simulate 100 replications of each initial condition to construct simulated histories of health (including death), disability and retirement benefits recipiency, and nursing home stays.[18] Using our simulated histories, we compute the fraction of time spent in bad health, the number of working years, the number of years claiming disability or retirement benefits, the number of years spent in a nursing home in the last two years, and life expectancy. We then equalize initial conditions across races by assigning each non-White person a random draw from the (gender-specific) distribution of initial conditions for White people.

17. Hispanic people have low rates of nursing home residence. As a result, while we can estimate an ethnicity effect for them, we cannot reliably estimate the Hispanic-specific differential effect of health on nursing home entry. Hence, when estimating Equation A6, we constrain the effect of health for Hispanic people to be the same as that for White people.

18. When simulating, we assign all individuals an initial age of 55 and do not update their marital status or education. Age evolves deterministically and we assume health insurance coverage remains at the initial condition unless either an individual enrols on SSDI or reaches the age of 65. This captures statutory eligibility for Medicare, which covers over 95% of the retirees in our sample.

E Marginal Effects for Micro-Simulation Inputs

This section contains tables for the marginal effects in our dynamic system.

Table A-18: Marginal effects from ordered logit on health

Very good health in period $t-1$		
Excellent health in period t	-0.660***	(0.00259)
Very good health in period t	0.345***	(0.00297)
Good health in period t	0.268***	(0.00196)
Fair health in period t	0.0459***	(0.000741)
Poor health in period t	0.00180***	(0.0000430)
Good health in period $t-1$		
Excellent health in period t	-0.747***	(0.00223)
Very good health in period t	-0.0745***	(0.00269)
Good health in period t	0.470***	(0.00237)
Fair health in period t	0.331***	(0.00244)
Poor health in period t	0.0202***	(0.000403)
Fair health in period $t-1$		
Excellent health in period t	-0.755***	(0.00222)
Very good health in period t	-0.214***	(0.00205)
Good health in period t	0.0971***	(0.00154)
Fair health in period t	0.658***	(0.00224)
Poor health in period t	0.214***	(0.00204)
Poor health in period $t-1$		
Excellent health in period t	-0.756***	(0.00222)
Very good health in period t	-0.227***	(0.00204)
Good health in period t	-0.0114***	(0.000310)
Fair health in period t	0.0862***	(0.00195)
Poor health in period t	0.909***	(0.00204)
Black		
Excellent health in period t	-0.0127***	(0.00177)
Very good health in period t	0.00994***	(0.00182)
Good health in period t	0.000888	(0.00130)
Fair health in period t	-0.000367	(0.00132)
Poor health in period t	0.00220*	(0.00125)
Hispanic		
Excellent health in period t	-0.00342*	(0.00197)
Very good health in period t	0.00735***	(0.00201)
Good health in period t	0.000628	(0.00156)
Fair health in period t	-0.00190	(0.00161)
Poor health in period t	-0.00266	(0.00162)
Being a man		
Excellent health in period t	0.00170	(0.00110)
Very good health in period t	-0.00118	(0.00117)
Good health in period t	-0.00108	(0.000913)
Fair health in period t	0.00297***	(0.00100)
Poor health in period t	-0.00242**	(0.00100)
Having health insurance		
Excellent health in period t	0.00220	(0.00218)
Very good health in period t	0.00342	(0.00240)
Good health in period t	0.000542	(0.00201)
Fair health in period t	-0.00437*	(0.00229)
Poor health in period t	-0.00180	(0.00229)
Age		
Excellent health in period t	-0.00188***	(0.0000648)
Very good health in period t	-0.000483***	(0.0000227)
Good health in period t	0.0000230	(0.0000191)
Fair health in period t	0.000476***	(0.0000243)
Poor health in period t	0.00186***	(0.0000474)
Being partnered		
Excellent health in period t	-0.00725***	(0.00192)
Very good health in period t	-0.000995***	(0.000259)
Good health in period t	0.000577***	(0.000156)
Fair health in period t	0.00194***	(0.000497)
Poor health in period t	0.00573***	(0.00153)
Being single		
Excellent health in period t	-0.00652***	(0.000802)
Very good health in period t	-0.000899***	(0.000118)
Good health in period t	0.000520***	(0.0000728)
Fair health in period t	0.00175***	(0.000218)
Poor health in period t	0.00515***	(0.000635)
Years of education		
Excellent health in period t	0.00269***	(0.000148)
Very good health in period t	0.000133***	(0.0000354)
Good health in period t	-0.000370***	(0.0000320)
Fair health in period t	-0.000815***	(0.0000506)
Poor health in period t	-0.00164***	(0.0000980)
1895-1909 cohort		
Excellent health in period t	-0.00710	(0.00533)
Very good health in period t	-0.000925	(0.000666)
Good health in period t	0.000556	(0.000415)
Fair health in period t	0.00179	(0.00130)
Poor health in period t	0.00568	(0.00429)
1910-1929 cohort		
Excellent health in period t	0.00356**	(0.00165)
Very good health in period t	0.000495**	(0.000230)
Good health in period t	-0.000283**	(0.000131)
Fair health in period t	-0.000955**	(0.000441)
Poor health in period t	-0.00281**	(0.00131)
1930-1949 cohort		
Excellent health in period t	0.00214*	(0.00113)
Very good health in period t	0.000295*	(0.000155)
Good health in period t	-0.000170*	(0.0000897)
Fair health in period t	-0.000569*	(0.000298)
Poor health in period t	-0.00169*	(0.000897)

Table A-19: Marginal effects for death next wave

Very Good	0.0113***	(0.00105)
Good	0.0230***	(0.00122)
Fair	0.0463***	(0.00129)
Poor	0.137***	(0.00199)
Black	0.000703	(0.00152)
Hispanic	-0.0151***	(0.00185)
Male	0.0380***	(0.00119)
Health Insurance coverage=1	-0.0000199	(0.00395)
Age	0.00311***	(0.0000820)
Partnered	0.00875**	(0.00342)
Single	0.0118***	(0.00119)
Years of education	-0.000569***	(0.000171)
1895-1909 cohort	0.0390***	(0.00470)
1910-1929 cohort	0.0235***	(0.00269)
1930-1949 cohort	0.0128***	(0.00213)

Notes: Very Good, Good, Fair, and Poor refer to discretized frailty.

Table A-20: Marginal effects for disability benefits recipiency

Very Good	0.0218***	(0.00230)
Good	0.0455***	(0.00255)
Fair	0.0833***	(0.00262)
Poor	0.145***	(0.00412)
Black	0.00392**	(0.00173)
Hispanic	-0.0102***	(0.00212)
Male	0.0165***	(0.00151)
Health Insurance coverage=1	0.00638***	(0.00207)
Age	0.000367*	(0.000208)
Partnered	0.00933***	(0.00304)
Single	0.00727***	(0.00154)
Years of education	-0.000672**	(0.000269)
Past disability recipient	0.125***	(0.00134)
1930-1949 cohort	-0.00620***	(0.00153)

Notes: Very Good, Good, Fair, and Poor refer to discretized frailty.

Table A-21: Marginal effects for retirement benefits recipiency

Very Good	0.0263***	(0.00600)
Good	0.0296***	(0.00668)
Fair	0.0286***	(0.00642)
Poor	0.0124*	(0.00722)
Black	-0.0214***	(0.00580)
Hispanic	-0.0429***	(0.00717)
Male	-0.0248***	(0.00430)
Health Insurance coverage=1	-0.0575***	(0.00744)
Age	0.0739***	(0.000666)
Partnered	0.0148	(0.0114)
Single	-0.00402	(0.00488)
Years of education	-0.0162***	(0.000794)
FRA dummy	0.0327***	(0.00536)
1910-1929 cohort	0.0949***	(0.0242)
1930-1949 cohort	0.0715***	(0.00513)

Notes: Very Good, Good, Fair, and Poor refer to discretized frailty.

Table A-22: Marginal effects for nursing home residence

Very Good	0.00173**	(0.000870)
Good	0.00611***	(0.000864)
Fair	0.0104***	(0.000798)
Poor	0.0454***	(0.00119)
Black	-0.00383***	(0.000729)
Hispanic	-0.00758***	(0.000927)
Male	0.00292***	(0.000663)
Age	0.00101***	(0.0000489)
Partnered	0.00194	(0.00196)
Single	0.00927***	(0.000630)
Years of education	0.0000771	(0.0000931)
Previously living in a nursing home	0.0587***	(0.00103)
1895-1909 cohort	0.0111***	(0.00246)
1910-1929 cohort	0.00597***	(0.00170)
1930-1949 cohort	0.00381***	(0.00148)

Notes: Very Good, Good, Fair, and Poor refer to discretized frailty.