medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Disparities in Vulnerability to Severe Complications from COVID-19 in the United States

Emily E. Wiemers[1]
Scott Abrahams[2]
Marwa AlFakhri[2]
V. Joseph Hotz[2]
Robert F. Schoeni[3]
Judith A. Seltzer[4]

May 27, 2020

Abstract

This paper provides the first nationally representative estimates of vulnerability to severe complications from COVID-19 overall and across race-ethnicity and socioeconomic status. We use the Panel Study of Income Dynamics (PSID) to examine the prevalence of specific health conditions associated with complications from COVID-19 and to calculate, for each individual, an index of the risk of severe complications from respiratory infections developed by DeCaprio et al. (2020). We show large disparities across race-ethnicity and socioeconomic status in the prevalence of conditions, including hypertension, which are associated with the risk of severe complications from COVID-19. Moreover, we show that these disparities emerge early in life, prior to age 65, leading to higher vulnerability to such complications. Our results suggest particular attention should be paid to the risk of adverse outcomes in midlife for non-Hispanic blacks, adults with a high school degree or less, and low-income Americans.

[1]Syracuse University
[2]Duke University
[3]University of Michigan
[4]University of California, Los Angeles

Corresponding author: Emily Wiemers, eewiemer@maxwell.syr.edu, Department of Public Administration and International Affairs, 320C Lyman Hall, Syracuse University, Syracuse, NY 13244

Funding Acknowledgement: This paper was prepared with support, in part, from the Aging Studies Institute at Syracuse University, the Duke Center for Population Health and Aging, which receives core support (P30AG034424) from the National Institute on Aging, and by the California Center for Population Research at the University of California at Los Angeles, which receives core support (P2C-HD041022) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The collection of data used in this study was partly supported by the National Institutes of Health under grant number R01 HD069609 and R01 AG040213, and the National Science Foundation under award numbers SES 1157698 and 1623684.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

INTRODUCTION

The emerging evidence shows that the health impact of COVID-19 differs substantially across demographic groups. Hospitalization rates from COVID-19 are higher for men than women, and for older adults than younger (Garg et al., 2020). Blacks are overrepresented among hospitalized patients while Hispanics and whites are underrepresented relative to their population percentages (Garg et al., 2020). Data to date show staggering differences in mortality rates by race. At the extreme is Washington D.C., where African Americans make up over 75% of deaths from COVID-19 but only 44% of the population (CDC, 2020a).[1]

Behind these disparities in hospitalizations and deaths from COVID-19 are disparities in physical conditions like density in housing and neighborhoods, in the use of public transportation, in occupational hazards, in health care access and quality of care, and in underlying health conditions (Blumenshine et al., 2008). Each of these factors has been important during the COVID-19 pandemic. Early evidence showed that testing in New York City was less prevalent in poor, dense, and non-white neighborhoods but infection rates conditional on testing were higher (Borjas, 2020). Less-educated, lower income, and less wealthy workers are more likely to work in occupations where working from home and remaining physically distant on the job are harder, which increases the likelihood of infection (Mongey, Pilosoph, & Weinberg, 2020). Race differences in hospitalization rates persist even after controlling for comorbidities and there appear to be differences in the timing of testing by race with blacks more likely to be tested later in the course of the disease, suggesting unequal access to quality health care (Azar et al., 2020).

This paper focuses on the role that underlying health conditions and differences in these conditions across demographic and socioeconomic groups play in vulnerability to severe complications from COVID-19. The CDC reports that over 90% of hospitalized patients have at least one underlying health condition (CDC, 2020b). There are large existing gaps in health and chronic conditions by socioeconomic status and race (Case & Deaton, 2020; Harris & McDade, 2018; Mead et al., 2008; NCHS, 2012) and large mortality gradients by education, race, and income,

[1] In this paper we do not have sufficient data to examine non-Hispanic American Indian or Alaska Native populations, so we do not emphasize differential mortality among these groups. However, the emerging gaps in mortality are very large. In Arizona, 22% of deaths from COVID-19 are among non-Hispanic American Indians though they make up only 2% of the state's population (CDC, 2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

especially in midlife (Case & Deaton, 2015; Chetty et al., 2016; Currie & Schwandt, 2016). Vulnerability to severe complications from COVID-19 collides with these large and long-standing disparities in health and mortality.

While existing surveillance systems have provided data on confirmed infections of COVID-19, and the resulting hospitalizations and deaths, we know very little about how this varies in the population, aside from by race-ethnicity, and that evidence is not available nationally. The lack of such information has limited the ability to assess the distributional impacts of the severity of the disease in the U.S. population. Central to predicting the likely impact of the COVID-19 pandemic across demographic and socioeconomic groups is knowing who is vulnerable to severe illness and hospitalization from being infected with the Coronavirus.

This paper provides evidence on the distribution of risks for severe illness from the COVID-19 in the U.S. population based on underlying health conditions and age. We do so by using a "vulnerability index" (see DeCaprio et al, 2020) recently developed by ClosedLoop, an artificial intelligence data science platform that has developed a predictive model to identify people with heightened vulnerability to severe complications from the virus. The model is based on medical claims data indicating severe respiratory infections from diseases other than COVID-19. We use one of their models to map the age and self-reported health conditions of adult members of the Panel Study of Income Dynamics (PSID) to form population-representative estimates of the distribution of vulnerability to severe complications from COVID-19 in the United States and how it varies by education, income, and race-ethnicity. We complement this analysis by examining the distribution of underlying health conditions reported in the PSID across these groups, focusing on the number of conditions at younger ages, middle age, and older ages by education, income, and race-ethnicity.

We find that disparities across race-ethnicity and socioeconomic status in the prevalence of underlying health conditions translate into large gaps in vulnerability to severe complications from COVID-19. Overall, adults with twelve years of education or less are twice as likely as adults with a college degree to face severe complications from COVID-19. Adults in the lowest income quartile face three times the risk of severe complications compared to those in the highest income quartile. These gaps in risk of severe complications from COVID-19 are due to underlying health conditions that emerge early in life for less-educated and low-income adults. By race-ethnicity, we find large gaps in vulnerability to severe complications appearing in midlife for non-Hispanic

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

blacks, whereas Hispanics actually face lower risks than non-Hispanic whites based on underlying health conditions and age.

To our knowledge, we are the first to provide population-representative estimates of the risks of severe complications from COVID-19 using the distribution of underlying health conditions in the population. While improved predictive models of individual risk are likely to emerge over the coming months and years, our approach provides an immediate means of translating multiple underlying health risk factors into population-based estimates of the likely impacts of the pandemic and to identify groups that are more vulnerable to its impact.

Understanding the distribution of underlying health conditions across demographic and socioeconomic groups and how these conditions accumulate in terms of vulnerability to complications from COVID-19 are crucial pieces in assessing the disparities in the likely effect of the disease. Population-representative estimates are essential for this. Because we focus only on disparities in vulnerability that stem from underlying health conditions, our work should be seen as a complement to analysis of disparities in exposure to the virus and access to high quality health care.

BACKGROUND

Frameworks developed by Blumenshine et al. (2008) and Kumar and Quinn (2012) that describe the role of socioeconomic factors and race/ethnicity, in influenza pandemics are useful for understanding the relationships we examine. First, socioeconomic factors influence exposure to the virus. Disadvantaged populations may be less likely to be able to socially distance for a variety of reasons, including working in different occupations, living in smaller homes with more occupants, located in more densely populated areas, depending more on childcare outside the home, and relying more on public transportation. Second, dimensions of socioeconomic status, including race-ethnicity, influence susceptibility to contracting infection and to complications from infection. Factors influencing susceptibility include, among others, age, smoking, preexisting medical conditions (e.g., diabetes, hypertension, asthma), nutrition, and stress, including stress due to discrimination. All of these factors are associated with socioeconomic status and with being a racial minority. Finally, more advantaged populations have better access to and receive higher quality medical care, which may lead to earlier and more effective treatment, reducing the risk of severe complications.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Underlying Health Conditions and COVID-19

Our approach uses an individual's underlying conditions to assess their risk for severe complications from COVID-19. This approach is supported by worldwide evidence that shows that severe complications from COVID-19 are associated with underlying health conditions. In China, Italy, and the United States, most people admitted to the hospital for COVID-19 or for severe complications from COVID-19 had at least one underlying health condition (Arentz et al., 2020; Bhatraju et al., 2020; Cummings et al., 2020; Emami et al., 2020; Grasselli et al., 2020; Ma et al., 2020; Richardson et al., 2020; Scodia et al., 2020; Yang et al., 2020). For example, Richardson et al. (2020), showed that 94% of patients hospitalized for COVID-19 in New York City, Long Island, and Westchester County, NY, had at least one underlying health condition. Among those already in severe respiratory distress when they were hospitalized, 82% had at least one underlying health condition (Cummings et al., 2020, for 2 northern Manhattan hospitals). Many people had more than one underlying health condition. Richardson et al. reported that 88% of patients had more than one underlying health condition, and the median number of conditions among New York hospitalized patients was between 2 and 4 (Cummings et al., 2020; Richardson et al., 2020). In a small study in Washington state, the prevalence of underlying health conditions was lower, with 33% of hospitalized patients having more than one pre-existing condition (Bhatraju et al., 2020). Differences across studies in the prevalence of comorbidities can be attributed in part to how investigators defined comorbidities, differences in the prevalence of comorbidities across communities, and local variation in access to hospital beds and public health guidance on who should be hospitalized.

In nearly every study, the most common underlying health condition among patients hospitalized for COVID-19 was hypertension (Arentz et al., 2020; Bhatraju et al., 2020; Cummings et al., 2020; Emami et al., 2020; Grasselli et al., 2020; Ma et al., 2020; Richardson et al., 2020; Scodia et al., 2020; Yang et al., 2020). Other common conditions included diabetes, cardiovascular disease, kidney disease, and obesity (Emami et al., 2020; Cummings et al., 2020; Grasselli et al., 2020; Ma et al., 2020; Richardson et al., 2020; Scordia et al., 2020).

Underlying health conditions were associated with more severe outcomes – such as the development of acute respiratory distress syndrome and death—even for hospitalized patients (Cummings et al., 2020; Ma et al., 2020; Scordia et al., 2020). A meta-analysis of seven studies from China shows that severely ill patients had a more significant number of comorbid conditions

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

than hospitalized patients whose illnesses were less severe (Yang et al., 2020). Case fatality rates in China were higher among those with comorbid conditions including cardiovascular disease, diabetes, chronic respiratory disease, hypertension, and cancer (Wu & Googan, 2020). In Italy, patients with hypertension who were admitted to the ICU were more likely to die than those without hypertension (Grasselli et al., 2020).

Why Our Estimates Might Understate Socioeconomic and Race-Ethnic Disparities

Although we show very large differences in the risk of severe complications from COVID-19 by race-ethnicity and socioeconomic status, our estimates likely understate disparities in vulnerability to severe complications based on underlying health conditions for several reasons. First, some dimensions of health that affect severe complications may not be measured and they, like nearly every disease, are likely more common among disadvantaged populations. Second, the specification of the DeCaprio et al. (2020) model did not allow differences in the effects of risk factors by race-ethnicity or other socioeconomic factors, as they were not present in the data they used. However, there is some suggestive evidence that effects of preexisting conditions on hospitalization from influenza are larger for populations from disadvantaged areas (Glezen et al., 2000). The same may be true for COVID-19. Third, socioeconomic status has an effect on hospitalization for influenza above and beyond its effect on preexisting conditions (Lowcock et al., 2012); again, this might also be true for COVID-19. Fourth, disadvantaged populations have higher rates of undiagnosed diseases, implying that the observed disparities in preexisting conditions understate true disparities (Barcellos et al., 2012; Chobanian, 2009; Pleasants et al., 2016; Smith, 2007; Zallman et al., 2013). Fifth, among those who have been diagnosed with a disease, rates of control of the disease are lower for disadvantaged populations (Chatterji et al., 2012; Chobanian, 2009; Pleasants et al., 2016; Zallman et al., 2013). For these reasons, our estimates of vulnerability to severe complications from COVID-19 based on underlying health conditions may understate disparities across race-ethnic and socioeconomic groups.

DATA

In this section we discuss the data and methods we use to analyze the underlying health conditions that the CDC and the developing literature on hospitalization and death from COVID-19 indicates as risk factors for severe illness from COVID-19. We begin with a discussion of the Panel Study of Income Dynamics (PSID) data that we use to determine the distribution of and

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

disparities in risk factors in the U.S. population of adults age 25 years and older. We consider disparities aross gender, race-ethnicity, educational attainment, and household income. We also describe an index of relative vulnerability to severe illness from respiratory infections recently developed by ClosedLoop (DeCaprio et al., 2020), which is a function of the CDC risk factors, recent hospitalization, and age. We discuss the construction of the vulnerability index and how we apply it to nationally representative data.

Panel Study of Income Dynamics

The Panel Study of Income Dynamics (PSID) is a national longitudinal survey that has interviewed the original sample members and their descendants every 1-2 years since 1968. The original sample included over samples of black and low-income families to facilitate the study of poverty. With the addition in 1997 of a sample of immigrants who arrived in the United States after 1968 plus the addition in 2017 of a sample of immigrants who arrived after 1997, the weighted sample of 26,455 individuals of all ages living in 9,607 families is representative of the national population in 2017.2 We study the adult sample age 25 and older consisting of 13,529 heads and spouses (or cohabiting partners)3 of households who are not missing data on age, race-ethnicity, educational attainment or household income.4 Throughout our analyses, we use the PSID, 2017 cross-sectional weights.

Demographics and Socioeconomic Status

Measures of race and Hispanic origin are combined to classify race-ethnicity as Hispanic, non-Hispanic black, non-Hispanic white, and non-Hispanic other race. There are insufficient sample sizes to examine separately races other than black and white; therefore, individuals who are non-Hispanic other race are included in analyses, but we do not report separate estimates for this group. Education is categorized as no more than 12 years (high school or less), 13-15 years (some college), and at least 16 years (a BA degree or more). Income is measured at the household level.

[2] https://psidonline.isr.umich.edu/data/Documentation/UserGuide2017.pdf

[3] To be designated as a cohabiting partner, the person had to have lived with the head for one year or more. Starting in 2017, the PSID began to refer to heads as “reference person 1” and spouses and long-term cohabiters as “reference person 2.” For ease in comparison to the CPS and, we continue to use the labels “heads and spouses.”

[4] There are 13,822 heads and spouses, age 25 and older, in the 2017 Wave of the PSID. We dropped 293 observations due to missing data, producing an analysis sample of 13,529 used in all of our analyses.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

We classify each adult according to the income quartile of the household to which they belong.[5]

The basic population characteristics of our analysis sample of household heads and spouses age 25 and older in the U.S. population are displayed in Table 1. We show the distributions of gender, race-ethnicity, educational attainment, and household income quartile overall and for three age groups (25-44, 45-64 and 65+). We also show the distribution of age across each race-ethnic, educational attainment, and household income category. In Appendix A, we compare the PSID data to that of the Current Population Survey (CPS) from the 2017 Annual Social and Economic Supplement (ASEC) for household heads and their spouses or cohabiting partners by age group across gender, race-ethnicity, and educational attainment. Overall, the samples produce almost identical distributions.[6] Prior work has shown that estimates of income from the PSID match well with those from the March CPS (McGonagle et al., 2012).

Differences in the age distribution across demographic and socioeconomic groups is notable because the risks from COVID-19 vary with age. Females are one year older, which is consistent with well-documented higher mortality rates of men than women. Non-Hispanic whites (NH white) make up about 71% of the adult U.S. population, while 10.7% are Non-Hispanic blacks (NH blacks), 13.2% are Hispanics and 4.9% are other non-Hispanic groups. NH whites are the oldest on average (53.7 years old), followed by NH blacks (49.4), and Hispanics are the youngest group at 46.7 years of age. Among NH whites, 28.2% are age 65 or older compared to only 16.8% of NH blacks, and 13.1% of Hispanics. With respect to educational attainment, 38.7% of the U.S. adult population has a high school degree or less, 23.5% have completed some college, but not a BA degree and 37.8% have a BA degree or higher. Due to educational differences across birth cohorts, those with a high school or less are older (54.4 years old) than either those with some college (51.2) or those with a BA degree or more (50.1). Finally, 18.3% of individuals in the adult population are in the lowest quartile of household income, while 22.7%, 27.5% and 31.5% are in the second, third and top quartiles, respectively. Again, we see age differences across these quartiles, with those in the lowest quartile being older (56.7 years) compared to the top two quartiles,

[5] The lowest quartile consisted of those households with reported incomes between $0 and $29,096; the second quartile, incomes between $29,096 and $56,760; the third quartile, incomes between $56,760 and $101,820; and the top quartile, incomes greater than $101,820. The measure is based on reports about the income households received in 2016 and is reported in 2016 dollars.

[6] There are slight differences between the PSID and CPS samples for the distributions of educational attainment, presumably because of slightly different questions.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

which average 50.4 years.

Self-Reported Health Conditions (*Risk Factors*)

In the 2017 Wave, the PSID included a battery of questions about whether the heads and spouses were ever told by a doctor or health professional they had particular health conditions. Many of these conditions are listed by the CDC as risk factors for becoming severely ill from infection of COVID-19. The health conditions (risk factors) in the PSID include the diagnoses for asthma, diabetes, heart disease, heart attack, hypertension, lung disease, neurological conditions, cancer, stroke, and kidney disease. In addition, the PSID asked respondents whether they had been hospitalized in 2016 (the reference year for most of the 2017 Wave survey); hospitalization is a risk factor in the DeCaprio et al. (2020) Model described below. Finally, respondents were asked to report their height and weight, which we used to compute each respondent's BMI. We then created an indicator of whether a respondent is severely obese (BMI $\geq$ 40). (See Appendix Table B-2 for the question wording for all self-reported health conditions in the PSID.) One risk factor that the CDC identifies that we are not able to include is living in a nursing home, for which the sample size in the PSID is insufficient.

PSID staff compared estimates of these self-reported health questions to gold-standard topical surveys to evaluate the quality of PSID data. The PSID estimates of the prevalence of health conditions for nearly all measures are close to those in the National Health Interview Survey (NHIS), a personal household interview study that has been used by the National Center for Health Statistics to monitor Americans' health since 1957.[7] PSID estimates also show time trends consistent with those from the NHIS (Insolera & Freedman, 2017). There are two exceptions. Chronic kidney disease is two to three times more common in the NHIS than in the PSID, in which 0.80 percent report kidney disease.[8] The PSID information about kidney disease, unlike all other conditions we examine, comes from reports about "other serious chronic conditions," without a specific prompt for kidney-related conditions. The NHIS information is obtained from a specific question about "weak or failing kidneys," and specific questions are more likely to elicit these reports. The other exception is the prevalence of "neurocognitive conditions," which we indicate by

[7] See https://www.cdc.gov/nchs/nhis/about_nhis.htm for more details about the NHIS. Also see https://psidonline.isr.umich.edu/Publications/Papers/tsp/2017-01_Health_Data_update_2015.pdf for a more detailed comparison of the health conditions measures collected in the PSID with other studies, including the NHIS.

[8] See https://www.cdc.gov/nchs/nhis/ADULTS/www/index.htm.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

whether a doctor or health professional had ever told the head or spouse that they had permanent loss of memory or mental ability" (see Appendix Table B-2). However, the PSID self-reported measure has external validity as demonstrated by its consistent positive association with the Eight Item Interview to Differentiate Aging and Dementia Screen (AD8) (Freedman, McFall, & Ryan, 2019).

Relative Vulnerability Index

In this section we discuss the vulnerability index for severe illness from the Coronavirus that DeCaprio et al. (2020) develop, including exactly how it was formed and what it characterizes, and how we use it to assess socioeconomic disparities in vulnerability**.**

According to the CDC, the risk of severe complications from COVID-19 resulting in hospitalization increases with a number of preexisting health conditions, i.e., risk factors or comorbidities, and with age, but the size of the effect of each risk factor is currently unknown (CDC, 2020). To circumvent the lack of current data on these effects, DeCaprio et al. (2020) assessed the risk of hospitalization for respiratory infections available in existing medical claims data. In particular, they looked at (in-patient) hospitalizations associated with: acute respiratory distress syndrome, pneumonia (other than caused by tuberculosis), influenza, acute bronchitis and other upper respiratory infections. Such conditions are coded in most hospitalization claims data sources. DeCaprio et al. (2020) used medical claims data on hospitalizations for Medicare recipients – using the Center for Medicare and Medicaid Services Limited Data Set, which contains a 5% sample of the Medicare population – and from Healthfirst, a New York (non-Medicare population) insurer. They mapped in-patient hospitalizations from these respiratory infections to individual-level data on 11 pre-existing health conditions (risk factors) coded in the Medicare and Healthfirst data, along with patients' gender, age and whether they had been hospitalized in the last three months.

As described in DeCaprio et al. (2020), the authors estimated several alternative versions of their model. In our paper, we use results from their "Survey Model," which is based on a logistic regression of incidence of hospitalizations due to the above conditions. DeCaprio et al. (2020) described other, more sophisticated, models that they estimated with claims data. These models, largely designed for use by health care providers, required claims data to implement. Their Survey Model is more suitable for use with the health conditions available in survey data, such as the

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

PSID.[9] Below, we shall simply refer to the Survey Model in DeCaprio et al. (2020) as the DeCaprio et al. model.

The DeCaprio et al. model is of the following form. Let d_i denote a 0/1 indicator of whether individual *i* has a severe respiratory infection and ends up hospitalized. Then,

(1) $$p_i \equiv Prob(d_i = 1) = \exp(VI_i^*)/[1 + \exp(VI_i^*)],$$

where d_i is a 0/1 indicator for whether *i* is hospitalized and VI_i denotes individual *i*'s *Vulnerability Index score* which is given by the following function:

(2) $$VI_i^* = \beta_0 + \beta_1 Age_i + \beta_2 Male_i + \beta_3 PHosp_i + \boldsymbol{\beta}_4 \boldsymbol{R}_i + \boldsymbol{\beta}_5 \boldsymbol{R}_i \cdot Age_i,$$

where Age_i is *i*'s current age, $Male_i$ is a 0/1 indicator of whether *i* is a male; $PHosp_i$ is a 0/1 indicator of whether *i* has been recently hospitalized, and $\boldsymbol{R}_i$ is a vector of 0/1 indicators for whether *i* has particular health conditions, or *risk factors*, R_{ki} *k*, *k* = 1,…, *K*. Given the logistic form of (1), it follows that VI_i^* is just the *log odds* of severe illness for individual *i* that is a function of their age, gender, recent hospitalization and health-related risk factors. It follows that

(3) $$\widetilde{VI}_i = \exp(VI_i^*)$$

is their *odds ratio* of risk for severe illness or the *relative vulnerability index* score for individual *i*.

Given the expression for VI_i^* in (2), technically the base group for the DeCaprio et al. model is a female with no underlying risk factors who is age equal to 0. But, to provide a more meaningful base group for vulnerability score, we use the following relative vulnerability index score in all of our calculations in the paper:

(4) $$VI_i = \exp(VI_i^* - \beta_0 - \beta_1 \cdot 30)$$

so, the score we report, VI_i, provides the odds that individual *i* has of a severe illness relative to that for a 30-year-old female with no risk factors. We report average VI_i by age, race-ethnicity, education, and household income.

In Appendix Table B-1, we display, in the first two columns, respectively, the risk factors in $\boldsymbol{R}_i$ for the DeCaprio et al. model and estimates of the β_j's for VI_i^* in (2). The third column in

[9] For one version of their model, they have set up a website (https://closedloop.ai/c19index/) where one can calculate values of what they refer to as the "C-19index" based on risk factors, hospitalization experience and age a user supplies. The Survey Model documented in DeCaprio et al. (2020) that we use is similar to, but not exactly the same as, the C-19index, as the website model is trained on some additional information that we do not use, e.g., the zip-code where one resides.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Appendix Table B-1 indicates which health risk factors are available in the PSID. As one can see, we are missing several risk factors from the DeCaprio et al. model. In our implementation of the vulnerability index with PSID data, we do not attempt to adjust the formula for absent variables; rather, we act as if the members of the PSID sample do not have these conditions. Because of this, the values of the vulnerability index scores for our sample will be biased downwards. As discussed in Appendix B, almost all of the risk factors used in the DeCaprio et al. model that are not available in the PSID, with the exception of liver disease (cirrhosis of the liver), either have very low marginal impacts on the vulnerability index (based on the coefficient estimates for these conditions) or the incidence of the risk factor is very low in the U.S. population.

Finally, to provide a sense of how the various risk factors affect the relative vulnerability index scores, we present, in Figure 1, the age profiles of the marginal contributions of each risk factor to relative vulnerability. These marginal contribution at age, Age_i, for risk factor, R_{ki}, is given by:

$$\frac{\partial VI_i}{R_{ki}} = (\beta_{4k} + \beta_{5k} Age_i) \cdot \exp(\beta_1 [Age_i - 30] + \beta_{4k} + \beta_{5k} Age_i) \tag{5}$$

where we evaluate all of the other risk factors, $R_{mi}, m \neq k$, at 0.

In Figure 1, we draw attention to those risk factors that have the most sizeable contributions to vulnerability at various ages by labeling them in the figure. As one can see from Figure 1, having hypertension and lung disease have the largest impacts on the relative vulnerability index generated by the DeCaprio et al. model at all ages, and both effects rise rapidly with age. Severe obesity – and, to a lesser extent, having had cancer – also consistently have a large impact on the vulnerability index at all ages. In early adulthood (25-44), asthma contributes a sizeable impact on relative vulnerability, but this risk factor is less important at midlife (45-64) and older ages (65+). With age, one also sees that heart disease and heart attacks contribute to relative vulnerability in the DeCaprio et al. model. At older ages (65+), neurological conditions start to have substantial impacts on relative vulnerability. Finally, recent hospitalizations, which presumably proxy for compromised health, have larger impacts on relative vulnerability as individuals age.

All of the other risk factors individually have more modest impacts, raising the vulnerability odds of severe disease relative to a 30-year old female with no risk factors 2 to 5 times and do not vary substantially with age. Nonetheless, the DeCaprio et al. model implies that the increase in relative vulnerability to severe illness will result, in large part, from the number of conditions

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

one has at any age.

Three final points about Figure 1. First, age, *per se*, has a very small impact on relative vulnerability. That is, growing older, when one has no risk factors, has virtually no impact on vulnerability. It is the accumulation of risk factors and the interactions of these risk factors with age that leads to higher vulnerability at older ages. Second, a male, who has no other health-related risk factors, is at slightly higher vulnerability to severe illness compared to females and this male-female differential in vulnerability rises modestly with age, so that, by age 72, a male's vulnerability, relative to a 30-year old female, is 1.14 times higher. Third, the distribution of and disparities in the relative vulnerability to severe illness from COVID-19 in the U.S. adult population reported below will be driven by the marginal contributions of individual risk factors displayed in Figure 1 based on the DeCaprio et al. model and by the prevalence of these risk factors in the population.

RESULTS

In this next section, we use the data from the PSID to document the prevalence of health-related risk factors and VI_i in the U.S. population by age and examine how these risk factors are differentially distributed by race-ethnicity and measures of individuals' socioeconomic status.

Prevalence of Health-Related Risk Factors in the United States and their Disparities

In Table 2, we present the prevalence of the health-related risk factors and how they vary by age, across demographic groups and by the socioeconomic status of adults in the U.S. population. We test whether differences across age, race-ethnicity, education, and household income are statistically significant and denote differences from the base category (age 25-44, NH white, BA or more, top income quartile) with asterisks. The risk factors are listed in order of their prevalence in the overall U.S. population based on the PSID data.

As one can see, hypertension is the most prevalent risk factor in the United States, with one third of the population having reported that they have this condition. Furthermore, the prevalence of hypertension increases with age, with 59% of those 65 and older having been diagnosed with hypertension. The next most prevalent risk factor is severe obesity, with 17% of the U.S. population having a BMI greater than or equal to 40. Unlike hypertension, the prevalence of severe obesity varies little by age. Diabetes and asthma are the third and fourth most prevalent risk factors in the population at 12% and 11%, respectively. The prevalence of diabetes increases with age, with only 4% of those in early adulthood (25-44 years old) having been diagnosed with diabetes but rising

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

to almost one-quarter (23%) of those over 65 having this disease. In contrast, the prevalence of asthma declines with age, going from 12% for the 25-34 year old group to 9% for those 65 and older. Some 9% of the adult population has had (or still has) some form of cancer and its incidence rises with age so that 1 in 5 (21%) of those over 65 reported having some form of cancer. Among the remaining health-related risk factors (lung disease, heart disease, heart attack, stroke, a neurological condition and a chronic kidney disorder), the prevalence ranges from 1% to 6%, with the prevalence of all of these conditions increasing with age. Finally, 10% of adults in the United States reported having an in-patient hospitalization in the preceding year of one night or more, and this rises to 17% of those 65 years old or older.

As the other columns in Table 2 make clear – and is well known – there is an unequal distribution of these health-related risk factors by race-ethnicity, educational attainment and income. Compared to NH whites, NH black adults have higher prevalence of almost all of the risk factors for COVID-19, with most of these differences being statistically significant. For example, NH blacks are almost one third more likely to have been diagnosed with hypertension than NH whites (44% vs 34%). Cancer is the only risk factor that is more prevalent for NH whites than NH blacks (10% for NH whites vs 6% for NH blacks). On the other hand, prevalence rates for the risk factors listed in Table 2 are lower for Hispanic adults than NH whites (or NH blacks). The prevalence of hypertension is 24% for Hispanics compared to 34% for NH whites. These racial differences in prevalence are well-known and have been documented in many other studies[10] (Hertz et al, 2005; Hicken et al., 2014; NCHS, 2018).

The prevalence of risk factors for COVID-19 is also quite different by education. Adults in the United States with a high school education or less have higher prevalence rates for almost all of the risk factors listed in Table 2 than do adults with a BA degree or more. Almost 40% of those with a high school degree or less suffer from hypertension compared to one quarter (26%) of adults with college degrees. The two notable exceptions to this pattern of inequality across educational attainment are cancer and chronic kidney disease, where prevalence is about the same for the two educational groups.

We also see substantial differences in risk factor prevalence by household income. Adults

[10] NCHS (2018) reports that the prevalence of hypertension is higher among Hispanics than non-Hispanic whites. This is based on age-adjusted rates. Crude rates are higher for Hispanics. The crude rates that we report are nearly identical to the crude rates reported in NCHS (2018) Table 22.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

in the bottom quartile of household income have higher prevalence rates for every risk factor compared to those adults in the top quartile of income. These differences are sizeable. Compared to those in the top income quartile, those in the bottom quartile are around twice as likely to have been diagnosed with hypertension (25% vs 46%), diabetes (8% vs 18%) and asthma (9% vs 15%), are around three times more likely to have had a heart attack (2% vs 7%) or a stroke (2% vs 7%) and are six times more likely to suffer from lung disease (2% vs 12%). And, those in the lowest income group are three times more likely than those in the highest income group to have had an in-patient hospitalization in the past year (6% vs 18%).

Table 2 also shows that the prevalence of each risk factor increases with age. The distribution of the total number of risk factors that individuals have also changes with age. In Figure 2, we display proportions of the population with 0, 1, 2 and 3 or more of the risk factors listed in Table 1 across three age groups. Among all adults, 41% have 0 risk factors, while almost 15% have 2 and another 14% have 3 or more, for an average of 1.14 risk factors in the U.S. adult population. The number of risk factors increases across age groups, with 57% of adults 24-44 having no risk factors while only 18% of those 65 years and older having none. Only 4% of younger adults have 3 or more risk factors, 30% of those 65 and older have 3 or more.

We showed that the distribution of age varies across race-ethnicity, education, and income groups in our sample. For this reason, in Tables 3 through 5, we examine the prevalence of each of the risk factors by race-ethnicity, socioeconomic status, and age and test whether differences in the prevalence of conditions across race-ethnicity and socioeconomic status are statistically significant for each age group. In Table 3, we show that differences in the prevalence of hypertension between NH blacks and NH whites emerge early in life. The differences are statistically significant in every age group, but large gaps appear in midlife. At age 45-64, 55% of NH blacks report having been diagnosed with hypertension compared to only 34% of NH whites. Large differences in the prevalence of diabetes between NH blacks and NH whites also emerge at age 45-64. In contrast, the health advantage of Hispanics in terms of hypertension occurs mainly at younger ages and differences between NH whites and Hispanics are not statistically significant at older ages. Tables 4 and 5 show similar patterns and similarly large disparities for less-educated and low-income Americans as we see for NH blacks. Health disparities become large in midlife, especially for the most common conditions like hypertension, obesity, and diabetes. However, Tables 4 and 5 show

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

evidence that there are statistically significant differences by education and income in the prevalence of nearly all health conditions, even among those age 25-44.

In Figure 3, we present the distribution of the number of risk factors for the overall population and by age across race-ethnic groups, educational attainment, and household income, using the same format as Figure 2. Consistent with the disparities by race and ethnicity in individual risk factors, among all adults (ages 25 and older), NH blacks are more likely to have 3 or more risk factors compared to NH whites (19% vs 15%), while only 9% of Hispanics have such a high number (Figure 3(a), 25+ age group). Among adults of all ages with 12 years of schooling or less, almost 1 in 5 (19%) have 3 or more risk factors, compared to just 8% of those with a college degree (Figure 3(b), 25+ age group). Finally, the differences in numbers of risk factors are particularly dramatic by income, with 26% of adults of all ages in the bottom quartile of household income having 3 or more risk factors compared to only 7% among those in the top income quartile (Figure 3(c), 25+ age group).

Figure 3 also shows the distribution of the number of risk factors by age across race-ethnicity, education, and household income. Consistent with Tables 3 through 5, Figure 3 shows that for NH blacks, those with 12 years of schooling or less, and those in the bottom quartile of income, risk factors accumulate at earlier ages. For example, by age 45-64, 43% of NH blacks have two or more risk factors compared to only 28% of NH whites. On average, adults, age 45-64, with household incomes in the top income quartile have fewer conditions than those, age 25-44, in the bottom income quartile and those in the oldest group, age 65+, have fewer conditions than those who are middle age (45-64) and have incomes in the bottom quartile. All differences between groups are statistically significant.

Disparities in Relative Vulnerability to Severe Illness in the U.S.

The dramatic differences in the prevalence of specific health-related risk factors and the distribution of the number of risk factors by race-ethnicity, education and income clearly suggest the potential for dramatic disparities in the vulnerability to severe illness due to COVID-19. Yet such disparities in health are likely to affect the susceptibility to a wide variety of health-related outcomes and overall morbidity. To assess how these disparities in health-related risks translate to vulnerability to severe illness from COVID-19, we use the DeCaprio et al. model relative vulnerability index (Equation 4) to translate the individual risk factors and age of individuals in our sample into the relative vulnerability to severe illness from COVID-19 of the U.S. population. We

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

then examine how this vulnerability varies by age and across race-ethnicity and socioeconomic status to provide the first assessment of the distribution and disparities of this vulnerability at the population level.

In Figure 4, we display average values of the relative vulnerability index by numbers of risk factors for the three age groups, 25-44, 45-64 and 65+, and the overall population of adults, over the age of 25. Recall that what is being measured in the index (VI) is the odds of having severe illness from COVID-19 relative to a 30 year-old female with no risk factors. The average value of the relative vulnerability index in the U.S. population of adults is 16.6 (not shown in Figure 4). But as can be seen in Figure 4, relative vulnerability varies substantially by number of risk factors and by age. Among all ages (the 25+ group), those with no risk factors are 2.6 times more vulnerable to severe illness from COVID-19 compared to a 30 year-old female. But as the number of risk factors increases, the relative vulnerability of severe illness increases, with those with 3 or more risk factors having, on average, 82.4 times greater odds of becoming severely ill from COVID-19. On average individuals in this high-risk group have 3.82 risk factors and 10% have 5 or more risk factors.

Figure 4 also makes clear that the increasing numbers of risk factors for those at older ages displayed in Figure 2, result in a progressive increase in relative vulnerability to becoming severely ill from COVID-19 across the age groups. For example, individuals who are 65 years old or older and who have 3 or more risk factors are 111.7 times more likely to become severely ill from COVID-19 than a 30 year-old healthy (no risk factors) female. Among individuals in this high-risk group, 25% have 5 or more risk factors and half are age 75 or older. While this estimate of relative vulnerability is extraordinarily high, the relative magnitudes of the differences by age and number of risk factors clearly illustrate how the accumulation of risk factors with age makes many older Americans much more vulnerable to COVID-19.

Figure 4 also illustrates how having multiple risk factors heightens the vulnerability to severe illness from COVID-19 at any age. For example, those 65 or older with no risk factors (18.1% of that age group) have an average vulnerability that is 6.3 times that of the reference person (30 year-old female with no risk factors), which is almost the same relative vulnerability to severe illness from COVID-19 as for 25-44 year-olds with 2 risk factors (7.4 relative vulnerability and 21.8% of the younger age group). In fact, the relative vulnerability of those 65 and older with no risk factors is more than 4 times lower than the average relative vulnerability of 25-44 year olds

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

with 3 or more risk factors, 28.1. (We note that younger adults with 3 or more risk factors constitutes 7% of those 25-34 years old). This comparison illustrates both the disturbing consequences of younger adults having multiple risk factors and the apparent benefits to older age adults of having fewer risk factors with respect to vulnerability of becoming severely ill from the Coronavirus.

The results in Figure 4 for relative vulnerability to severe illness from COVID-19 represent the differences in the distributions of numbers of risk factors for different age groups *and* the impact that particular risk factors have on relative vulnerability at different ages based on the DeCaprio et al. model displayed in Figure 1. Recall for example that hypertension and lung disease consistently have the largest impacts on the relative vulnerability scores at all ages and that the impacts of both went up markedly with age. At the same time, as we noted in our discussion of Table 2, the prevalence of hypertension on the U.S. population was high (33%) and increased across age groups, while, relatively speaking, the incidence of lung disease was much lower (6%) and increased less markedly with age. Thus, the relative vulnerability index allows one to characterize the combined impact of (and differences in) prevalence of risk factors and their impact on population-level vulnerability to severe illness from the Coronavirus than simply looking at the prevalence of risk factors can provide. While it would be tempting to try to isolate the contributions of particular risk factors to the vulnerability in the population, such an analysis would require modeling which is beyond the scope and focus of this paper. The primary complication to attributing the differences in vulnerability to particular risk factors is that individuals often have multiple conditions. For example, those who reported that they had been diagnosed with hypertension had, on average, 1.4 other risk factors, while those who did not have hypertension had an average of 0.5 other risk factors.

In Figure 5 we display the average relative vulnerability scores by age for race-ethnic, education, and income groups in the adult population. Recall the differences in number of risk factors among NH whites, NH blacks and Hispanics displayed in Tables 2 and 3 and Figure 3(a). We found that NH blacks had more risk factors, on average that NH whites (Table 2), that differences in the prevalence of hypertension between NH blacks and NH whites grow large by age 45-64 (Table 3), and that NH blacks had a higher number of risk factors after age 45 than NH whites (Figure 4(a)). We also found that the prevalence of risk factors was lower among Hispanics, particularly at younger ages. As shown in Figure 5(a), this same pattern across racial and ethnic groups

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

holds for the relative vulnerability to severe illness from the Coronavirus. By age 45, sizeable differences in relative vulnerability to severe illness emerge between NH blacks and NH whites and these gaps remain large at older ages. Hispanics have lower relative vulnerability at all ages. Across the whole sample (25+), the difference in relative vulnerability between NH blacks and NH whites is muted. This is due to the age distributions we show in Table 1 where those age 65 or older make up 28.2% of the population of NH whites but only 16.8% of the population of NH blacks.

Figure 5(b) has the same format for showing the differences in average relative vulnerability between those adults with 12 or fewer years of education and those who are college educated. On average, the relative vulnerability to severe illness from the Coronavirus is twice as high for those with high school or less (25.8) than for those with a college degree (12.7) and this sizeable gap by education is present for each age group, reflecting, in part, the differences in incidence of number of risk factors displayed in Figure 4(b).

Finally, Figure 5(c) presents the differences in relative vulnerability for those in the bottom and top quartiles of household income. Among all adults (ages 25+), there is more than a 3-fold difference in the relative vulnerability for severe illness from COVID-19, with those in the bottom quartile of income having 33.8 higher odds of severe illness compared to the reference person, while for those in the top income quartile, the odds are only 9.3 higher. This gap by income in relative vulnerability is even more pronounced for middle-age adults (27.5 vs 8.4). Again, the gap reflects the sizeable differences in prevalence of risk factors (Figure 4(c)) by income. These differences contribute significantly to the dramatic differences in the odds of severe illness from the COVID-19 pandemic in the United States.

DISCUSSION

A reasonable question to ask is whether the segment of the U.S. population that the DeCaprio et al. model's index suggests are more vulnerable to severe illness from COVID-19 resemble those who, in fact, have become severely ill from the virus and required hospitalization. While, the pandemic appears to still be in progress in the United States and the information about the characteristics of those who have been hospitalized is limited, we report on an exercise below that attempts a tentative answer to this question.

In this exercise, we identify the subsample of individuals who all have a high relative vulnerability score based on the DeCaprio et al. model index. To do this, we rank order all individuals

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

in the PSID data according to their vulnerability index (*VI*). We choose the subsample of all individuals who were in the top 55% of our rank-ordered sample, i.e., those who are above the 45th percentile of the vulnerability index in our data. We chose this threshold because of the similarity between the characteristics of individuals above this threshold and the characteristics of those hospitalized from COVID-19 (CDC, 2020b) on dimensions including the rate of hypertension, the percent with zero health conditions, and the percent age 65 and older. Using alternative thresholds around the point produces similar conclusions.

Based on our analysis above, we know that this subsample consists of adults who are older and have more risk factors, on average, than the overall population. As of this writing, the rates of hypertension, zero underlying health conditions, and the percent of people age 65 and over among those hospitalized from COVID-19 were 58%, 7.9% and 40.9%, respectively (CDC, 2020b). In our subsample from the PSID in the top 55% of the vulnerability index distribution, the corresponding rates of hypertension, zero underlying health conditions, and percent over 65 are 59.3%, 10.7%, and 43.6%, respectively.

Furthermore, the segment of our sample that is in the top 55% of the vulnerability index is markedly different that the overall population of adults in the United States. In particular, the average age among individuals in the top 55% of vulnerability is almost 10 years older than the overall population (61.9 years old vs 52 years old in the overall population) and there are many fewer young adults age 25-44 (12.2% vs 37.3%) than the overall population. Those in the top 55% of the vulnerability index are more likely to have 3 or more risk factors than the overall population (26% vs 14.2%) and less likely to have zero health conditions (11.1% vs 40.9%). Among younger adults (25-44) in our high-risk sample, 22.7% have 3 or more risk factors compared to only 4.1% of the overall population of young adults. In short, those in the top 55% of the vulnerability index are more likely to have those risk factors that the CDC has indicated place individuals at higher risk of complications from the COVID-19.

Finally, we note that the average vulnerability index score for individuals in the top 55% of vulnerability is 30.9 which is almost double the average score of 17.8 for the overall population. Comparing this score with the scores we calculate in Figure 5, the average vulnerability score in this high-risk group is very similar to the average of adults age 45-64 from the lowest income quartile in the overall population. That is, the average vulnerability index of low-income adults in midlife is roughly the same as that of a sample of individuals whose characteristics match those

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

hospitalized with COVID-19. The average vulnerability of NH blacks, adults with a high school degree or less, and low-income adults age 65+ is nearly double this level of risk.

While the above comparison is certainly limited in what it can say about the predictive validity of the vulnerability index for the estimating the actual hospitalization rates from COVID-19, it does suggest that the index certainly appears to be a reasonable guide for them. And, coupled with population characteristics beyond those currently available in existing data, like race, and especially education and income, our analyses make clear that hospitalizations are likely to be very unequally distributed across the U.S. population.

CONCLUSION

This paper provides the first nationally representative estimates of vulnerability to severe complications from COVID-19 overall and across race-ethnicity and socioeconomic status. We use the PSID to examine the prevalence of specific health conditions associated with complications from COVID-19 and to calculate, for each individual, an index of the risk of severe complications from respiratory infections developed by DeCaprio et al. (2020). We show large disparities across race-ethnicity and socioeconomic status in the prevalence of conditions, including hypertension, which are associated with adverse outcomes from COVID-19, and in the overall risk of severe complications. Moreover, we show that these disparities emerge early in life, prior to age 65.

Our results suggest particular attention should be paid to the risk of adverse outcomes in midlife for non-Hispanic blacks, adults with a high school degree or less, and low-income Americans. These results are especially important as states and municipalities start to reopen businesses and public services. Mongey, Pilosoph, & Weinberg (2020) showed that disadvantaged groups are less likely to be able to socially distance at work. The evidence that we present shows large disparities in underlying health conditions among working-age adults for these same groups. Combined, these results suggest that localities with predominantly non-white populations and those with high levels of poverty or a high concentration of less-educated adults face potentially devastating effects of the virus if there are high rates of exposure.

Several caveats are important. The first is that we are likely understating the disparities in the risk for severe complications from COVID-19 by race-ethnicity, education, and income based on underlying health conditions for the reasons we outlined in the Background section. However, even as a lower bound, the disparities in risk we find are very large. Second, we have not yet looked at the disparities in the risk of severe complications from COVID-19 across interactions of

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

race-ethnicity and education or race-ethnicity and income. There is ample evidence of large mortality gaps across these groups (Case & Deaton, 2020), and this is an area for more extensive exploration. Finally, our analysis of disparities in the risk of severe complications from underlying health conditions does not capture the true risk of hospitalization and death from COVID-19 because it does not account for factors influencing exposure to the virus or access to high quality care. All of these factors, including the underlying health conditions we examine, are influenced by systemic inequalities in society and in our health care system. Such inequalities will make it difficult to isolate the influence of health conditions on disparities in the overall effect of COVID-19 pandemic on the U.S. population.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

REFERENCES

Arentz, M.,Yim, E., Klaff, L., et al. (2020). Characteristics and outcomes of 21 critically ill patients with COVID-19 in Washington State. *JAMA*. Published online March 19, 2020. doi:10.1001/jama.2020.4326.

Azar, K. M., Zijun S., Romanelli, R. J., Lockhart, S. H., Smits, K., Robinson, S., Brown, S., & Pressman, A. R. (2020). Disparities In Outcomes Among COVID-19 Patients In A Large Health Care System in California. *Health Affairs* 39(7). doi.org/10.1377/hlthaff.2020.00598 .

Barcellos, S. H., Goldman, D. P., & Smith, J. P. (2012). Undiagnosed Disease, Especially Diabetes, Casts Doubt On Some Of Reported Health 'Advantage' Of Recent Mexican Immigrants. *Health Affairs* 31(12), 2727–37. doi.org/10.1377/hlthaff.2011.0973

Bhatraju, P. K., Ghassemieh, B. J., Nichols, M., Kim, R., Jerome, K. R., Nalla, A. K., Greninger, A. L., Pipavath, S., Wurfel, M. M., Evans, L., Kritek, P. A., & West, T. E. (2020). Covid-19 in Critically Ill Patients in the Seattle Region – Case Series. *New England Journal of Medicine*, 382, 2012-20122. doi:10.1056/NEJMoa2004500

Blumenshine, P., Reingold, A., Egerter, S., Mockenhaupt, R., Braveman, P., & Marks, J. (2008). Pandemic Influenza Planning in the United States from a Health Disparities Perspective. *Emerging Infectious Diseases* 14(5), 709–15. doi.org/10.3201/eid1405.071301.

Borjas, G. J. (2020). Demographic Determinants of Testing Incidence and COVID-19 Infections in New York City Neighborhoods. *NBER Working Paper No. 26952*.

Case, A., & Deaton, A. (2020). *Deaths of Despair and the Future of Capitalism*. Princeton University Press.

Case A., & Deaton, A. (2015). Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. *Proceedings of the National Academy of Sciences* 112(49), 15078-15083. doi.org/10.1073/pnas.1518393112.

Center for Disease Control and Prevention. (2020a, May 20). Weekly Updates by Select Demographic and Geographic Characteristics, Table 2b. https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm.

Center for Disease Control and Prevention. (2020b). COVID-19 Laboratory-Confirmed Hospitalization. Preliminary data as of May 16, 2020. https://gis.cdc.gov/grasp/COVID-Net/COVID19_5.html.

Chatterji, P., Joo, H., & Lahiri, K. (2012). Racial/Ethnic- and Education-Related Disparities in the Control of Risk Factors for Cardiovascular Disease Among Individuals With Diabetes. *Diabetes Care* 35(2), 305–12. doi.org/10.2337/dc11-1405.

Chetty, R., Stepner, M., Abraham, S., Lin, S., Scuderi, B., Turner, N., Bergeron, A., & Cutler, D. (2016). The Association Between Income and Life Expectancy in the United States, 2001-

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2014. *JAMA* 315(16), 1750-1766. doi:10.1001/jama.2016.4226.

Chobanian, A. V. (2009). The Hypertension Paradox — More Uncontrolled Disease despite Improved Therapy. *New England Journal of Medicine* 361(9), 878–87. doi.org/10.1056/NEJMsa0903829.

Cummings, M. J., Baldwin, M. R., Abrams, D., Jacobson, S. D., Meyer, B. J., Balough, E. M., Aaron, J. G., Claassen, J., Rabbani, L. E., Hastie, J., Hochman, B. R., Salazar-Schicchi, J., Yip, N. H., Brodie, D., & O'Donnell, M. R. (2020). Epidemiology, Clinical Course, and Outcomes of Critically Ill Adults with COVID-19 in New York City: A Prospective Cohort Study. (2020). *The Lancet*, published online May 19, 2020, S0140673620311892. doi.org/10.1016/S0140-6736(20)31189-2.

Currie, J., & Schwandt, H. (2016). Inequality in mortality decreased among the young while increasing for older adults, 1990-2010. *Science* 352(6286), 708-712. DOI:10.1126/science.aaf1437.

DeCaprio, D., Gartner, J., McCall, C. L., Burgess, T., Kothari, S., & Sayed, S. (2020). Building a COVID-19 Vulnerability Index. medRxiv.org; doi.org/10.1101/2020.03.16.20036723 .

Emami, A., Javanmardi, F., Pirbonyeh, N., & A. Akbari. (2020). Prevalence of Underlying Diseases in Hospitalized Patients with COVID-19: a Systematic Review and Meta-Analysis. *Archives of Academic Emergency Medicine* 8(1): e35. doi.org/10.22037/aaem.v8i1.600.g748.

Freedman, V. A., McFall, B. H., & Ryan, L. (2019). Adding the AD8 Dementia Screen to the Panel Study of Income Dynamics. *PSID Technical Series Paper #19-01*. University of Michigan, Ann Arbor

Garg S, Kim L, Whitaker M, et al. (2020). Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 – COVID-NET, 14 States, March 1–30, 2020. *Morbidity and Mortality Weekly Report (MMWR)*. 2020; 69:458–464. dx.doi.org/10.15585/mmwr.mm6915e3.

Glezen, W. P., Greenberg, S. B., Atmar, R. L., Piedra, P. A., & Couch, R. B. (2020). Impact of Respiratory Virus Infections on Persons With Chronic Underlying Conditions. *JAMA* 283(4), 499–505. doi.org/10.1001/jama.283.4.499.

Grasselli, G., Zangrillo, A., Zanella, A., Antonelli, M., Cabrini, L., Castelli, A., Cereda, D., Coluccello, A., Foti, G., Fumagalli, R., Iotta, G., Latronico, N., Lorini, L., Merler, S., Natalini, G., Piatti, A., Ranieri, M. V., Scandroglio, A. M., Storti, E., Cecconi, M., & Pesenti, A. (2020). Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA* 323(16), 1574-1581. doi:10.1001/jama.2020.5394.

Harris, K. M. & McDade, T. W. (2018). The Biosocial Approach to Human Development, Behavior, and Health Across the Life Course. *RSF: The Russell Sage Foundation Journal of the Social Sciences* 4(4), 2-26. doi.org/10.7758/RSF.2018.4.4.01.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Hertz, R. P., Unger, A. N., Cornell, J. A., & Saunders, E. (2005). Racial Disparities in Hypertension Prevalence, Awareness, and Management. *Archives of Internal Medicine* 165(18), 2098-2104. doi:10.1001/archinte.165.18.2098.

Hicken, M. T., Lee, H., Morenoff, J., House, J. S., & Williams, D. R. (2014). Racial/ethnic disparities in hypertension prevalence: reconsidering the role of chronic stress. *American journal of Public Health*, 104(1), 117–123. doi.org/10.2105/AJPH.2013.301395.

Insolera, N. E. & Freedman, V. A. (2017). Comparing Health Estimates in the PSID and NHIS, 2001–2015. Institute for Social Research, University of Michigan, Technical Series Paper #17-01.

Kumar, S., & Quinn, S. C. (2012). Existing Health Inequalities in India: Informing Preparedness Planning for an Influenza Pandemic. *Health Policy and Planning* 27(6), 516–26. doi.org/10.1093/heapol/czr075.

Lowcock, E. C., Rosella, L. C., Foisy, J., McGeer, A. & Crowcroft, N. (2012). The Social Determinants of Health and Pandemic H1N1 2009 Influenza Severity. *American Journal of Public Health* 102(8), e51–58. doi.org/10.2105/AJPH.2012.300814.

Ma., C., Gu, J., Hou, P., Zhang, L., Bai, Y., Guo, Z., Wu, H., Zhang, B., Li, P., Zhao, X. (2020). Incidence, clinical characteristics an dprognostic factor of patients with COVID-19: a systematic review and meta-analysis, medRxiv.. doi.org/10.1101/2020.03.17.20037572.

McGonagle, K., Schoeni, R. F., Sastry, N., & Freedman, V. A. (2012) The Panel Study of Income Dynamics: Overview, recent renovations, and potential for life course research. *Longitudinal and Life Course Studies* 3, 268-284. dx.doi.org/10.14301/llcs.v3i2.188.

Mead, H., L. Cartwright-Smith, K. Jones, C. Ramos, B. Siegel, & K. Woods. (2005). Racial and Ethnic Disparities in U.S. Health Care: A Chartbook. The Commonwealth Fund Pub. No. 1111. https://www.commonwealthfund.org/sites/default/files/documents/_usr_doc_Mead_racialethnicdisparities_chartbook_1111.pdf

Mongey, S., Pilossoph, L., & Weinberg, A. (2020). Which Workers Bear the Burden of Social Distancing Policies? *NBER Working Paper No. 27085*.

National Center for Health Statistics. (2018). *Health, United States, 2018.* Hyattsville (MD): National Center for Health Statistics.

National Center for Health Statistics. (2012). *Health, United States, 2011: With Special Feature on Socioeconomic Status and Health.* Hyattsville (MD): National Center for Health Statistics (US); 2012 May. Report No.: 2012-1232.

Pleasants, R. A., Riley, I. L., & Mannino, D. M. (2016). Defining and Targeting Health Disparities in Chronic Obstructive Pulmonary Disease. *International Journal of Chronic Obstructive Pulmonary Disease* 11, 2475–96. doi.org/10.2147/COPD.S79077.

Richardson, S., Hirsch, J. S., Narasimhan, M., Crawford, J. M., McGinn, T., Davidson, K. W., and

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

the Northwell COVID-19 Research Consortium. (2020). Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. *JAMA*. Published online April 22, 2020. Corrected on April 24, 2020. doi.org/10.1001/jama.2020.6775.

Scodia, J. A., (2020). Epidemiology and clinical features of COVID-19: A review of current literature. *Journal of Clinical Virology* 127 (104357). doi.org/10.1016/j.jcv.2020.104357.

Smith, J. P. (2007). Nature and Causes of Trends in Male Diabetes Prevalence, Undiagnosed Diabetes, and the Socioeconomic Status Health Gradient. *Proceedings of the National Academy of Sciences* 104(33), 13225–31. doi.org/10.1073/pnas.0611234104.

Yang, J., Zheng, Y., Gou, X., Pu, K., Chen, Z., Guo, Q., Ji, R., Wang, H., Wang, Y., Zhou, Y. (2020). Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *International Journal of Infectious Diseases* 94, 91-95. doi.org/10.1016/j.ijid.2020.03.017.

Wu Z, McGoogan JM. (2020) Characteristics of and Important Lessons from the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA.* 2020; 323(13):1239–1242. doi:10.1001/jama.2020.2648.

Zallman, L., Himmelstein, D. H., Woolhandler, S., Bor, D. H., Ayanian, J. Z., Wilper, A. P., & McCormick, D. (2013). Undiagnosed and Uncontrolled Hypertension and Hyperlipidemia among Immigrants in the US. *Journal of Immigrant and Minority Health* 15(5), 858–65. doi.org/10.1007/s10903-012-9695-2.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 1. Proportions of Demographic and SES Groups by Age across Age Categories & Mean Age

	Prop. across Age Categories			Prop. within Age Categories					
	25-44	45-64	65+	25-44	45-64	65+	Prop., All Ages (25+)	Mean Age	Sample sizes
Overall				0.373	0.383	0.244		52.0	13,529
Gender									
Female	0.362	0.381	0.256	0.516	0.528	0.557	0.531	52.5	7,420
Male	0.385	0.385	0.230	0.484	0.472	0.443	0.469	51.5**	6,109
Race and Ethncity									
NH White	0.331	0.387	0.282	0.632	0.719	0.825	0.712	53.7	7,252
NH Black	0.416	0.416	0.168	0.120	0.117	0.074	0.107	49.4***	4,274
Hispanic	0.509	0.360	0.131	0.180	0.124	0.071	0.132	46.7***	1,550
NH Other	0.523	0.322	0.154	0.068	0.041	0.031	0.049	47.4	453
Education									
HS or less	0.312	0.405	0.283	0.324	0.408	0.448	0.387	54.4***	5,450
Some college	0.388	0.385	0.227	0.245	0.237	0.219	0.235	51.2**	3,577
BA or more	0.425	0.360	0.215	0.431	0.355	0.333	0.378	50.1	4,502
Household Income									
Bottom Quartile	0.307	0.328	0.365	0.151	0.157	0.275	0.183	56.7***	2,651
Second Quartile	0.393	0.312	0.295	0.239	0.185	0.274	0.227	52.4***	3,269
Third Quartile	0.409	0.363	0.228	0.301	0.261	0.257	0.275	50.4**	3,757
Top Quartile	0.366	0.484	0.150	0.309	0.397	0.194	0.315	50.4	3,852
Sample sizes	6,962	4,552	2,015						

Data Source: 2017 Wave, PSID. Sample: PSID heads and spouses, 25 years and older. Weights: PSID individual cross-section weights.

Notes: Stars (*) are for tests of risk factors being of a subgroup being significantly different from base group, where: * P-value ≤ 0.10; ** P-value ≤ 0.05; *** P-value ≤ 0.01.

Tests of differences in age distributions: (a) Female vs Male: p = 0.0218; (b) NH Blacks vs NH Whites: p = 0.000; [c] Hispanics vs NH Whites: p = 0.000; (d) HS or less vs BA or more: p = 0.0000; (e) Some Coll vs BA or more: p = 0.0429; (d) Second income quartile vs Highest quartile: p = 0.0000; (f) Third income quartile vs Highest quartile: p = 0.0000.

Table 2. Age, Race/Ethnicity, and SES in prevalence of health-related risk factors

		Age			Race-ethnicity			Education		Household Income	
Risk factor	Overall	25-44	45-64	65+	NH White	NH Black	Hispanic	HS or less	BA or more	Bottom Quartile	Top Quartile
Hypertension	0.33	0.14	0.36***	0.59***	0.34	0.44***	0.24***	0.39***	0.26	0.46***	0.25
Severe obesity	0.17	0.16	0.18*	0.16	0.16	0.19***	0.20***	0.19***	0.14	0.18***	0.13
Diabetes	0.12	0.04	0.14***	0.23***	0.11	0.16***	0.14**	0.15***	0.09	0.18***	0.08
Asthma	0.11	0.12	0.10**	0.09***	0.11	0.14**	0.09**	0.11	0.10	0.15***	0.09
Hospitalization	0.10	0.07	0.09**	0.17***	0.10	0.12*	0.09	0.13***	0.07	0.18***	0.06
Cancer	0.09	0.02	0.07***	0.21***	0.10	0.06***	0.05***	0.09	0.09	0.10**	0.08
Lung disease	0.06	0.03	0.05***	0.10***	0.06	0.07	0.03***	0.08***	0.03	0.12***	0.02
Heart disease	0.06	0.01	0.05***	0.14***	0.06	0.07	0.02***	0.07***	0.04	0.11***	0.03
Heart attack	0.04	0.01	0.04***	0.10***	0.05	0.05	0.02***	0.06***	0.02	0.07***	0.02
Stroke	0.03	0.01	0.03***	0.08***	0.03	0.06***	0.02	0.05***	0.02	0.07***	0.02
Neurological conditions	0.03	0.01	0.02***	0.05***	0.03	0.02**	0.01***	0.04***	0.01	0.06***	0.00
Chronic Kidney Disorder	0.01	0.00	0.01**	0.01***	0.01	0.01	0.00**	0.01	0.01	0.01*	0.00
Sample Sizes	13,529	6,962	4,552	2,015	7,252	4,274	1,550	5,450	4,502	2,651	3,852

Data Source: 2017 Wave, PSID. Sample: PSID heads and spouses, 25 years and older. Weights: PSID individual cross-section weights. Income quartiles are determined across households using family weights.

Stars (*) are for tests of risk factors being of a subgroup being significantly different from base group, where: * P-value ≤ 0.10; ** P-value ≤ 0.05; *** P-value ≤ 0.01.

Table 3. Prevalence of risk factors by Race and Age

	Race-ethnicity								
	NH White			NH Black			Hispanic		
Risk factor	25-44	45-64	65+	25-44	45-64	65+	25-44	45-64	65+
Hypertension	0.14	0.34	0.56	0.18***	0.55***	0.78***	0.11*	0.27***	0.63
Severe obesity	0.15	0.17	0.16	0.18	0.21**	0.17	0.19**	0.23**	0.16
Diabetes	0.04	0.12	0.20	0.05	0.21***	0.32***	0.04	0.18***	0.42***
Asthma	0.13	0.11	0.09	0.15	0.14*	0.09	0.10**	0.07**	0.09
Hospitalization	0.07	0.09	0.16	0.07	0.13***	0.22*	0.08	0.07	0.19
Cancer	0.03	0.09	0.21	0.01***	0.04***	0.22	0.02	0.02***	0.21
Lung disease	0.03	0.06	0.11	0.05**	0.07	0.12	0.02	0.04	0.05**
Heart disease	0.01	0.05	0.14	0.03**	0.06	0.18	0.00	0.02***	0.05***
Heart attack	0.01	0.04	0.11	0.01	0.06*	0.12	0.00**	0.03	0.07
Stroke	0.01	0.03	0.07	0.02*	0.06***	0.15**	0.00*	0.03	0.11
Neurological conditions	0.01	0.03	0.05	0.01	0.02	0.04	0.00***	0.01***	0.07
Chronic Kidney Disorder	0.00	0.01	0.01	0.01	0.01	0.02	0.00	0.00	0.00***
Sample Sizes	3,432	2,419	1,401	2,318	1,543	413	969	447	134

Data Source: 2017 Wave, PSID. Sample: PSID heads and spouses, 25 years and older. Weights: PSID individual cross-section weights.

Stars (*) are for tests of risk factors being of a subgroup being significantly different from base group, where: * P-value ≤ 0.05; ** P-value ≤ 0.01; *** P-value ≤ 0.001.

Table 4. Prevalence of risk factors by Educ and Age

	Education					
	HS or less			BA or more		
Risk factor	25-44	45-64	65+	25-44	45-64	65+
Hypertension	0.17***	0.40***	0.62***	0.11	0.29	0.53
Severe obesity	0.19***	0.20***	0.16	0.13	0.14	0.13
Diabetes	0.04	0.17***	0.26***	0.04	0.10	0.16
Asthma	0.12	0.10	0.10*	0.11	0.10	0.08
Hospitalization	0.09***	0.11***	0.19***	0.06	0.05	0.13
Cancer	0.02	0.06*	0.21	0.03	0.08	0.22
Lung disease	0.04***	0.08***	0.13***	0.01	0.02	0.06
Heart disease	0.01***	0.06***	0.15*	0.00	0.03	0.12
Heart attack	0.01***	0.05***	0.12**	0.00	0.01	0.08
Stroke	0.01**	0.04***	0.09*	0.00	0.01	0.06
Neurological conditions	0.02***	0.03***	0.07**	0.00	0.01	0.04
Chronic Kidney Disorder	0.00	0.01	0.01*	0.00	0.01	0.03
Sample Sizes	2,492	2,017	941	2,492	1,378	632

Data Source: 2017 Wave, PSID. Sample: PSID heads and spouses, 25 years and older.
Weights: PSID individual cross-section weights.

Stars (*) are for tests of risk factors being of a subgroup being significantly different from base group, where: * P-value ≤ 0.10; ** P-value ≤ 0.05; *** P-value ≤ 0.01.

Table 5. Prevalence of risk factors by Household Income and Age

	Household Income					
	Bottom Quartile			Top Quartile		
Risk factor	25-44	45-64	65+	25-44	45-64	65+
Hypertension	0.17**	0.49***	0.67***	0.13	0.29	0.46
Severe obesity	0.19***	0.18*	0.18*	0.11	0.14	0.13
Diabetes	0.06***	0.21***	0.27***	0.03	0.09	0.14
Asthma	0.16***	0.19***	0.11*	0.10	0.08	0.08
Hospitalization	0.12***	0.18***	0.22***	0.05	0.05	0.11
Cancer	0.03	0.08	0.19	0.03	0.08	0.21
Lung disease	0.05***	0.16***	0.13***	0.01	0.02	0.06
Heart disease	0.03***	0.10***	0.18***	0.01	0.02	0.10
Heart attack	0.01	0.06***	0.13**	0.01	0.02	0.08
Stroke	0.02**	0.06***	0.11***	0.00	0.02	0.05
Neurological conditions	0.03***	0.08***	0.07***	0.00	0.00	0.01
Chronic Kidney Disorder	0.00	0.02**	0.01	0.00	0.00	0.02
Sample Sizes	1,288	825	538	1,851	1,598	403

Data Source: 2017 Wave, PSID. Sample: PSID heads and spouses, 25 years and older. Weights: PSID individual cross-section weights. Income quartiles are determined across households using family weights.

Stars (*) are for tests of risk factors being of a subgroup being significantly different from base group, where: * P-value ≤ 0.10; ** P-value ≤ 0.05; *** P-value ≤ 0.01.

Figure 1. Marginal Contribution of Each Risk Factor Available in PSID to Relative Vulnerablity Index (*VI*) by Age

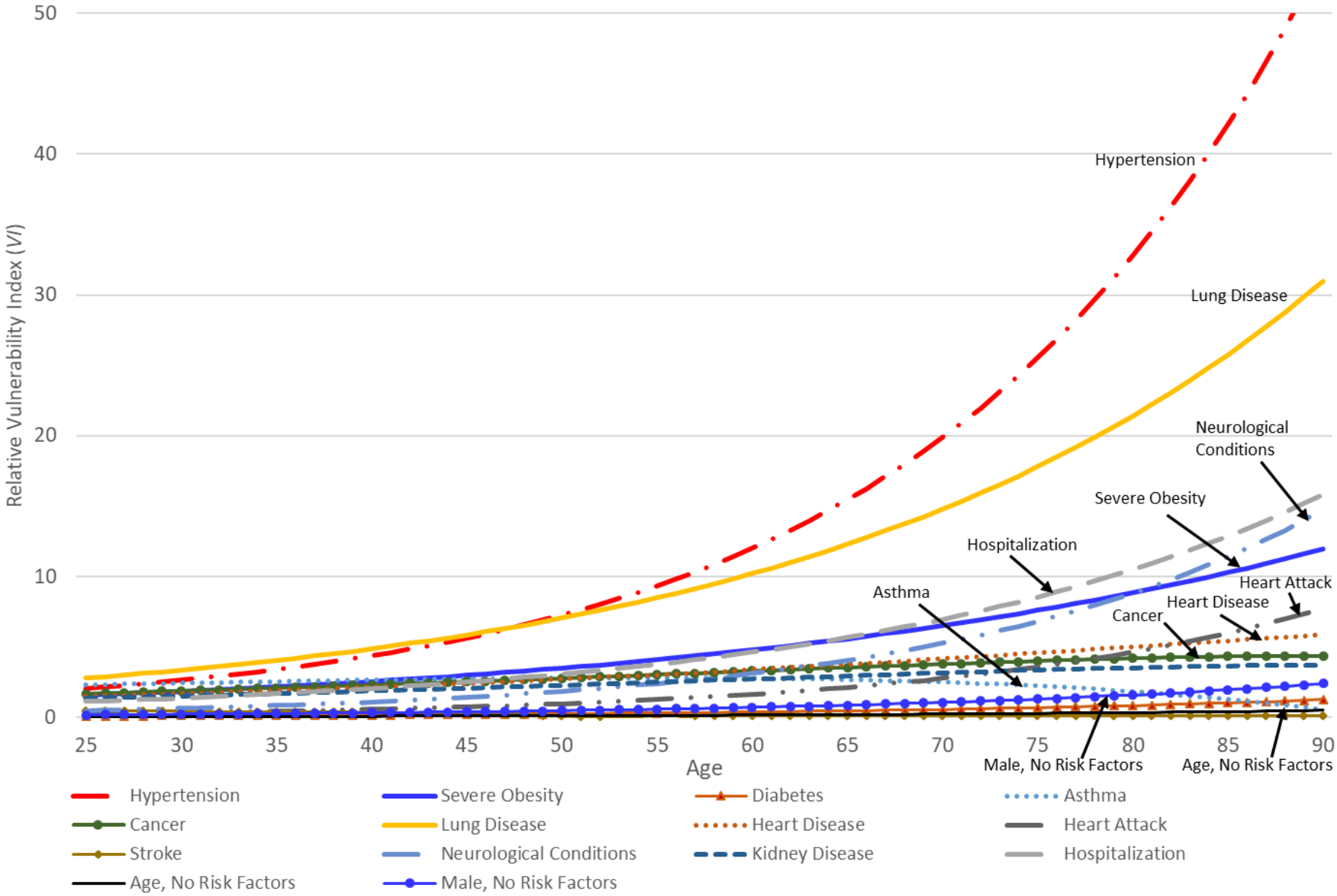
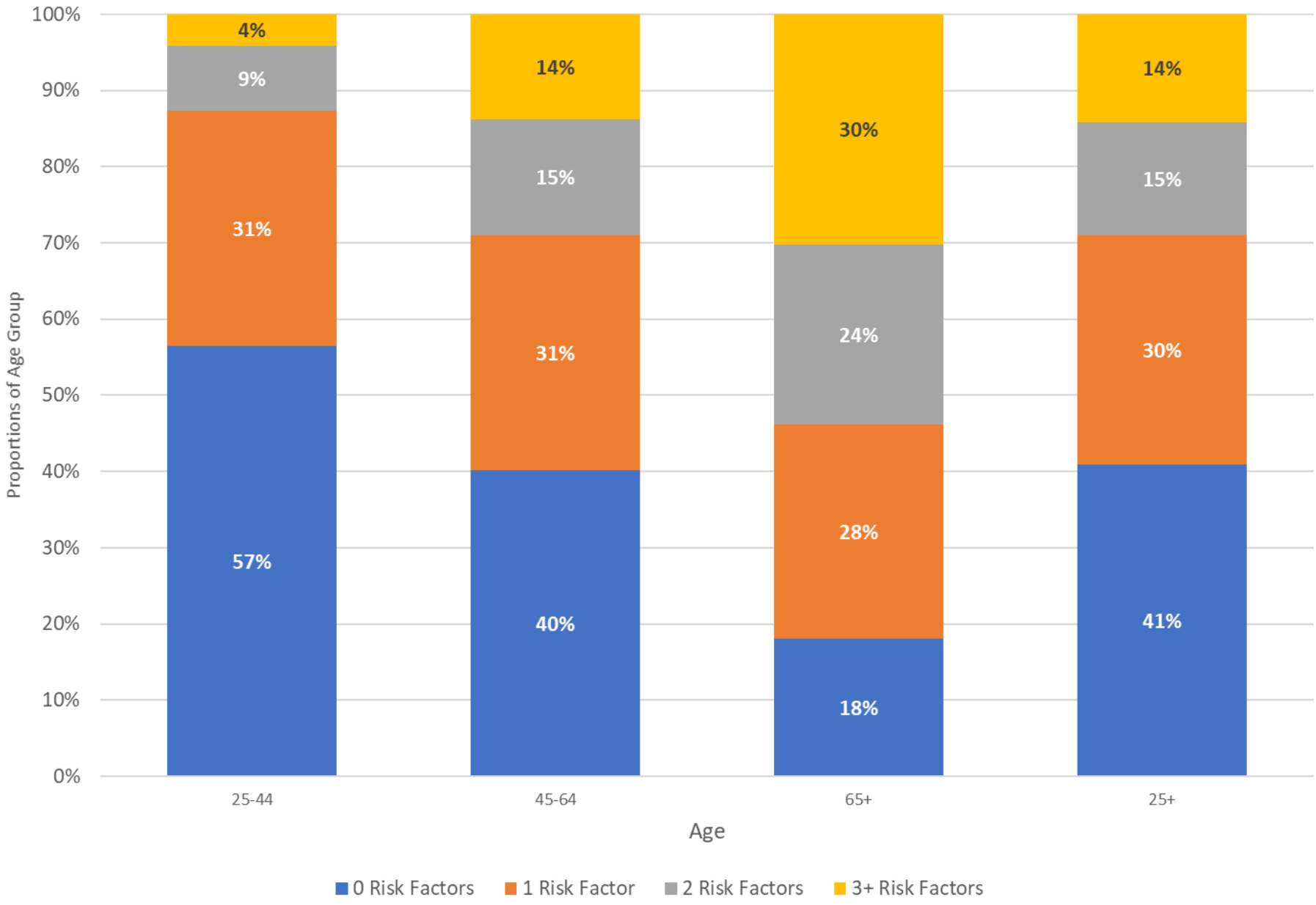


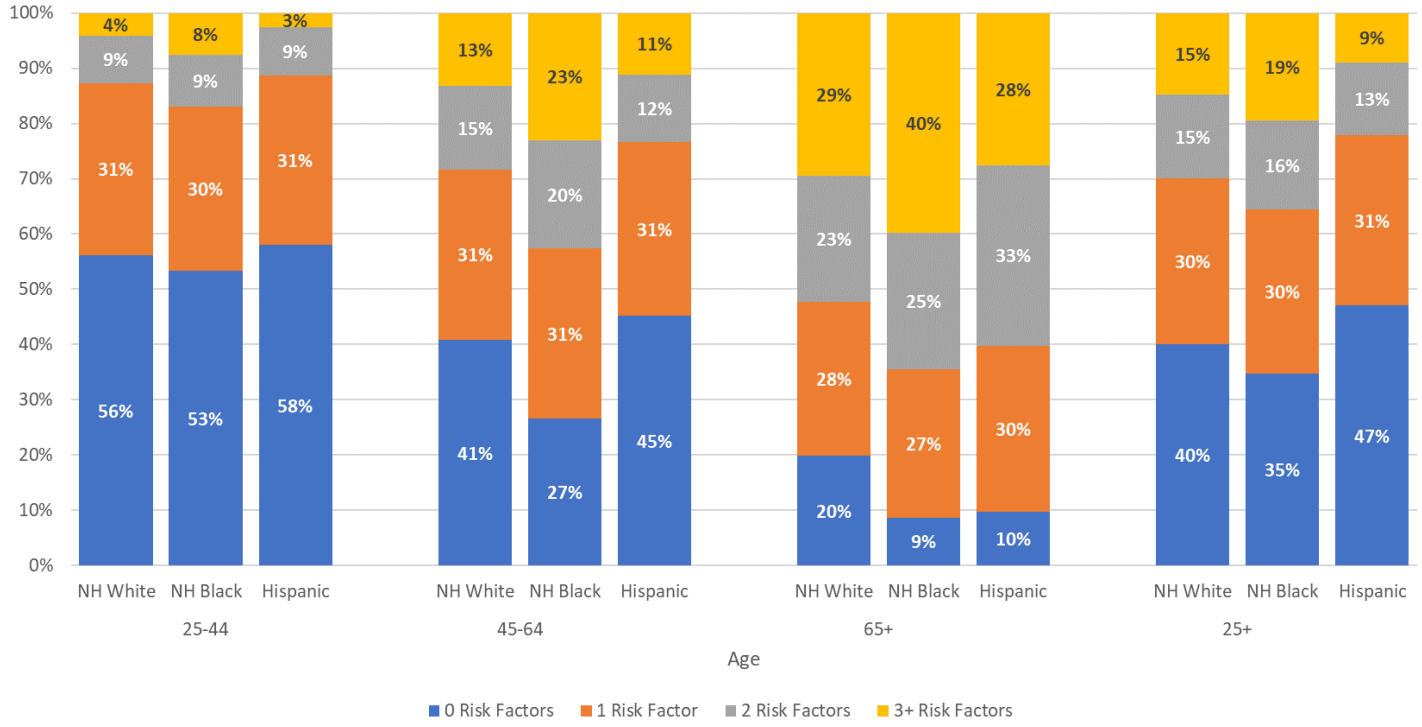
Figure 2. Proportions of Population with Various Numbers of Risk Factors by Age



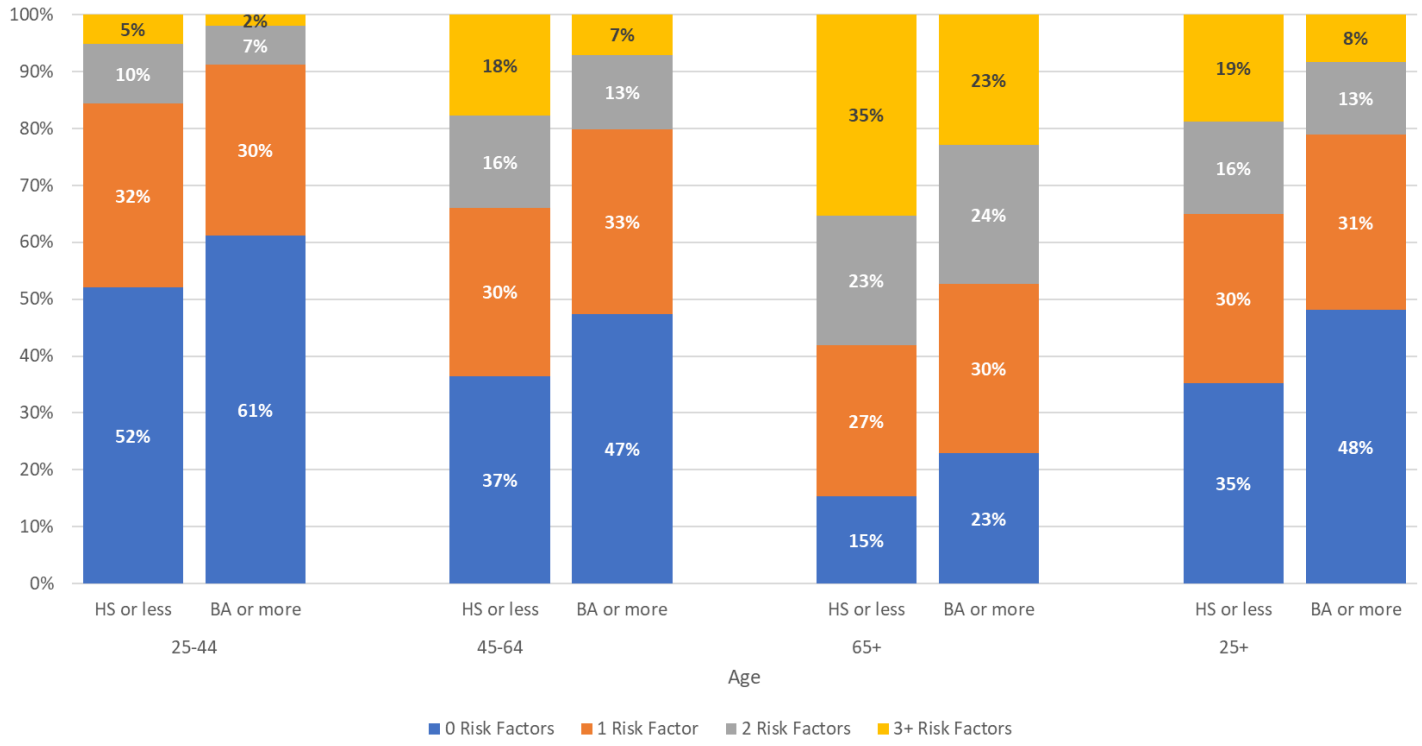
medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 3. Distribution of Number of Risk Factors

(a) by Race-Ethnicity & Age



(b) by Education & Age



medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

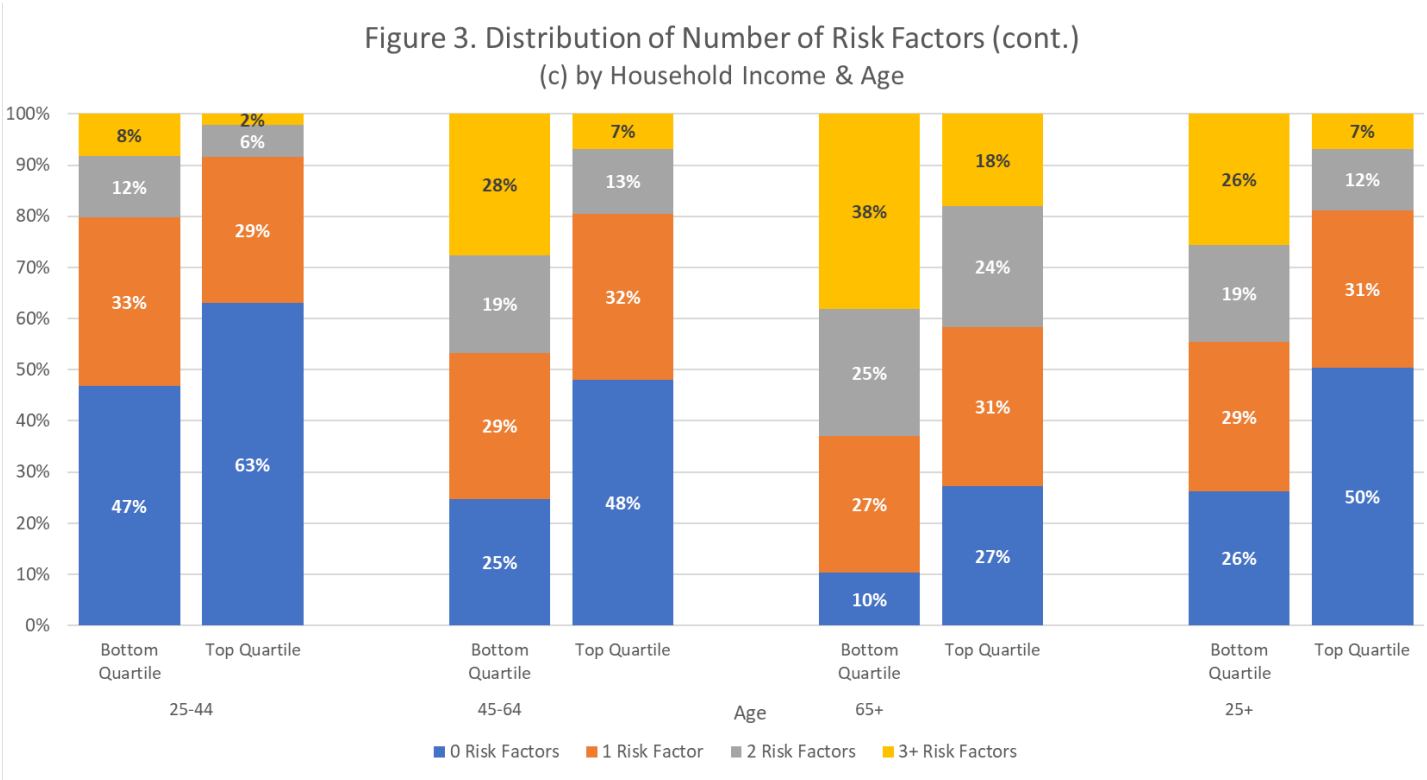
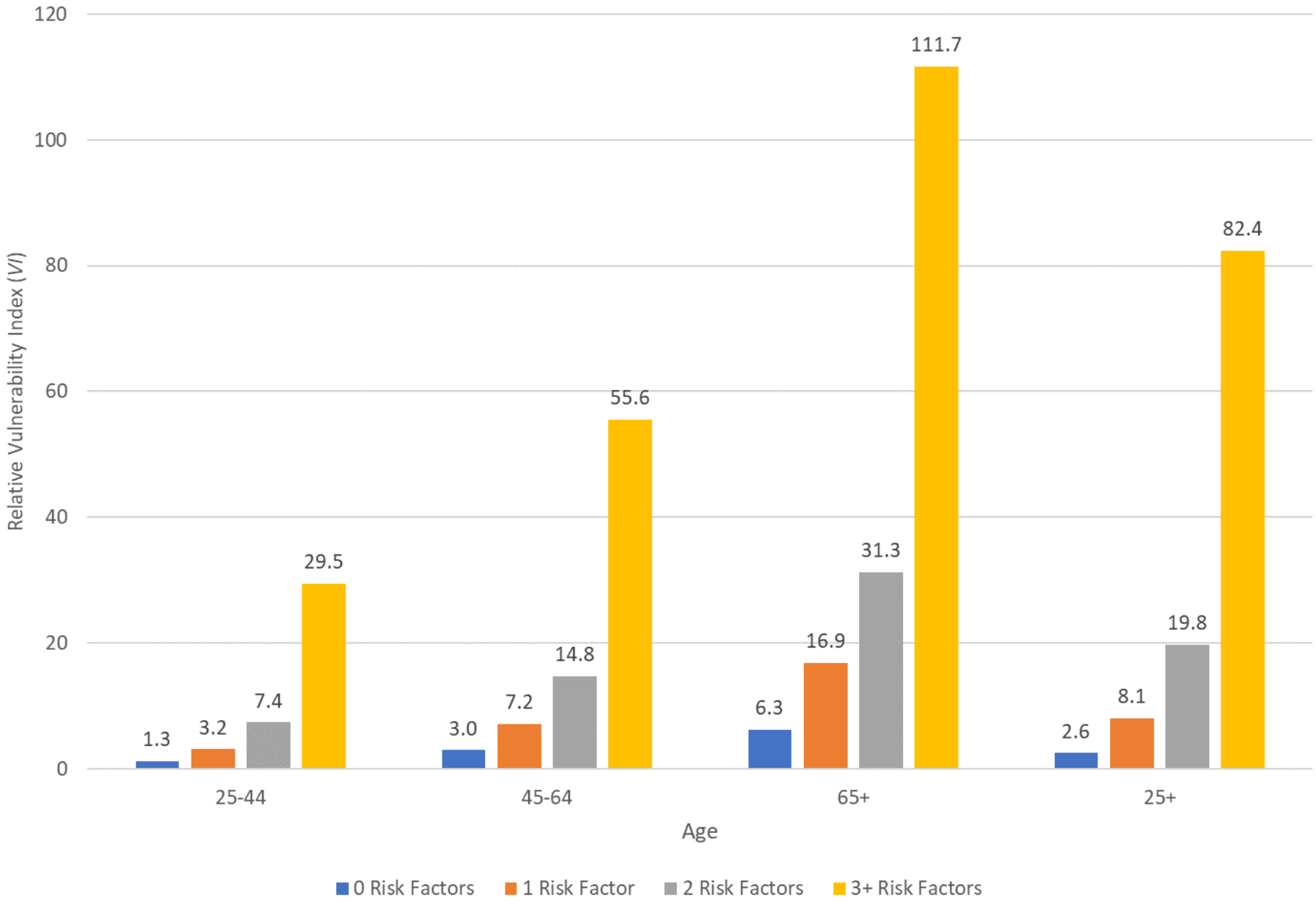


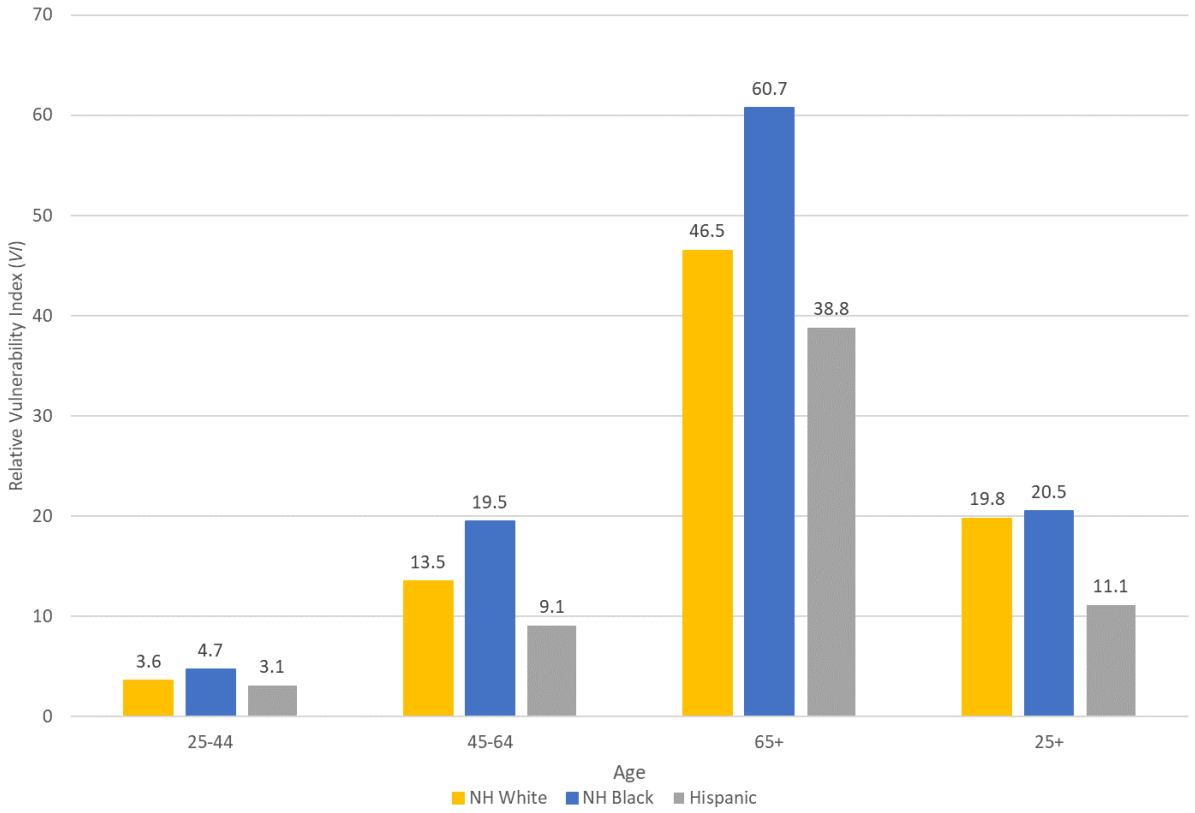
Figure 4. Average Relative Vulnerability Index (*VI*) by Number of Risk Factors & Age



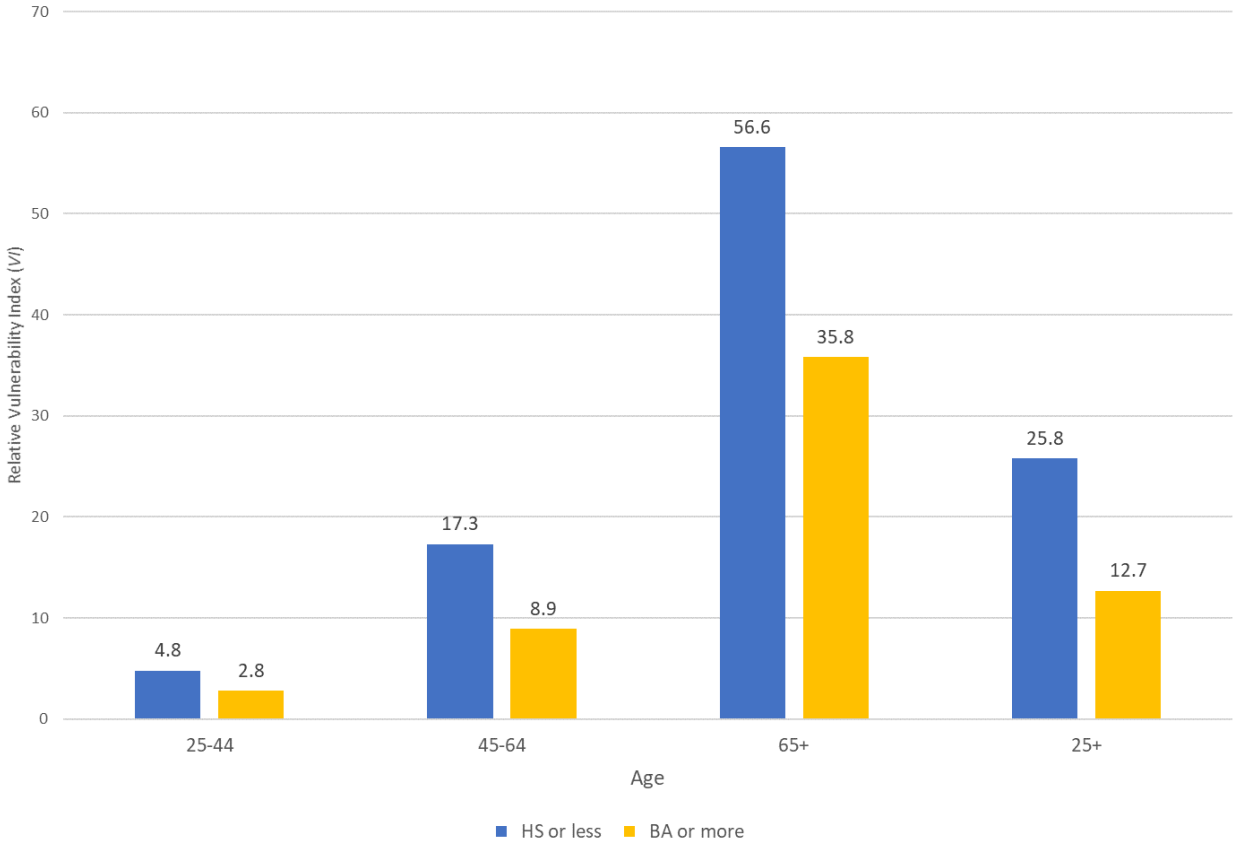
medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 5. Average Relative Vulnerability Index (*VI*)

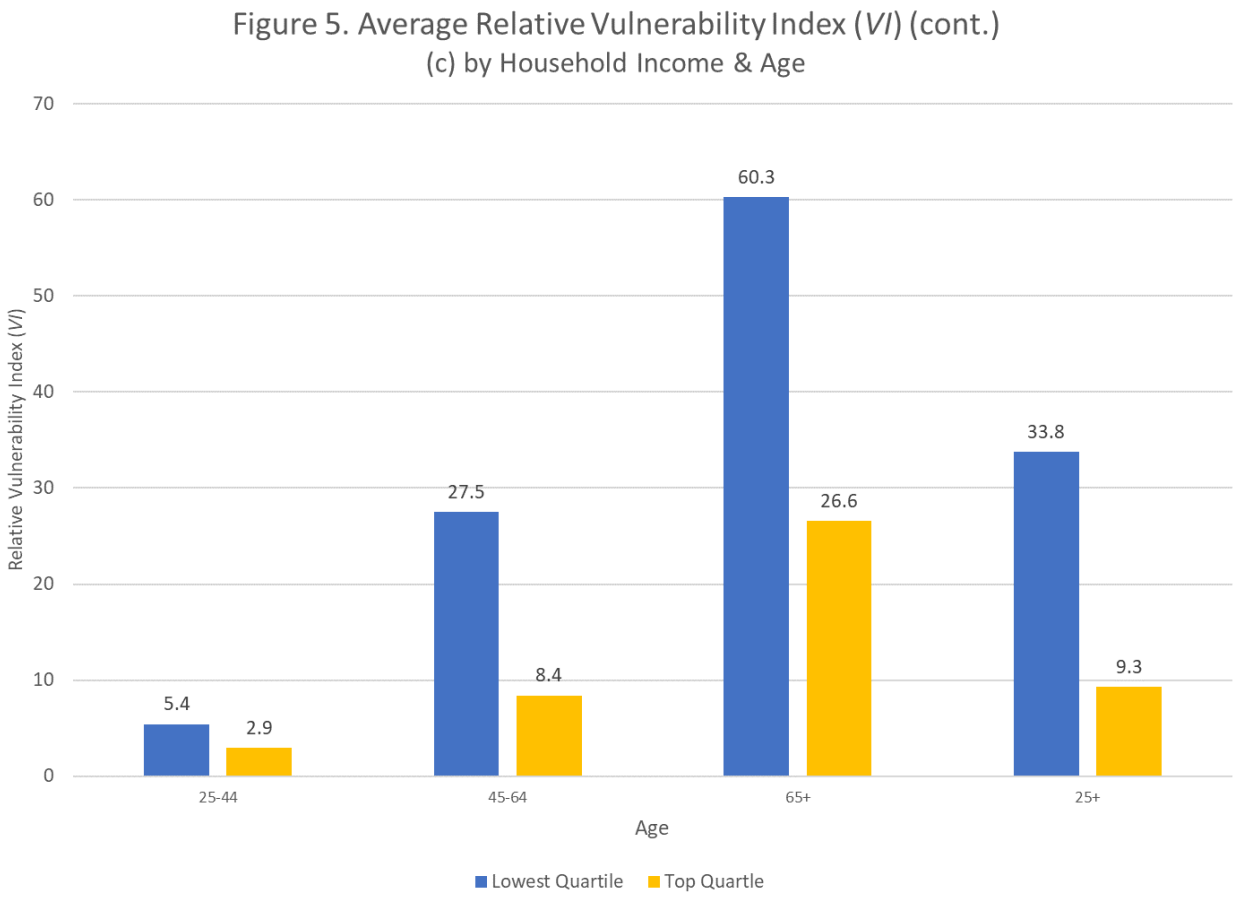
(a) by Race-Ethnicity & Age



(b) by Education & Age



medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Appendix A: Comparison of PSID Sample with Current Population Survey (CPS)

Table A1: Proportions of Demographic and SES Groups by Age across Age Categories & Mean Age for PSID Sample

	Prop. across Age Categories			Prop. within Age Categories					
	25-44	45-64	65+	25-44	45-64	65+	Prop., All Ages (25+)	Mean Age	Sample sizes
Overall				0.373	0.383	0.244		52.0	13,529
Gender									
Female	0.362	0.381	0.256	0.516	0.528	0.557	0.531	52.5	7,420
Male	0.385	0.385	0.230	0.484	0.472	0.443	0.469	51.5	6,109
Race and Ethncity									
NH White	0.331	0.387	0.282	0.632	0.719	0.825	0.712	53.7	7,252
NH Black	0.416	0.416	0.168	0.120	0.117	0.074	0.107	49.4	4,274
Hispanic	0.509	0.360	0.131	0.180	0.124	0.071	0.132	46.7	1,550
NH Other	0.523	0.322	0.154	0.068	0.041	0.031	0.049	47.4	453
Education									
HS or less	0.312	0.405	0.283	0.324	0.408	0.448	0.387	54.4	5,450
Some college	0.388	0.385	0.227	0.245	0.237	0.219	0.235	51.2	3,577
BA or more	0.425	0.360	0.215	0.431	0.355	0.333	0.378	50.1	4,502
Sample sizes	6,962	4,552	2,015						

Data Source: PSID, 2017 Wave. **Sample:** Heads and spouses of households, ages 25 and older. **Weighted.**

Table A2: Proportions of Demographic and SES Groups by Age across Age Categories & Mean Age for CPS Sample

	Prop. across Age Categories			Prop. within Age Categories					
	25-44	45-64	65+	25-44	45-64	65+	Prop., All Ages (25+)	Mean Age	Sample sizes
Overall				0.348	0.403	0.249		52.4	101,462
Gender									
Female	0.349	0.397	0.254	0.532	0.523	0.539	0.530	52.5	53,840
Male	0.346	0.409	0.245	0.468	0.477	0.461	0.470	52.4	47,622
Race and Ethncity									
NH White	0.303	0.408	0.289	0.597	0.695	0.796	0.686	54.3	65,151
NH Black	0.378	0.425	0.198	0.116	0.113	0.085	0.107	50.6	11,378
Hispanic	0.491	0.376	0.132	0.186	0.123	0.070	0.132	46.8	15,976
NH Other	0.466	0.371	0.163	0.101	0.069	0.049	0.075	48.1	8,957
Education									
HS or less	0.288	0.411	0.300	0.304	0.375	0.442	0.367	55.0	38,019
Some college	0.354	0.414	0.232	0.275	0.278	0.251	0.270	51.8	27,560
BA or more	0.403	0.386	0.211	0.421	0.348	0.307	0.363	50.4	35,883
Sample sizes	38,732	40,401	22,329						

Data Source: CPS, 2017 Annual Social & Economic Supplement. **Sample:** Heads and spouses of households, ages 25 and older. **Weighted.**

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Appendix B: Formula for the DeCaprio et al. Model and the Corresponding Risk Factors & Other Variables Available in the PSID

Table B-1 displays the variables used and coefficient estimates for the DeCaprio et al. Model[11] along with the corresponding health-related risk factors that are available in the PSID. In Table B-2, we provide the variable names and question wording from the 2017 Wave of the PSID for these risk factors.

As noted in the text, the PSID does not have all of the variables that DeCaprio et al. (2020) used in the estimation of the logistic regression for this Survey Model. In our implementation of the model, we acted as if the indicators for the risk factors not collected in the PSID to a value of 0. As we note in the text, this means that our estimates of the vulnerability index scores of individuals in the PSID we understate their risk of serious disease. The extent of this understatement and how the extent of understatement will vary in the population depends on both the prevalence of the particular risk factors in the population and their marginal impacts on the vulnerability index score, which is determined by the coefficients on these variables.

We examined what is known about the prevalence of these conditions in the U.S. Here is a summary:

Sickle Cell Disease: (from CDC https://www.cdc.gov/ncbddd/sicklecell/data.html)

- Sickle Cell Disease affects approximately 100,000 Americans.
- SCD occurs among about 1 out of every 365 Black or African-American births.
- SCD occurs among about 1 out of every 16,300 Hispanic-American births.
- About 1 in 13 Black or African-American babies is born with sickle cell trait (SCT).

Hemodialysis (a form of **Dialysis**) (from the National Kidney Foundation https://www.kidney.org/news/newsroom/factsheets/KidneyDiseaseBasics)

- In 2017, 746,557 Americans had kidney failure, and needed dialysis or a kidney transplant to survive.
- Nearly 500,000 of these patients received dialysis at least 3 times per week to replace kidney function.

Pneumonia (from CDC https://www.cdc.gov/dotw/pneumonia/index.html)

- In the United States, more than 250,000 people have to seek care in a hospital due to pneumonia each year.
- Most of the people affected by pneumonia in the United States are adults.

Liver Disease (**Cirrhosis** of the liver) (from CDC https://www.cdc.gov/nchs/fastats/liver-disease.htm)

[11] The first two columns of Table B-1 are taken from Table 4 of DeCaprio et al. (2020) for their Survey Model.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

- Number of adults with diagnosed liver disease: 4.5 million
- Percent of adults with diagnosed liver disease: 1.8%

Acute Rheumatic Fever and Heart Disease (from UpToDate.com https://www.up-todate.com/contents/acute-rheumatic-fever-epidemiology-and-pathogenesis)

- While Rheumatic fever and heart disease remains a major cause of cardiovascular disease in developing nations, it has declined dramatically in industrialized nations like the U.S. because of access to vaccines and antibiotics to treat Group A streptococcus.
- While it can occur at any age, although most cases occur in children 5 to 15 years of age.
- In the U.S., the incidence of Rheumatic fever is less than 2 cases per 100,000 of school-age children.

Thus, with the exception liver disease (Cirrhosis of the liver), the risk factors in the DeCaprio et al. Model, but not in the PSID, are of very low prevalence and/or incidence. Thus, excluding them is not likely to have a significant impact on the relative vulnerability scores, VI_i. The possible exception is Sickle cell anemia, which has a fairly sizeable impact in the calculation of the index for NH Blacks at ages up to around 50. However, as noted above, the prevalence of Sickle cell anemia for NH Blacks and for some of the Hispanics in our sample who reported that their "race" was Black cannot be determined for our PSID sample. Thus, the values for the vulnerability index in the PSID data will be (slightly) understated, on average, for NH blacks and some Hispanics.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.28.20115899; this version posted May 30, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table B-1: Risk Factors & Other Conditions from the DeCaprio et al. Model

Variables in DeCaprio et al. Survey Model	Coefficient Estimate	Variables available in the PSID
intercept	-6.740	
Age	0.041	Age
Male	0.171	Male
Prior Admissions	0.682	Hospitalized
Prior ER Visits	0.413	
Chronic obstructive pulmonary disease (COPD) or emphysema, cystic fibrosis, or chronic bronchitis	1.167	Lung Disease
Asthma	1.393	Asthma
Obesity (BMI ≥ 40)	0.935	Severe Obesity (BMI ≥ 40)
Diabetes (other than when you were pregnant)	0.096	Diabetes
Hypertension (or high blood pressure)	0.832	Hypertension
Congestive Heart Failure	0.982	Heart Disease
Heart attack	0.159	Heart Attack
Rheumatic heart disease	0.788	
Stroke	0.285	Stroke
Sickle cell anemia/HIV infection/Transplant	2.582	
Chronic kidney disease	0.966	Kidney Disease
Hemodialysis	1.369	
Liver disease	0.055	
Pneumonia, acute bronchitis, influenza or other acute respiratory infection	0.696	
Cancer	1.091	Cancer
Neurocognitive conditions	0.294	Neurological Condition
Pregnancy, childbirth & puerperium	0.789	
COPD x Age	-0.002	Lung Disease x Age
Asthma x Age	-0.015	Asthma x Age
Obesity x Age	-0.004	Severe Obesity x Age
Diabetes x Age	0.000	Diabetes x Age
Hypertension x Age	0.005	Hypertension x Age
Congestive heart failure x Age	-0.007	Heart Disease x Age
Myocardial infarction x Age	0.003	Heart Attack x Age
Rheumatic heart disease x Age	-0.008	
Stroke x Age	-0.003	Stroke x Age
Sickle cell/HIV/Transplate x Age	-0.028	
Chronic kidney disease x Age	-0.008	Kidney Disease x Age
Hemodialysis x Age	-0.018	
Liver disease x Age	0.001	
Pneumonia, acute bronchitis, influenza or other acute respiratory infection x Age	-0.005	
Cancer x Age	-0.009	Cancer x Age
Neurocognitive conditions x Age	0.004	Neurological Condition x Age
Pregnancy, childbirth & puerperium x Age	-0.003	

Table B-2: Health-Related Risk Factors and Other Variables, Variable Names, and Question Wording from the 2017 Wave of the PSID

Risk Factor/Variable	PSID Variable	Question Wording
Age	ER34504	What is your current age?
Gender	ER32000	What is your gender?
Past Hospitalization (in 2016)	ER68511, ER69638	Were you a patient in a hospital overnight or longer at any time during 2016?
Lung Disease	ER68454, ER69581	Has a doctor or other health professional ever told you that you had chronic lung disease such as bronchitis or emphysema?
Asthma	ER68449, ER69576	Has a doctor or other health professional ever told you that you had asthma?
Severe Obesity	ER68568, ER68569, ER69695, ER69696, ER68566, ER68567	Constructed BMI using height and weight variables and designating obese if BMI > 40
Diabetes	ER68459, ER69586	Has a doctor or other health professional ever told you that you had a heart attack?
Hypertension	ER68444, ER69571	Has a doctor or other health professional ever told you that you had diabetes or high blood sugar?
Heart Disease	ER68439, ER69566	Has a doctor or other health professional ever told you that you had coronary heart disease, angina, congestive heart failure?
Heart Attack	ER68433, ER69560	Has a doctor or other health professional ever told you that you had a heart attack?
Stroke	ER68427, ER69554	Has a doctor or other health professional ever told you that you had a stroke?
Kidney Disease	ER68498, ER69625	Are there other conditions that you have? [3: Kidney disease]
Cancer	ER68479, ER69606	Has a doctor or other health professional ever told you that you had cancer or a malignant tumor?
Neurological condition	ER68469, ER69596	Has a doctor or other health professional ever told you that you had permanent loss of memory or mental ability?