The Lifecycle Formation of Health Inequality*

Kaveh Danesh[†] **Jonathan Kolstad**[‡] **William Parker**[§]
Johannes Spinnewijn[¶]

September 22, 2026

Abstract

Life expectancy differs substantially across income groups, but mortality reveals little about when and how underlying health inequalities emerge. Using population-wide Dutch panel data linking income to detailed histories of diagnosed chronic conditions, we construct a mortality-weighted index that summarizes individual-level chronic-condition profiles on a common scale. We then use this index to study when and how health inequality forms over the lifecycle. About half of the old-age difference in chronic disease burden across the income distribution is already present by age 40, while the mortality differences that drive differences in life expectancy emerge decades later. Differential accumulation of chronic conditions accounts for 58 percent of the gap, with diabetes and cardiovascular disease being the main sources of divergence through midlife. The remaining 42 percent reflects income sorting by health. Mental health conditions play the largest role in sorting early in the lifecycle, and poor health is disproportionately associated with movements into the bottom of the income distribution.

*We thank Rob Alessie, Sandy Black, Zach Bleemer, Raj Chetty, Janet Currie, Liran Einav, Amy Finkelstein, John Friedman, Ingrid Huitfeldt, Henrik Kleven, Camille Landais, Adriana Lleras-Muney, Ellen Meara, Canishk Naik, Carol Propper, Hannes Schwandt, Bram Spinnewijn, Michael Stepner, Frank van Dijk, Winnie Van Dijk and seminar participants at the LSE, IFS, University of Oslo, Bank of Spain, BI Norwegian Business School, CPB Netherlands, Bocconi, Bologna, EIEF, Zurich, Oxford, Royal Holloway, London Business School, NHH Bergen, LSE Health, KULeuven, Tinbergen Amsterdam, Opportunity Insights, Columbia, Princeton, Yale, NYU, and at the CEPR Public Economics symposium, the NBER Aging meeting, the Dutch health econometrics workshop, the lolaHESG conference, and the RFBerlin Economics of Aging conference for valuable comments and suggestions. We also thank the Dutch Central Bureau of Statistics and Thomas Minten for help with the data and Boris Beyen, Ben Dahmen, Rasmus Duret, Diego Ferreras-Garrucho, and Marco Visentin for excellent research assistance. We thankfully acknowledge financial support from the International Inequalities Institute. An earlier version circulated as a working paper under the title 'The Chronic Disease Index: Analyzing Health Inequalities Over the Lifecycle'.

[†]UC Berkeley and UCSF
[‡]UC Berkeley and NBER
[§]London School of Economics
[¶]London School of Economics and CEPR

1 Introduction

Inequality in health is a major source of socioeconomic disparities and a central policy challenge facing many countries. Efforts to 'close the health gap' are a key element of numerous policy agendas (e.g., World Health Organization (1985, 2008, 2017), as well as the EU's recent "Joint Action on Health Inequalities"). Alongside these health-inequality agendas, chronic disease has become a policy priority in its own right. This focus reflects the scale and persistence of the problem: chronic conditions account for a large share of morbidity, mortality, and healthcare spending, and their burden is strongly correlated with socioeconomic status. Recent evidence and policy initiatives like the Make America Healthy Again Commission in the U.S. illustrate this broader concern.[1]

A large body of work in public health, epidemiology, sociology and economics has studied health inequalities and delivered detailed insights into how socioeconomic factors associate with self-reported health, diseases and mortality (see for example Cutler, Lleras-Muney, and Vogl 2011; Lleras-Muney, Schwandt, and Wherry 2024). A more recent and rapidly growing literature uses linkages between mortality, tax and other administrative registers to provide an increasingly granular perspective on inequalities in life expectancy (e.g., Chetty et al. 2016; Godøy and Huitfeldt 2020; Finkelstein, Gentzkow, and Williams 2021; Chen, Persson, and Polyakova 2022; Black et al. 2024; Chetty et al. 2025). Yet these mortality gradients represent the endpoint of a long health process. As a result, they provide limited guidance about *when* health inequality is formed, *through which channels* it develops, and *how* the forces shaping health inequality vary over the lifecycle and across the income distribution.

In this paper, we study the formation of health inequality over the lifecycle using population-wide administrative data from the Netherlands. We build a panel for the full Dutch population over a 20-year period, combining multiple administrative registries with detailed data on health and socioeconomic outcomes. Using pharmaceutical data, we construct comprehensive profiles of diagnosed and treated chronic conditions for approximately 17 million people. We aggregate these profiles into a Chronic Disease Index (CDI), a mortality-weighted measure of chronic disease burden that can be compared across individuals and followed across ages.

1. In 2023, 76% of U.S. adults reported at least one chronic condition and 51% reported multiple chronic conditions (Watson et al. 2025). Most U.S. healthcare spending is incurred by people with chronic and mental health conditions (Centers for Disease Control and Prevention 2026a). The recently launched Make America Healthy Again Commission now directs federal health agencies to focus on reversing chronic disease, including mental health disorders, obesity, and diabetes (Executive Office of the President 2025).

While the focus on chronic disease can be motivated by its scale and policy relevance, the empirical value of the CDI comes from combining three features that are rarely available together: a comprehensive health measure, repeated observation of the same individuals, and population-wide scale. First, the index allows us to study chronic illness as a dynamic marker of health that appears well before mortality is realized. Second, by linking the index to longitudinal income and socioeconomic records, we can not only document health inequality at different ages, but also separate distinct dynamic pathways between health and income. Finally, the population-wide scale of the data allows us to examine precisely how these pathways vary across ages, income, and chronic conditions.

The lifecycle perspective is natural at the conceptual level, but challenging to establish empirically. Indeed, the idea that health status in later life reflects long-run life experience and investments in health has been well-developed since Grossman (1972). Similarly, the observations that chronic conditions can have important health implications and differ in their prevalence across socioeconomic groups throughout life are not new (see, e.g., Loucks et al. 2007; Martinson 2012; Bauer et al. 2014). What is less studied is how these chronic conditions develop dynamically and interact with income over the lifecycle, before resulting in death.

To study lifecycle health dynamics in our setting, we first address two measurement challenges. The first is whether diagnosed and treated chronic conditions provide a reliable and comparable measure of health across income groups. The second is how to aggregate high-dimensional chronic disease profiles into a common health scale that captures the component of health relevant for the mortality gradients we seek to understand.

For the first challenge, measuring chronic disease itself, we build on mappings from dispensed medications to chronic diseases in the medical literature (Huber et al. 2013). This allows us to observe treated chronic illness repeatedly and for the full population. Because this is a healthcare-based measure, it captures diagnosed and treated disease rather than all latent disease. This distinction is not merely conceptual: standard chronic condition measures can understate true prevalence when individuals are unaware of their condition or when it is not properly diagnosed or treated (Buttorff, Ruder, and Bauman 2017; Centers for Disease Control and Prevention 2026b). We therefore treat comparability of diagnosis and treatment across income groups as an empirical question and assess it directly. In particular, we study healthcare expenditures and mortality associated with chronic conditions, as well as concordance with self-reported conditions and testing rates in survey data.

Conditional on observed chronic conditions, treatment intensity and mortality differ little across income groups in the Netherlands, with suggestive evidence of under-diagnosis appearing only at the very bottom of the income distribution. The Dutch healthcare system has thus broadly equalized access to *healthcare*—an important finding in its own right, and one that is valuable for our analysis. It allows us to focus primarily on differences in the *prevalence* of chronic conditions rather than differences in their *treatment*. Yet equalized access does not imply equal health: substantial income differences in chronic-condition prevalence remain, and these differences account for a sizeable share of mortality gaps at different ages.

For the second challenge, aggregating multidimensional chronic-condition profiles, we use predicted mortality risk as the common scale. Since life expectancy gradients are the endpoint we seek to understand, mortality risk provides a natural scale for this aggregation. We anchor this scale at age 70, when mortality rates are high and chronic conditions are strongly predictive of subsequent survival. We adapt a double-Lasso procedure, flexibly accounting for rich comorbidity patterns, including multiple lags and interactions, and controlling for a rich set of socioeconomic differences, which may correlate with chronic conditions, but affect mortality through other channels (Belloni, Chernozhukov, and Hansen 2014). Both adjustments matter for the estimated mortality effects, further underlining the value of rich, population-wide data for this analysis.

The estimated weights are then used to construct the Chronic Disease Index for any individual at any age, expressing their observed chronic-disease profile in mortality-equivalent units. The CDI thus provides a transparent way to compare health across ages and income groups. Although the measure has important caveats as we discuss, we show that it captures meaningful variation in health well before death, allowing us to trace when inequality emerges and through which channels it evolves.

Our lifecycle analysis delivers three main sets of results:

First, we show that the health gap between income groups emerges much earlier than mortality gradients alone would suggest. Mortality *after age 70* accounts for about 80% of life expectancy gaps across income groups. In contrast, nearly half of the old-age gap in chronic disease burden is formed *before age 40*. Put differently, the gap is equivalent to a substantial age differential: by age 35, low-income individuals have about the average CDI of high-income individuals aged 50. The drivers of mortality inequality thus operate long before mortality itself becomes

measurable. As a result, policies focused only on late-life mortality risk may arrive after much of the underlying health inequality has already formed.

Second, exploiting the panel structure of the data, we decompose the observed health gap into two channels: *differential aging*—the faster accumulation of chronic conditions among lower-income individuals—and *health-based income sorting*—the tendency for individuals who develop chronic illness to move into lower income groups. We find that both channels contribute substantially to the health gap at old age, but that differential aging is the larger component: it accounts for 58 percent of their combined contribution, compared with 42 percent for health-based sorting. The aging mechanism evolves gradually over the lifecycle; chronic disease develops at a faster rate for lower-income individuals, and this difference increases with age. The sorting mechanism is also present throughout the lifecycle, but is especially important at the start of working life. The channels also differ sharply across conditions: mental health conditions operate primarily through sorting, whereas cardiovascular disease and diabetes contribute to the gradient mainly through differential aging.

Third, leveraging the granularity of the data, we show that the pronounced non-linearity of the health gradient—with poor health concentrated at the bottom of the income distribution—is driven by health-based sorting effects. This corresponds to the non-linearity of life expectancy documented in prior work (e.g., Mortensen et al. 2016), but shows that this pattern appears much earlier in the health process, and identifies the channel through which it develops. The aging effects are approximately linear across the income distribution: all income groups experience faster chronic disease accumulation than the group above, with similar increments. However, the sorting effects are highly nonlinear, worsening average health at the bottom of the income distribution while improving the health of all other groups. That is, poor health does not merely shift individuals downward in the income distribution; it disproportionately moves them toward the bottom.

Our results have direct implications for policy design to address health inequality, which we further substantiate through counterfactual analysis. In particular, we quantify the returns to intervening at different points in the lifecycle. In a simulation that equalizes the accumulation of chronic disease from age 20 onward, the life expectancy income gap falls by 31%. Equalizing this process from age 40 onward still closes 27% of the gap. However, waiting until age 60 reduces the gain substantially, closing only 8% of the gap. These results reinforce that equal access to healthcare, while important, is insufficient to close the health gap once large differences in chronic

disease burden have already accumulated. They instead highlight the value of policies that slow the incidence of chronic disease among lower-income individuals before those differences become entrenched.

Related literature This work bridges three strands of literature that have separately studied mortality gradients, the lifecycle dynamics of health, and the forces underlying socioeconomic health inequality.

First, a rapidly growing literature in economics studies social gradients by combining administrative data from mortality registers and tax records.[2] Following the seminal contribution by Chetty et al. (2016) in the US, the income gradients have been extended to other countries (Kinge et al. 2019; Chen, Persson, and Polyakova 2022), for different age groups (Kennedy Moulton et al. 2022), across geography (Chetty et al. 2016; Godøy and Huitfeldt 2020; Finkelstein, Gentzkow, and Williams 2021), etc. Our contribution is to use administrative data to measure health before death and study how the health gap leading to mortality differences at late age evolves over the lifecycle. Most relatedly, Chetty et al. (2025) are currently studying the mechanisms underlying time trends in the US, following a framework closely related to ours. Shui et al. (2025) explicitly build on our work to study health gaps over the lifecycle in the Netherlands based on biomarkers observed in the Lifelines study.

Second, the question of how health itself evolves dynamically over the lifecycle has been central to health economics since Grossman (1972), but the empirical evidence has been arguably lagging behind. Seminal empirical work has documented lifecycle patterns in socioeconomic health gradients (Case, Lubotsky, and Paxson 2002; Case and Deaton 2005), but data limitations have made it difficult to observe detailed within-individual health dynamics by age and socioeconomic status (e.g., Currie and Stabile 2003; Case, Fertig, and Paxson 2005). More recent structural work has used survey data and calibrated models to study the evolution of health and its interaction with economic outcomes, including Galama and van Kippersluis (2019), Hosseini, Kopecky, and

2. Preceding the work using administrative data, the longstanding health inequality literature can be categorized by the type of data used: **a) longitudinal surveys focused on old-age individuals**, including HRS, ELSA, & SHARE (e.g. Smith 2004; Avendano et al. 2009; Banks, Muriel, and Smith 2010; Zaninotto et al. 2020; Russo et al. 2024); **b) cross-sectional population surveys at all ages** such as NHIS, BRFSS, & EHIS (e.g. Bleich et al. 2012; Singh et al. 2017) and NHANES & HSE (e.g. Loucks et al. 2007; Seeman et al. 2008; Martinson 2012; Nesson and Robinson 2017; Campbell et al. 2023; Abdalla et al. 2025), which include biomarkers too; **c) long-term birth cohort studies** such as NSHD & NLSY (e.g. Pais 2014; Khanolkar et al. 2021; Bolt 2022); or **d) studies that sample from an electronic health record**, often specific to a disease or geographically focused, such as the TOGETHER study (London, UK), the Lifelines study (Groningen, NL), as well as data provided by healthcare networks including Geisinger Health System (PA, US) and Sentara Healthcare (VA/NC, US) (e.g. Dharmayat et al. 2020; Zhu, Dekker, and Mierau 2023; Poulsen et al. 2024; Lu et al. 2024).

Zhao (2022), De Nardi, Pashchenko, and Porapakkarm (2023) and Hosseini, Kopecky, and Zhao (2025). In particular, Borella et al. (2024) study health gaps in the US across race and ethnicity from age 55 onwards using the Health and Retirement Study and link these to subsequent economic outcomes. Relative to this literature, we provide a data-driven account based on rich administrative panel data for the full population, allowing us to trace the accumulation of specific chronic conditions over the lifecycle with minimal additional structure.

Finally, our paper speaks to the broader interdisciplinary literature studying the drivers of the health gap (Cutler, Lleras-Muney, and Vogl 2011; Marmot 2015; Mackenbach 2019; Murray et al. 2020). Much of this work in the economics literature studies specific causal pathways.[3] Our analysis does not replace this causal work. Instead, it provides an empirical decomposition of the lifecycle health gap into differential accumulation of chronic conditions and health-based income mobility, clarifying where more targeted causal analysis is most needed.

This paper proceeds as follows. Section 2 discusses the data environment and presents some motivating descriptives. Section 3 presents a conceptual framework to analyse health gaps, both in a static and dynamic sense. Section 4 describes the measurement of chronic disease and the construction of our health measure. Section 5 uses this measure to study how the health gap evolves over the lifecycle, and presents counterfactuals of alternative aging scenarios. Section 6 concludes.

2 Data and Motivating Descriptives

The Netherlands provides a particularly valuable setting for studying lifecycle health inequality: population-wide administrative records allow us to link income, demographic characteristics, mortality, healthcare use, and detailed histories of dispensed medications for nearly the full resident population over time. We first describe the data sources and construction of the core income and health variables, then present two descriptive patterns that motivate the subsequent analysis.

3. This includes the pathway from socioeconomic determinants to health outcomes (e.g., Currie 2009; Adda, Banks, and von Gaudecker 2009; Sullivan and von Wachter 2009; Black, Devereux, and Salvanes 2015) and from health shocks to socioeconomic outcomes (e.g., Currie 2009; Dobkin et al. 2018; Stepner 2019), as well as specific mediators such as access to medical care (e.g., Finkelstein and McKnight 2008; Deaton and Paxson 2004), obesity (e.g., Bolt 2022), health behaviors (e.g., Pampel, Krueger, and Denney 2010; Darden, Gilleskie, and Strumpf 2018), early life factors (e.g., Case, Fertig, and Paxson 2005; van den Berg, Lindeboom, and Portrait 2006), social structures and stress (e.g., Marmot et al. 1991; Sapolsky 2005). See also the review in O'Donnell, Van Doorslaer, and Van Ourti (2015).

2.1 Data Sources

We build a panel for the full Dutch population using microdata from the national statistical agency (Dutch: *Centraal Bureau voor de Statistiek*, CBS). The panel runs from 2003 to 2021 and combines multiple administrative registries and survey sources. For practical purposes, it covers the full resident Dutch population (17.2 million people in 2016). The basis is the Municipal Register (Dutch: *Gemeentelijke Basisregistratie*, GBA), which includes an individual identifier that forms the linkage between datasets.

Selected descriptive statistics on socioeconomic and health outcomes are included in Appendix Table C.1. The table reports moments for the full sample, for selected calendar years (2006 and 2016), and for selected ages (40 and 70). It summarizes demographics, education, income and wealth, chronic condition prevalence, healthcare use, and mortality outcomes.

Mortality and Medicines The two main health outcomes we consider are mortality and chronic illness. Mortality is derived from death certificates in the mortality register, completed by either a physician or pathologist and collated by CBS for statistical purposes.[4] To measure chronic conditions, we use dispensed-medication records, mapping prescription histories to a set of chronic-condition indicators that can be observed repeatedly for the full population. Section 4 describes this measurement approach in detail and assesses the extent to which medication-based prevalence captures comparable chronic disease burden across income groups. The dataset for prescription medication is administered by the National Health Care Institute (Dutch: *Zorginstituut Nederland*). This contains medicines dispensed to an individual, outside the hospital setting. Medicines are classified by the Anatomical Therapeutic Chemical (ATC) code, at ATC3 digit level: for example, the code *N06A* corresponds to antidepressants. This prescription dataset is available from 2006 onwards.

Other health data Annual healthcare costs are collated by the commercial data provider *Vektis* using information from health insurers. The data provided to CBS relates to costs insurable under the Dutch Health Insurance Act (Dutch: *Zorgverzekeringswet*, ZVW). This includes costs insurable under compulsory standard insurance, which represented 90% of all hospital, medical

4. The mortality register also includes cause of death, defined as "The disease or injury that initiated the train of morbid events leading directly to death." We are instead most interested in the underlying conditions; for instance, hyperlipidemia rather than an acute myocardial infarction. Further, cause-of-death coding is often less reliable for chronic rather than acute diseases (Harteloh, de Bruin, and Kardaun 2010).

practice, and pharmacy costs in the Netherlands from 2009-2021 (Statistics Netherlands 2024). Annual totals are itemised by category, including general-practitioner, medication, mental-health, and hospital costs. These data cover the resident population from 2009 onwards. We also use the frequency and duration of hospitalisations and the primary-diagnosis ICD code from the hospital discharge register (Dutch: *Landelijke Medische Registratie*, LMR) for 2011–2017. These hospitalization variables do not enter the core CDI measure; we use them in robustness analyses testing whether acute-care information adds explanatory power for income differences in mortality.

In addition to the administrative data, we merge representative surveys on self-reported health and health behaviors at the individual level. These include the large-scale 'Health Monitor' surveys (Dutch: *Gezondheidsmonitor*, GEMON), fielded in 2012 and 2016 as repeated cross-sections covering around 400,000 individuals. We also link an annual cross-sectional 'Health Inquiry' survey of approximately 9,000 individuals per year (Dutch: *Gezondheidsenquete*, GECON). These surveys report self-assessed physical and mental health and participation in screening activities such as blood-pressure tests.

Household Income Income data are collated by CBS from the tax authorities, and is available at the population level from 2003 onwards. In this work, we focus on *Standardized disposable household income*.[5] This measure is constructed by CBS. In line with prior literature, negative or zero disposable income households are omitted: these represent less than 1% of observations. For comparability, we also follow prior work in ranking individuals based on lagged income (Chetty et al. 2016; Kinge et al. 2019): we take the mean of $(Y_{t-4}, Y_{t-3}, Y_{t-2})$ and use pre-retirement incomes (Y_{60}, Y_{61}, Y_{62}) for those aged 65 and above. Lagging income reduces the influence of contemporaneous health shocks on the income ranking, but our analysis shows that health-based income mobility remains important even with this lagged measure. We therefore test the sensitivity of the lifecycle decomposition to alternative income definitions, including shorter lag lengths restrictions to individuals who remain in the same income group for longer durations.

To form measures of relative income, we rank lagged income within gender, birth cohort, and calendar year. Much of the analysis uses a binary classification: we define income percentiles

5. Disposable income is defined by CBS as all gross income and government insurance/benefit transfers, less insurance premia, and taxes on income and wealth. This measure is standardized by CBS at the household level by dividing the sum of members' disposable income by a 'household equivalence factor', which adjusts for differences in the size and composition of households.

1–50 as **Low income** and percentiles 51–100 as **High income**.[6] We also consider alternative markers of individuals' socioeconomic status, including parents' income, education, and wealth.

Other Administrative Data The CBS microdata environment is comprehensive in its administrative data collection. Beyond birth year, gender at birth, and birth country, it includes linkages to biological parents, household membership and composition, and residential postcode. Highest education attained is taken from a combination of administrative records for younger cohorts and labour-force survey data for older cohorts. Education coverage is approximately 60% for those aged 40 and falls to 20% for 70-year-olds.

2.2 Motivating Features of the Data

This section presents two facts that motivate the paper's focus on lifecycle health dynamics. First, income gradients in life expectancy are large in the Netherlands, but most of the resulting life expectancy shortfall appears late in life, after the ages at which lifetime income differences are largely established. Second, chronic disease is already strongly correlated with income in midlife and explains a meaningful share of mortality differences, making it a natural measure for studying how health inequality forms before mortality itself is realized.

Dynamics of Inequality in Life Expectancy

The principal measure of health inequality is the gap in life expectancy across income (e.g., Chetty et al. 2016; Kinge et al. 2019). Figure 1 presents life expectancy at age 40 by income percentile separately for women and men. Life expectancy rises strongly with income for both genders: from 77.5 years at p4 to 89.0 years at p100 for women, and from 71.9 to 86.1 years for men. The corresponding gradients are 11.5 and 14.2 years, respectively, with the sharpest gains concentrated at the bottom of the distribution in both series.[7] Women have higher life expectancy throughout the income distribution, while the income gradient is steeper for men.

6. Notably, the bottom income decile includes some households with high net assets: this suggests some have targeted low incomes perhaps for tax reasons, rather than reflecting overall financial resources. However, it also predominantly includes individuals with very little financial resources, whom we wish to include. Nevertheless, in several aspects of the analysis, we treat the bottom decile separately to account for its distinct composition.

7. The bottom three percentiles of income include those with very low incomes who do not claim any sort of social insurance, but tend to have other financial resources. These are included for transparency; similar patterns were found in related work, e.g. Kinge et al. (2019) omit p1-p3 from their analysis.

Despite universal coverage and broad healthcare access in the Netherlands, these gender-specific gradients are comparable in magnitude to those documented in the US (Chetty et al. 2016).

Given this inequality in life expectancy, a natural follow-up question is when the underlying health gap emerges. Panel A of Figure 2 decomposes the life expectancy shortfall relative to the top income percentile into cumulative age windows.[8] At the peak of the gradient—the income bin averaging percentiles 3 and 4—only 2% of this overall shortfall is generated before age 50 and 20% before age 70; the remaining 80% is realized after age 70. This suggests that mortality differences are most relevant later in life, while the underlying health inequalities may emerge much earlier. The earnings dynamics literature also suggests that most of the lifetime earnings variation is already established by middle age (e.g., Karahan, Ozkan, and Song 2019). This temporal disconnect further confirms the concern that mortality differences are too delayed as an outcome to understand the dynamic relationship between income and health over the lifecycle.

The Role of Chronic Disease in Health Inequality

While we require a health measure that varies earlier in life, any candidate measure should explain a substantial share of the income gradient in mortality. We find that chronic conditions identified from medicine-dispensation data satisfy this criterion throughout the age range examined, and perform well relative to other health-related measures. As mentioned, we discuss the measurement of chronic conditions and its relation to income in detail in Section 4.

Panel A of Figure 3 shows the evidence at age 70, regressing mortality on low-income and a varying set of controls. The five-year mortality rate for low-income individuals is 44 deaths per 1000 higher than for high-income individuals. Controlling for the 22 chronic conditions observed in the prior year reduces this gap to 31 per 1000, implying that measured chronic conditions account for roughly one third of the mortality gap. Adding further lags of chronic conditions does little to further explain the gap, consistent with the persistence of chronic disease over time. Other health-related measures—including non-chronic prescriptions, hospital utilization, and healthcare expenditures—increase the predictive power of the regression, but add relatively little to explaining the income gap once chronic conditions are included.

Panel B of Figure 3 shows that the same pattern holds between ages 40 and 70: chronic conditions explain between 25 and 45% of mortality gaps at all ages. However, as Panel A of Figure 2

8. Following Chetty et al. 2016, we truncate the life-table calculation at age 110. Hence the age window 20-110 yields the overall life-expectancy shortfall at age 20.

already showed, mortality gaps at younger ages contribute little to life expectancy differences, precisely because mortality is rare before later life. Chronic disease is therefore valuable not because of its importance for young-age mortality gaps per se, but because chronic conditions can reveal the underlying health differences that precede the old-age mortality gaps, which do contribute to differences in life expectancy.[9]

Together, these facts motivate our focus on chronic disease as the empirical object for studying when health inequality emerges and how it evolves over the lifecycle.

3 Conceptual Framework

The purpose of this section is to provide a framework for thinking about health inequality in terms of chronic disease, which can be observed early in life and linked directly to the underlying accumulation of health disadvantage. This framework clarifies which components of the health gap can be studied through the construction of a Chronic Disease Index. In particular, we will focus on differences in chronic disease prevalence rather than treatment, on how these chronic conditions accumulate and translate into mortality, and on how they interact with income over the lifecycle.

3.1 Chronic Disease and Mortality

Consider a stylized, linear model of mortality for an individual i at age a:

$$M_{i,a} = \alpha_{i,a} + CC_{i,a}\beta_{i,a} + \epsilon_{i,a} \tag{1}$$

where $M_{i,a}$ denotes mortality risk, and $\epsilon_{i,a}$ is a random error term. The intercept $\alpha_{i,a}$ captures non-chronic health (e.g. infectious diseases) and external factors (e.g., accidental death risk) affecting mortality unrelated to chronic diseases. The vector $CC_{i,a}$ denotes the individual's chronic conditions and the relevant interactions between them. The slope $\beta_{i,a}$ captures the mortality relevance of those chronic conditions.

Comparison across Income We can compare how mortality risk differs across different groups, through their differential burden of chronic disease. The mortality gap at a given age a between

9. See Appendix E for more detail on testing the role of chronic conditions in mortality and healthcare expenditure gaps.

individuals with low versus high income can be decomposed as:

$$\underbrace{M_{L,a} - M_{H,a}}_{\text{Mortality Gap}} = [\alpha_{L,a} - \alpha_{H,a}] + CC_{H,a}\underbrace{[\beta_{L,a} - \beta_{H,a}]}_{\text{Treatment Gap}} + \underbrace{[CC_{L,a} - CC_{H,a}]}_{\text{Prevalence Gap}}\beta_{L,a} \tag{2}$$

where $Z_{Y,a} \equiv E\left(Z_{i,a}|Y_{i,a} \in Y\right)$ denotes the average outcome for individuals with income $Y_{i,a}$ in income group Y at age a. This decomposition illustrates the potential importance of three different factors. First, different income groups can be subject to different chronic conditions ($CC_{i,a}$). Potential reasons include differences in genetics, health behaviours, environmental exposure, and work conditions across income groups, as well as the reverse channel in which individuals' earnings potential depends on their health and ability to work. We refer to these jointly as *prevalence effects*. Second, chronic conditions may have differential health implications ($\beta_{i,a}$) across income groups, for example due to differential access to healthcare or differential treatment of chronic conditions. We refer to these as *treatment effects*. Finally, individuals are exposed to other non-chronic health and external factors ($\alpha_{i,a}$), which may differ across income groups.

Separating the prevalence and treatment gaps is a key conceptual distinction: the former reflects differences in chronic disease burden, while the latter reflects differences in the mortality consequences of a given disease profile. A large prevalence gap points toward interventions that reduce the burden of chronic disease among low-income individuals, while a large treatment gap points toward improving access, take-up, or management of chronic conditions conditional on disease. The respective gaps provide arguably a lower bound on the impact that interventions on chronic conditions for individuals with low income can have on the health gap, simply because improving individuals' health may improve their income trajectory, which can also further improve their health.[10]

Comparison across Ages We can also consider how the relationship between chronic disease and mortality risk varies with age. Returning to equation 1, the expected change in mortality

10. To illustrate this, consider an intervention that changes the incidence of chronic conditions:

$$\frac{\partial M_{i,a}}{\partial CC_{i,a}} = \beta_{i,a} + \left[\frac{\partial \alpha_{i,a}}{\partial Y_{i,a}} + CC_{i,a}\frac{\partial \beta_{i,a}}{\partial Y_{i,a}} + \frac{\partial CC_{i,a}}{\partial Y_{i,a}}\beta_{i,a}\right]\frac{\partial Y_{i,a}}{\partial CC_{i,a}}.$$

The second term captures how much the improvement in health improves an individual's SES and how this further improves her health. This can be through either one of the three factors in the health production function: the incidence of chronic conditions, its treatment or the residual health part. This indirect effect is arguably positive and would add to the direct effect, i.e., $\frac{\partial M_{i,a}}{\partial CC_{i,a}} \geq \beta_{i,a}$. A similar argument can be made for interventions that improve the treatment of chronic conditions.

risk for individuals at ages A and A' can be written as:

$$\underbrace{M_{A'} - M_A}_{\text{Change in Mortality Risk}} = [\alpha_{A'} - \alpha_A] + \underbrace{[CC_{A'} - CC_A]}_{\text{Disease Incidence}}\beta_A + CC_{A'} \underbrace{[\beta_{A'} - \beta_A]}_{\text{Age-Specific Weights}}, \tag{3}$$

where $Z_A \equiv E\left(Z_{i,a}|a = A\right)$ denotes the average outcome for individuals with the same age A. The first term captures changes in residual health and external mortality factors across ages. The second term captures the incidence of new chronic diseases, which tend to accumulate as individuals age and are valued using their mortality consequences at age A. The third term captures changes in the mortality consequences of a given disease profile across ages. The decomposition therefore separates disease incidence from changes in age-specific mortality weights. This age variation matters for the lifecycle interpretation. As Figures 2A and 3B have illustrated, mortality differences become much larger at older ages, reflecting the higher baseline mortality risk in that part of the lifecycle. If the goal is to link chronic disease burden to life expectancy differences, older-age mortality weights are therefore empirically more relevant than weights estimated at young ages, where mortality is rare and mortality coefficients are necessarily smaller and noisier.

3.2 A Chronic Disease Index

To study how health differs across individuals and evolves across ages, we require a common scale on which chronic-condition profiles can be compared across individuals and ages. To implement this idea, we construct a Chronic Disease Index (CDI) using a common set of mortality weights β_A at the reference age A, which we estimate through:

$$M_{i,A} \quad = \quad X_{i,A}\gamma_A + CC_{i,A}\beta_A + \varepsilon_{i,A}. \tag{4}$$

Importantly, since chronic conditions are correlated with socioeconomic status, and because non-chronic determinants of mortality may differ across socioeconomic groups, the control vector $X_{i,A}$ includes income and other SES controls. Otherwise, the estimated chronic-condition weights would partly absorb residual SES-related mortality risk, mechanically building socioeconomic differences into the index. Section 4.4 implements this principle using a double-selection procedure to select the relevant controls before estimating the chronic-condition weights. Under the restriction $E\left(\varepsilon_{i,A} \mid CC_{i,A}, X_{i,A}\right) = 0$, β_A captures the association between chronic conditions

and mortality net of these observables.

We then use these estimated weights to construct a scalar index for each individual i at any age a:

$$CDI_{i,a} \equiv \hat{\alpha}_A + CC_{i,a}\hat{\beta}_A. \tag{5}$$

The coefficient vector $\hat{\beta}_A$ gives the mortality-relevant weight attached to each chronic disease profile at the reference age. The term $\hat{\alpha}_A$ is common across individuals and places the index in age-A mortality-risk units; in the empirical implementation, it corresponds to the non-chronic component evaluated at common control values. Differences in the CDI are therefore driven by differences in observed chronic disease profiles, valued using the same reference-age mortality weights.

The decompositions above underline which components the CDI is designed to isolate - differences in the prevalence of chronic disease when making static comparisons of the disease burdens across groups or differences in the incidence of chronic diseases when making dynamic comparisons. This interpretation has two important qualifications.

First, we construct the CDI using observed, medication-treated chronic disease. This is the empirical counterpart to the residual term in equation 2: if latent chronic disease is not measured in $CC_{i,a}$, part of the mortality gap may appear as residual mortality rather than as a prevalence gap. In addition, the common-weight restriction across income groups requires diagnosed chronic conditions to have similar mortality implications across groups. We therefore examine whether differential under-diagnosis or differential treatment effects affect the interpretation of CDI gaps in Sections 4.2 and 4.3.[11]

Second, the CDI fixes mortality weights at a single reference age. This restriction is deliberate as it defines the CDI as a measure of chronic disease burden on a common mortality-relevant scale, rather than as a measure of actual age-specific mortality risk. Holding the weights fixed means that changes in the CDI over the lifecycle reflect changes in chronic disease profiles, $CDI_{i,a'} - CDI_{i,a} = (CC_{i,a'} - CC_{i,a})\hat{\beta}_A$. The index therefore abstracts from changes in the age-specific mortality consequences of a given disease profile. This is what makes lifecycle

11. Although we find little evidence that either mechanism materially affects our conclusions, any remaining under-diagnosis among low-income individuals would mean that measured $CC_{L,a}$ understates their true chronic disease burden. Similarly, worse treatment or disease management for low-income individuals would mean that common mortality weights understate the mortality relevance of a given low-income disease profile. Both forces would make the measured CDI gap a lower bound: the underlying chronic disease burden, and its mortality relevance, would be larger for low-income individuals than the observed medication-based index implies.

comparisons interpretable: the same chronic disease profile is valued in the same way at each age, so differences in the CDI can be read as differences in chronic disease burden rather than a mixture of disease accumulation and changing mortality weights. Still, we examine in Section 4.3 whether the relation between the chronic conditions and mortality materially changes with age and we account for this when translating simulated CDI paths into survival and life expectancy in the counterfactual analysis in Section 5.4.

3.3 Dynamic Framework: Aging and Sorting

The CDI is not just comparable across individuals and across ages, we can also observe it repeatedly for the same individual, just like income. This allows us to separate differences in health trajectories and income mobility underlying the gap in disease burden across groups. In particular, the observed change in health $dCDI_{Y,a}$ across ages for income group Y can be decomposed as follows:

$$\begin{aligned} dCDI_{Y,a} &\equiv E(CDI_{i,a+1}|Y_{i,a+1} = Y) - E(CDI_{i,a}|Y_{i,a} = Y) && (6) \\ &= \underbrace{E\left(dCDI_{i,a}|Y_{i,a} = Y\right)}_{\text{Biological aging}} + \underbrace{E(CDI_{i,a+1}|Y_{i,a+1} = Y) - E(CDI_{i,a+1}|Y_{i,a} = Y)}_{\text{Health-Based Sorting}}. && (7) \end{aligned}$$

The first term captures how the health of individuals in the income group Y worsens with age driven by the incidence of new chronic conditions, which we will refer to as *(biological) aging*. The second term captures the differences in health due to compositional changes of the income group with age.

The decomposition in equation (7) assumes a balanced panel where all individuals are observed at ages $a+1$ and a. In practice, different cohorts are observed at different ages and thus individuals enter and exit the sample at different ages. We account for these entry and exit effects in our empirical decomposition. In particular, individuals can also exit the sample due to death, and this attenuates the age-profile of the average health measure as those with higher chronic disease burden face higher mortality rates (see Appendix G).

By separating out the health-based sorting and other compositional effects, we can meaningfully compare the biological aging for different income groups. We will re-construct the health measure

for a specific income group as the accumulation of the aging effects between age a_0 and a as

$$CDI^{aging}_{Y,a} = CDI_{Y,a_0} + \sum_{\tilde{a}=a_0}^{a-1} E\left(dCDI_{i,\tilde{a}} | Y_{i,\tilde{a}} = Y\right). \tag{8}$$

We note the parallel between this simulated aging and period life expectancy (e.g., Chetty et al. 2016), which aggregates income-specific mortality rates at different ages. Instead of aggregating mortality rates, the simulated aging aggregates the income-specific incidence of chronic conditions at different ages.

This decomposition approach can also be applied to other measures of socioeconomic status, such as education, parental income, or wealth. While these differences in biological aging should never be interpreted as causal effects of socioeconomic status, they allow us to quantify the potential value of targeting health interventions that reduce the incidence of chronic conditions for a specific group. By focusing on income, our estimate of health-based sorting sheds empirical light on whether health is a meaningful factor of income mobility.

4 Measuring Chronic Disease

The previous section motivates chronic disease as the empirical object through which we study lifecycle health inequality. This section takes that framework to the data. We first describe how we identify the empirical counterpart of $CC_{i,a}$ from dispensed medications and we assess their mortality relevance as well as the limits of this mapping. We then show that income differences in measured chronic disease primarily reflect prevalence differences rather than differential diagnosis, treatment, or disease severity. Finally, we use these results to construct the CDI as a mortality-relevant summary measure of chronic disease burden that can be compared across income groups and ages.

4.1 Identification of Chronic Conditions

We define chronic conditions as 22 medication-identifiable groups, including diabetes, cardiovascular disease, respiratory disease, psychiatric conditions, and chronic pain; see Table 1 for the complete list. The population-wide dispensation records provide annual observations from 2006 onwards, allowing us to follow treated chronic disease repeatedly at scale. Section 2 described the underlying data and its advantages relative to hospital records and surveys.

These data contain the ATC3 codes of all prescribed medications dispensed outside the hospital setting. Several studies have used prescription medication ATC data as indicators of chronic disease. We build on Huber et al. (2013) as our basis to translate medication data into chronic disease indicators.[12] Their mapping extends prior work using medical expertise, reducing Type I errors by focusing on ATC codes that are used more exclusively for the treatment of a given chronic disease. We make a number of additional modifications, using the panel nature of the data to further reduce Type I errors, and reflecting the fact that our data are resolved to the ATC3 level, whereas the mapping in Huber et al. (2013) sometimes uses finer ATC5 resolution.

Our mapping from ATC codes to chronic conditions is given in Table 1, with further description of specific refinements given in Appendix D. Most chronic diseases are related to only one ATC3 code, with cardiovascular disease being the notable exception mapping to several medications. Further, we check whether the medication data contain mortality-relevant ATC codes that are not included in our chronic-condition mapping. We do this by estimating a Lasso regression of five-year mortality on ATC codes directly, allowing the data to select medication groups predictive of mortality. All 25 ATC codes selected by the LASSO procedure are already captured by our refined mapping, suggesting that the mapping does not omit additional medication-based health signals that are important for mortality prediction.

Cancer The main chronic disease that is not comprehensively captured by this approach is cancer. Only around 5% of cancer diagnoses are treated with pharmacy-dispensed medication, and digestive tract and skin cancers account for more than 60% of the cancer diagnoses detected in the medication data. In a companion paper (Danesh et al. 2026), we study the income gradient in cancer using data from the Netherlands Comprehensive Cancer Organization (Dutch: *Integraal Kankercentrum Nederland*, IKNL). These data are rich, but they capture *incident* diagnoses rather than repeated prevalence, and are available for a shorter panel than the medication data used to construct the CDI. We use the 2016–2019 overlap between the cancer registry, medication data, and mortality records to compare the income gradient in the CDI with the mortality-relevant gradient implied by cancer histories in Appendix Section F.1. Appendix Figure F.5 shows that while cancer is a material part of the overall health inequality story, the CDI gradient is almost

12. Huber et al. (2013) apply their mapping to establish new prevalence estimates using Swiss administrative data. Examples in the Netherlands are Lamers and van Vliet (2004), who construct a mapping from ATC codes to 22 chronic conditions for the purposes of risk adjustment in the social health insurance sector, and van Ooijen, Alessie, and Knoef (2015), who use pharmacy-derived chronic conditions to predict an index of self-reported health status over the lifecycle.

five times larger than the cancer-associated mortality gradient, indicating that the CDI captures a substantially broader set of mortality-relevant chronic disease differences across the income distribution.

4.2 Underdiagnosis

Our approach identifies chronic conditions that are actively treated through medication. It therefore does not capture undiagnosed disease, and it does not distinguish the severity of a condition or the effectiveness of treatment conditional on diagnosis. If under-diagnosis (or under-treatment) is more important for low-income groups, we would underestimate both the gap in chronic disease burden and the role of chronic conditions in explaining mortality differences. The relevant question is therefore not whether the CDI measure captures all morbidity, but whether differential under-diagnosis is large enough to materially distort the income gradient.

Our evidence described below suggests that this concern is limited in the Dutch context. This is consistent with the institutional setting, where access to healthcare is broad: only 0.4% of poor households report unmet medical needs, compared to 5.1% in the UK (Eurostat 2023) and 8.5% of *all* households in the US (National Center for Health Statistics 2022).

Survey evidence We compare our medication-based detection of diabetes and cardiovascular disease to self-reported conditions in the *Gezondheidsmonitor* health survey. For diabetes, the concordance is high: of those who self-reported as having diabetes, 85% were detected as taking diabetes medication; conversely, of those detected as taking diabetes medication, 95% self-reported having diabetes. The precision is lower for the set of medications indicating cardiovascular disease, since these include conditions beyond hypertension, while the survey question refers specifically to high blood pressure. The corresponding concordance patterns are reported in Appendix Figure C.1. More importantly, the concordance between self-reported conditions and medicated conditions is stable across income deciles: individuals who self-report diabetes or high blood pressure are similarly likely to receive medication across the income distribution, suggesting limited differential under-treatment conditional on diagnosis.

Differential testing To explore the potential for differential testing leading to under-diagnosis, we use evidence from the smaller-scale *Gezondheidsenquete* health survey, which asks whether

respondents have had a cholesterol, blood pressure, or blood sugar test in the past 12 months. We consider reported test rates among individuals not already prescribed with the relevant medication. The testing and subsequent-prescription patterns are reported in Appendix Figure C.2. Testing is comprehensive: by age 70, around half of those not currently medicated for a condition are tested for it in a given year. Testing trajectories are similar across incomes, with low-income individuals testing at marginally higher rates across ages and conditions. Since low-income individuals may also have higher underlying risk, we also study new prescriptions in the year following the survey. Detection rates are low overall—10–15% for cardiovascular disease and hyperlipidemia, and 2–3% for diabetes—and prescription rates per test are only marginally higher for low-income individuals. This suggests that lower-income individuals are being tested at a rate commensurate with their underlying risk. This is in line with the evidence in (Danesh et al. 2026), showing that the incidence of cancer types differs dramatically across income groups, while the stage of diagnosis for a given cancer type is much more comparable.

Excess mortality The survey evidence above addresses under-diagnosis and under-testing for selected chronic conditions. We complement this with a broader, indirect test for disengagement from the healthcare system using the administrative data. Among individuals with no measured chronic conditions, we distinguish those who take other prescribed medications from those who take no medication at all. The former group is at least engaged with the primary healthcare system, while the latter may include individuals whose conditions are undiagnosed because they are not actively seeking care. Up to age 60, mortality rates among individuals without measured chronic conditions are very similar regardless of whether they take other medication, suggesting little evidence of widespread under-diagnosis during prime ages (See Appendix Figure C.3A). At older ages, mortality rates for the two groups begin to diverge, while the no-medication group becomes smaller (Fig. C.3B) and increasingly concentrated among the lowest-income individuals (Fig. C.4).

The income pattern is concentrated at the bottom of the distribution, as shown in the remaining panels of Appendix Figure C.3. During prime ages, excess mortality among individuals with no medication is visible mainly in the bottom income decile; it is largely absent in the second income decile and remains absent in higher income groups. At older ages, excess mortality among individuals with no medication appears across a broader range of income groups, but the group taking no medication is small at those ages—around one in ten individuals by their

seventies (Fig. C.3B). Taken together, these patterns suggest that our measure may understate chronic disease burden for the lowest-income individuals, especially at older ages, but that the resulting bias is unlikely to overturn the main income-gradient patterns.

4.3 Differential Prevalence Versus Treatment

The motivating evidence in Section 2.2 already showed that measured chronic conditions explain a meaningful share of the mortality gap. As described in Section 3.1, chronic disease can contribute to mortality inequality in two ways: low-income individuals may have higher diagnosed prevalence, or the same diagnosed conditions may have different mortality implications. We now test empirically which channel drives that contribution.

Panel A of Figure 4 reports the prevalence of all 22 chronic conditions by income group at age 70. The income groups are defined as the bottom decile (D1), deciles 2–5, and deciles 6–10. We show the bottom decile separately because the under-diagnosis evidence described above suggests that any medication-based understatement of chronic disease burden is most likely to be concentrated in this group. The overall pattern is clear: the burden of common chronic conditions is higher among low-income individuals. This holds in aggregate and for most specific conditions. For example, prevalence for females is higher in D2–D5 than in D6–D10 for diabetes (14% versus 9%), respiratory illness (14% versus 11%), and cardiovascular disease (54% versus 46%). These conditions show the same pattern for males, and are important in the lifecycle analysis below.

We next assess whether diagnosed chronic conditions have different mortality implications across income groups. Such differences could arise if conditions differ in severity, if treatment differs conditional on diagnosis, or if under-diagnosis varies across the income distribution. We therefore allow the mortality effects of chronic conditions to vary by income group and report the condition-specific estimates in Panel B of Figure 4. These estimates show little systematic evidence that diagnosed conditions carry larger mortality penalties for lower-income individuals, especially when compared with the much larger differences in prevalence shown in Panel A of Figure 4. This already suggests that differences in treatment or severity conditional on diagnosis are not the main empirical channel.

To make this comparison more formal, Panel A of Appendix Figure E.1 reports an Oaxaca-Blinder decomposition following equation 2. The decomposition separates the mortality gap into the part explained by differences in chronic-condition prevalence, the part explained by

differences in the mortality coefficients of those conditions, and residual differences. At age 70, prevalence differences account for about 98.6% of the prevalence-plus-treatment component of the decomposition, while treatment/coefficient differences account for only 1.4%. This confirms that the chronic-disease component of mortality inequality primarily reflects differences in diagnosed prevalence rather than differential mortality effects conditional on diagnosis.

The mortality evidence is therefore suggestive of broadly similar treatment or disease management conditional on diagnosis, but it is not a direct measure of healthcare inputs. Healthcare expenditures provide a more direct corroborating check on treatment intensity. We find that low-income individuals have higher healthcare expenditures than high-income individuals, but this gap is mostly explained by their higher burden of measured chronic conditions. Conditional on chronic conditions, there is little evidence that high-income individuals receive more healthcare; if anything, low-income individuals appear to receive slightly more care for the same measured chronic disease profile. These patterns are reported in Appendix Figure E.2 and further corroborated by the Oaxaca-Blinder decomposition of expenditures in Panel B of Appendix Figure E.1.

Comparison Across Ages The Oaxaca-Blinder decompositions provide similar results at younger ages, indicating that differential prevalence dominates differential treatment at all ages. As developed in Section 3.1, our aim is to place chronic disease burden at different ages on a common mortality-relevant scale. We estimate the weights at age 70 and then apply them across the lifecycle. Accordingly, CDI differences earlier in life represent differences in chronic disease burden evaluated using mortality at age 70 as the reference, rather than predictions of mortality at those earlier ages.

The caveat is that the mortality implications of specific conditions may differ earlier in life, and the age-70 estimation sample excludes individuals who die before reaching that age. Appendix Figure C.5 helps assess this choice by comparing condition-specific five-year mortality coefficients at ages 40, 50, and 60 with the corresponding age-70 coefficients. Conditions with larger age-70 coefficients generally also have larger coefficients at earlier ages, while earlier-age estimates are much smaller and noisier because mortality is rare.

Taken together, the results suggest that mortality differences associated with chronic disease primarily reflect differences in prevalence rather than differences in treatment, supporting the

use of common mortality weights in the CDI. Moreover, age 70 is a sensible reference age, when mortality rates are high and chronic disease is strongly predictive of mortality.

4.4 Construction of the Chronic Disease Index

Given the evidence in the prior subsection, we construct a summary health measure comparable across ages by estimating how chronic conditions predict five-year mortality at age 70 under the common-weight structure in equation 4. We nevertheless adjust flexibly for income and other socioeconomic predictors so that the chronic-condition weights do not absorb mortality risks arising through non-chronic socioeconomic channels.

We adapt the double-selection procedure of Belloni, Chernozhukov, and Hansen (2014). Their insight is that estimating the mortality model in a single Lasso step could omit relevant socioeconomic controls if those controls are highly collinear with the chronic conditions themselves. Instead, the control set should include variables selected both for predicting mortality and for predicting the focal health variables. In our setting, the focal object is not a single chronic condition but a vector of 22 chronic-condition indicators, $CC_i = \{c_i^1, \ldots, c_i^{22}\}$. We therefore run separate selection steps for mortality and for each chronic condition, before selecting the relevant interactions among chronic conditions.

Omitting age subscripts for convenience, the selection procedure is:

$$\begin{aligned} M_i &= X_i'\theta_m + \zeta_i, \\ c_i^k &= X_i'\theta_c^k + \nu_i^k \qquad \forall\ k \in \{1, \ldots, 22\}, \\ M_i &= f(CC_i)'\theta_v + \varepsilon_i, \end{aligned} \tag{9}$$

where M_i is an indicator for five-year mortality, c_i^k is an indicator for the kth chronic condition, and X_i denotes the full set of potential socioeconomic controls. The function $f(CC_i)$ contains the chronic-condition features considered for the index: three years of chronic-condition indicators, within-condition interactions across lags, and two-way cross-condition interactions for the most recent lag.

The selected socioeconomic controls are the union of variables selected in either the mortality equation or any of the chronic-condition equations:

$$X_i^* = \left\{ x_j \in X_i : \widehat{\theta}_{m,j} \neq 0 \text{ or } \widehat{\theta}_{c,j}^k \neq 0 \text{ for some } k \right\}.$$

The selected chronic-condition terms are the union of the baseline chronic-condition indicators and the additional interaction terms selected in the final Lasso step:

$$f(CC_i)^* = CC_i \cup \left\{ z_j \in f(CC_i) : \widehat{\theta}_{v,j} \neq 0 \right\}.$$

We then include these selected variables in the final prediction model:

$$M_i = X_i^{*\prime}\beta_X + f(CC_i)^{*\prime}\beta_{CC} + \zeta_i. \tag{10}$$

The final model is estimated as a linear probability model, separately by gender, with calendar-year fixed effects absorbed. We estimate the model on a 50% random sample of 70-year-olds in income deciles 2–10, excluding the bottom decile from the estimation because of the under-diagnosis concerns discussed in Section 4.2 above, and then apply the resulting weights to all individuals. Diagnostic results are reported for the hold-out sample, see Appendix F.

Using the resulting estimates, we calculate the health measure for an individual at any age a as:

$$CDI_{i,a} \equiv \bar{X}^{*\prime}\widehat{\beta}_X + f(CC_{i,a})^{*\prime}\widehat{\beta}_{CC}, \tag{11}$$

where $\bar{X}^{*\prime}\widehat{\beta}_X$ captures mean socioeconomic effects. The measure therefore re-weights chronic illnesses observed at any point in life according to their contribution to five-year mortality risk at age 70.

4.5 Results and Caveats

The CDI captures substantial variation in chronic disease burden: among 70-year-olds, predicted five-year mortality differs sharply between the healthiest and sickest parts of the CDI distribution. The predicted mortality rate is below 4.8% for the bottom decile, while it is above 17.2% for the top decile of the distribution (Appendix Figure F.2).

The double-selection procedure in the construction of the CDI matters quantitatively. Overall, a 'naively' estimated measure without socioeconomic controls would overstate the chronic disease health gap by approximately 16%, as shown in Panel A of Appendix Figure F.3. Diagnostic checks confirm that, after accounting for the socioeconomic correction, prediction errors are close to zero across the range of CDI predictions and do not diverge between low- and high-income

individuals, as shown in Panel B of Appendix Figure F.3.

Figure 5 shows the resulting condition-specific CDI weights and compares them with coefficients from simpler specifications. For each condition j, the CDI weight is as follows:

$$\hat{\beta}^j_{70} = E[CDI_{70} \mid c^j_{69} = c^j_{68} = c^j_{67} = 1, c^{-j} = \overline{c^{-j}}] - E[CDI_{70} \mid c^j_{69} = c^j_{68} = c^j_{67} = 0, c^{-j} = \overline{c^{-j}}]$$

Thus, the figure illustrates the socioeconomic correction for each condition separately, and the importance of accounting for comorbidity patterns rather than attributing mortality risk to each condition in isolation. The corrections seem most consequential for high-mortality and high-prevalence conditions. For high-mortality conditions like dementia and anemia, the CDI weights are substantially lower than simpler estimates that omit socioeconomic corrections. The adjustments for comorbidities are even larger, in particular for high-prevalence conditions. For example, the estimates for diabetes and cardiovascular disease roughly double when comorbidities are ignored.

Of course, the condition weights remain associational. Some conditions, such as migraine or intestinal inflammatory diseases, appear protective in the mortality model, which may reflect selection into diagnosis and treatment as well as residual confounding. While the socioeconomic and comorbidity adjustments reduce this concern, the weights should still be interpreted as mortality-relevant associations rather than causal effects of each condition.

We also find that the CDI correlates with self-reported health among respondents in the *Gezondheidsmonitor* Health Monitor survey ($r = 0.36$). Low-income individuals report worse health at any given CDI level, but the relationship between the CDI and self-reported health is similar across income groups (Appendix Figure F.4). The CDI model's overall predictive power is relatively low, as shown in Appendix Table F.1. While prediction models that also include hospital use, non-chronic prescriptions, and healthcare expenditures would have higher predictive power, they do not materially help to explain the mortality gap across income groups, as discussed in the motivating evidence (see Figure 3). Taken together, the CDI is a mortality-relevant summary measure of chronic disease burden, designed to compare observed chronic-condition profiles on a common scale across individuals, ages, and socioeconomic groups. It is not an exhaustive account of health or a causal parameter, and it is not intended to maximise individual-level mortality prediction. Additional details on the prediction model, including selected predictors, lag structure, and robustness to alternative specifications, are provided in Appendix Section F.

5 Health Inequality over the Lifecycle

We now use the CDI to study how the health gap evolves over the lifecycle. We first document when the gap materializes, then decompose its growth into differential aging and health-based sorting, and explore the contribution of specific conditions by age and channel. We then examine why these mechanisms generate a strongly non-linear gradient across the income distribution, and provide a counterfactual analysis of disease burden and healthcare costs under alternative aging scenarios.

5.1 Health Gap over the Lifecycle

We begin with the comparison across ages motivated by the conceptual framework in Section 3.1. Panel A of Figure 6 plots the average CDI for the low- and high-income groups at different ages, pooling all observations in our sample for each age.[13] The striking feature is how early the gap appears. At age 20, the difference is still small—0.08 percentage points—but already interpretable as a difference in age-70 five-year mortality risk. By age 30, the gap has grown to more than 0.30 percentage points, and it continues widening through midlife. By around age 60, however, the gap no longer opens appreciably among surviving individuals, even though mortality differences themselves are still largely realized later. This stabilization shows that the measured chronic-disease gradient is largely in place before most mortality differences are observed. Still, the plateau should be interpreted cautiously because mortality selection may change the composition of survivors at older ages.[14]

The pattern in Panel A of Figure 6 can be expressed horizontally as a "biological age" gap. By age 35, low-income individuals have an average chronic disease burden equivalent to that of high-income individuals aged 50. This comparison illustrates the scale of the accumulated health disadvantage: lower-income individuals reach the same mortality-relevant chronic disease burden roughly 15 years earlier. This again confirms that the health gradient is not simply a late-life phenomenon that appears when mortality becomes common; it forms during early adulthood and is largely established before old-age mortality is realized.[15]

13. Because these profiles combine different birth cohorts at each age, they are descriptive age profiles rather than trajectories followed by a single cohort; Panel F of Appendix Figure C.6 shows similar divergence within selected cohorts observed from 2009 to 2021.

14. This evidence is complemented with direct reference-age checks in Appendix Figure C.5 and Appendix Table F.1, suggesting that the lifecycle pattern does not depend on fitting the CDI weights at exactly age 70.

15. This early health gap is also visible in self-reported health, as shown in Appendix Figure F.4. However,

Panels B of Figures 2 and 6 make this timing point more directly. Comparing below-median and above-median income groups in Figure 6, we find that by age 40 the CDI gap is already 48 percent of the gap observed at age 70. In other words, about half of the chronic disease burden gap at age 70 has already materialized by age 40. Mortality rates give a very different impression: at age 40, the mortality gap is only 7 percent of the mortality gap at age 70. Comparing individuals across the income distribution in Figure 2, we already established how the life-expectancy shortfall is sharply concentrated at the bottom of the income distribution and mostly realized after age 70 as part of the motivating evidence in Panel A. Panel B of Figure 2 adds the corresponding lifecycle evidence for the CDI. The distributional shape mirrors the life-expectancy shortfall profiles in Panel A. At the same P3–P4 income bin used above, excess CDI is 7 percent of its age-70–79 level at ages 20–29, 48 percent at ages 30–39, and 76 percent at ages 40–49; by ages 50–59 it is already as large as at ages 70–79. By comparison, only 2 percent of the eventual life-expectancy shortfall at this point is realized before age 50 and 20 percent before age 70, with the remaining 80 percent realized after age 70. Thus, the gradients are similar in shape but differ sharply in timing.[16]

Overall, the CDI demonstrates that much of the mortality-relevant health gap is already embedded in chronic disease profiles decades earlier.

Separate Chronic Conditions We consider six of the chronic conditions that contribute the most to the CDI gap and compare their contributions at ages 30 and 70 in Panel A of Figure 7. The contribution is $\kappa_a^j = (CC_{L,a}^j - CC_{H,a}^j) \cdot \hat{\beta}_{70}^j$, where $CC_{L,a}^j$ and $CC_{H,a}^j$ are the shares of low- and high-income individuals with condition j, and $\hat{\beta}_{70}^j$ is the condition's CDI weight defined in Figure 5.

At age 30, mental-health conditions—especially psychoses and psychological disorders—account for the largest shares of the gap. By age 70, respiratory illness and diabetes have become the most important, while cardiovascular disease, pain, and psychoses make similarly sized contributions. Thus, the condition-specific composition of the CDI gap shifts markedly between early and later adulthood.

The contribution profiles are generated by income differences in the prevalence of these same six

self-reported health is categorical rather than cardinal, subject to sampling error, and available only in repeated cross-sections, so it does not provide the same mortality-equivalent interpretation as the CDI.

16. The peak in the life-expectancy shortfall occurs around percentiles 3–4, whereas the excess CDI profile peaks somewhat higher in the income distribution, around percentiles 7–8. This imperfect alignment is consistent with some differential underdiagnosis at the bottom of the distribution, but suggests that its effect on the overall gradient is moderate. The excess-mortality evidence in Section 4.2 provides a more direct check on this concern.

chronic conditions. Appendix Figure C.7 traces those underlying lifecycle prevalence profiles over the lifecycle. Cardiovascular disease and diabetes are rare at young ages but rise steeply after midlife, with the low-income group pulling away increasingly strongly from around ages 40 to 50. Pain and respiratory disease show earlier and more persistent income gaps: respiratory disease is already more prevalent among low-income individuals in childhood and early adulthood, while pain rises for both groups from young adulthood but remains consistently higher for the low-income group. The mental-health conditions follow a different pattern. Psychological disorders increase sharply from early adulthood and remain substantially more common among low-income individuals throughout adulthood, while psychoses show a pronounced low-income excess from early adulthood through midlife before declining at older ages.[17]

Further Heterogeneity We can extend our analysis to other socioeconomic measures and for different sub-groups, but the overall empirical patterns are robust. Panel B of Appendix Figure C.6 shows that the differences are somewhat larger for men than for women with different dynamic patterns around child-bearing ages. Panels C and D of the same figure plot the average CDI by education groups and mother's household income respectively, noting that this is equivalent to father's household income for those parents who remain together. The determinants of parental income typically long predate health and economic outcomes of the adult children, making it a less endogenous proxy for socioeconomic background than own income. Education is also likely less endogenous than own income, but still may itself respond to health earlier in life. These socioeconomic measures are more stable at older ages, but the coverage for both variables decreases with age. We again find very large differences overall. We find somewhat larger differences in CDI already at 20, especially when considering high-school drop outs to the others, and a less dramatic opening of the gap in CDI in early adulthood. Panel E shows the split by net wealth, which is again endogenous over the lifecycle, but allows us to make a meaningful comparison beyond age 78 (which is the oldest age at which we can measure pre-retirement income). Interestingly, like for income, the gap in CDI between wealth groups does not meaningfully increase after retirement ages. Moreover, for both wealth groups, the average CDI decreases for the individuals who survive beyond 85 and older. Finally, Panel F zooms in on specific cohorts, splitting them into low- and high-income groups in 2009 and then

17. Panel A Appendix Figure C.8 provides the full age profiles of the contributions to the CDI gap underlying the endpoint comparison in Figure 7, showing how the condition-specific contributions evolve smoothly between ages 30 and 70, rather than changing discretely between the two reported ages.

showing how the CDI diverges for the respective income groups until 2021. Overall, these panels highlight that the difference in age-gradients between socioeconomic groups partly depends on compositional changes, and thus do not only capture differences in biological aging.

5.2 Differential Aging and Health-Based Sorting over the Lifecycle

The lifecycle profiles above show that the health gap opens early and grows steadily through adulthood. However, these age profiles do not by themselves reveal why the gap grows. As discussed in Section 3.3, a steeper CDI profile for low-income individuals could reflect faster accumulation of chronic conditions within income groups, but it could also reflect health-based movement across the income distribution. Individuals in poor health may be more likely to enter or remain in low-income groups, while healthier individuals may be more likely to move into higher-income groups.

We separate these two forces using the decomposition in equation (7), distinguishing differential aging from health-based sorting. We also account for the unbalanced nature of the sample across ages by separating out attrition due to mortality and cohort effects; the full decomposition is described in Appendix G.

Table 2 reports the age-specific and overall lifecycle results. Both differential aging and health-based sorting contribute substantially to the CDI gap observed at old age. Over the lifecycle, the two components together contribute 2.40 percentage points.[18] Differential aging accounts for 58 percent of this contribution (1.39 percentage points), and health-based sorting for the remaining 42 percent (1.01 percentage points). Their combined contribution exceeds the observed 1.21-percentage-point increase in the CDI gap because cohort effects and mortality attrition dampen the measured gap. Mortality attrition is mechanically important: low-income individuals with high CDI values are more likely to die, which lowers the average CDI of surviving low-income individuals relative to the initial underlying population. Attrition and cohort effects are particularly important at older ages and contribute to the stabilization of the CDI gap after age 65. Appendix Figure G.1 plots the four decomposition components as a function of age for the low- and high-income groups separately.

Panel A of Figure 8 illustrates the decomposition graphically. We simulate the mean CDI for each income group first using only the aging component, following equation (8), and then using

18. The CDI is expressed on the scale of fitted five-year mortality risk at age 70. A difference of 0.01 in the CDI therefore corresponds to one percentage point on that scale.

both aging and health-based sorting channels. The aging-only curves show a steadily widening gap over the lifecycle. The increase is modest in early adulthood: the aging gap rises by 0.10 percentage points between ages 20 and 30. It becomes much larger later in life, increasing by 0.47 percentage points between ages 60 and 70. Once sorting effects are included, the rapid opening of the CDI gap during early adulthood becomes apparent. This pattern indicates that early-life divergence is shaped importantly by health-based sorting, while differential aging accumulates more steadily over the lifecycle.

This distinction changes the interpretation of the lifecycle health gap. Health-based sorting is especially important in early adulthood, when individuals enter the labour market and establish their income rank. Although household composition also changes substantially at these ages, the sorting estimates remain similar when we control for it. The early-adult pattern therefore does not appear to be driven mechanically by individuals leaving the parental household or by differential household formation at those ages (see Appendix Table G.2).[19] Differential aging, by contrast, captures the persistent excess accumulation of chronic conditions among lower-income individuals. The decomposition therefore shows that the old-age health gap reflects both faster biological deterioration and the movement of less healthy individuals into lower income groups.

The decomposition so far has described vertical gaps in CDI at a given age. Appendix Figure C.9 uses horizontal differences between the simulated aging-only CDI profiles to express the remaining gap in terms of biological age. These biological-age profiles intentionally exclude mortality attrition: they describe how a low-income person ages over the lifecycle, rather than those who survive long enough to be observed. The biological-age gap increases for young adults before slowly narrowing around midlife. Using observed CDI, low-income individuals reach the average chronic disease burden of high-income 50-year-olds by age 35. Using the aging-only simulation, this threshold is reached around age 40. Health-based sorting therefore explains part of the early biological-age gap, while differential aging accounts for the persistent divergence that accumulates over the rest of the lifecycle.[20]

19. The importance of sorting at the start of the career is consistent with evidence that health affects earnings at labour-market entry (e.g., O'Donnell, Van Doorslaer, and Van Ourti 2015). Looking at finer age bins in Appendix Figure G.1, we also see the importance of health-based sorting near the end of the career, consistent with evidence that poor health contributes to premature labour-market exit (e.g., Blundell et al. 2021; Kolsrud et al. 2024).

20. Appendix Figure G.2 shows the corresponding profiles after adding mortality attrition back in for the low- and high-income groups.

Separate Chronic Conditions Panel B of Figure 7 repeats the decomposition into differential aging and health-based sorting, but for the six main chronic conditions separately. The contrast between the condition-specific decompositions is sharp. Cardiovascular disease and diabetes are mostly differential-aging conditions: 84 and 82 percent of their selected-condition contribution is attributed to differential aging. Respiratory illness is also more strongly related to differential aging, though less exclusively (62 percent). By contrast, psychoses and psychological disorders are overwhelmingly associated with health-based sorting, with sorting accounting for 82 and 88 percent of their contributions respectively. Pain sits between these cases, with a slight majority of its contribution associated with sorting. Thus the conditions that dominate the old-age CDI gap are mainly diseases that accumulate differentially with age, while the mental-health conditions that matter earlier in life operate primarily through health-based sorting.[21]

Robustness Our baseline income categorization uses average income two to four years prior to the age at which health is measured. This follows prior work and mitigates the influence of very recent health shocks on income. However, lagging income is not sufficient to remove persistent health-based sorting: individuals with worse health may remain in lower-income groups for many years. The dynamic decomposition is designed precisely to separate these persistent sorting effects from differential aging. We test the robustness of the results using alternative income lags and moving-average windows, a Markov restriction requiring individuals to remain in the same income group for two consecutive years, and contemporaneous income definitions. The aging-sorting split is stable across the sensitivity checks: relative to the baseline aging share of 57.7 percent, the Markov restriction raises the share by 1.3 percentage points, while using contemporaneous income lowers it by 1.4 percentage points (Appendix Figure G.3A). We discuss these robustness checks in more detail in Appendix G.

5.3 Non-linearity across the Income Distribution

Since the income gradient in CDI is non-linear, we next examine how the dynamic decomposition varies across the income distribution and which components account for the CDI's non-linearity. Panel B of Figure 2 already showed some important non-linearity in the lifecycle profile of chronic disease burden: excess CDI relative to the top income percentile is concentrated at the

21. Panel B of Appendix Figure C.8 focuses on the timing over the lifecycle and shows the condition-specific relative contributions to the aging shares at different ages; cardiovascular disease and diabetes make their largest contributions to differential aging in midlife, whereas respiratory illness becomes increasingly important at older ages.

bottom of the distribution and especially so for the mid-age ranges. This motivates moving beyond the low- versus high-income comparison and decomposing the gradient across finer income ranks.

Panel B of Figure 8 decomposes cumulative CDI changes across income ventiles (i.e., 20 quantiles). The cumulative aging effect is roughly twice as large in the bottom income ventile as in the top ventile, with a broadly smooth gradient across the intervening ventiles. Health-based sorting increases the CDI only in the first five ventiles, with the greatest effects in the lowest ventiles, and decreases it in the remaining ventiles. Mortality attrition lowers the CDI throughout the distribution, with the largest effects at the bottom, and thus offsetting some of the non-linearity due to health-based sorting.

Together, these results show that differential aging explains the broad socioeconomic gradient, whereas health-based sorting creates its sharp bottom-tail concentration.[22]

5.4 Counterfactual Analysis

Income-related inequalities in health and longevity are a longstanding policy concern. Understanding when these inequalities emerge and through which dynamic processes they develop helps identify the margins—and the stages of the lifecycle—at which interventions may have the greatest potential. Our decomposition points to two such margins: reducing the faster accumulation of chronic disease among low-income individuals, and limiting the extent to which poor health is associated with downward income mobility.

The relevance of these processes extends beyond individuals' economic resources and health, since they are also associated with government spending on healthcare and social transfers. Appendix Figure H.1 provides descriptive evidence on these broader economic gradients. Panel A ranks households by income percentile and shows how lower-income individuals face greater chronic disease burden and higher healthcare spending. Between the 10th and 90th income percentiles, cumulative healthcare spending between ages 25 and 65 falls by approximately one half. Panel B provides the complementary perspective by ranking individuals according to their CDI within gender–age cells. Total household disposable income falls and social transfers rise with chronic disease burden, with both gradients becoming particularly pronounced in the upper

22. The corresponding quintile-by-age decomposition in Panel B of Appendix Figure G.3 shows that bottom-quintile sorting is strongest in early adulthood, when individuals are forming households, entering stable employment, and establishing their own income ranks.

tail of the CDI distribution. Total social transfers between ages 25 and 65 average €176,200 at the bottom quintile of the CDI distribution, more than tripling to €660,100 at the very top. These patterns provide descriptive context for the counterfactual exercise and illustrate the broader economic stakes associated with health inequality.

Against this broader backdrop, our counterfactual analysis focuses more narrowly on changes in chronic-disease accumulation. It constructs income-group-specific CDI trajectories and asks how the later-life path of the low-income group would differ if its chronic-disease accumulation followed the high-income aging rate from different starting ages. This is necessarily a partial exercise, rather than a full simulation of the dynamic channels linking income, health, and mortality. It nevertheless quantifies the potential importance of differential aging, and highlights the role of timing: if mortality differences are mostly realized late in life while chronic-disease gaps form much earlier, interventions that wait until late adulthood will miss much of the relevant disease accumulation.

For the counterfactuals, we need a mapping from counterfactual CDI paths into life expectancy. The fixed age-70 weights place chronic disease burden at every age on a common CDI scale; they do not determine how a given CDI level maps into mortality at each age. To make that link, we regress one-year mortality on the CDI by age and gender including income fixed effects.[23] We then aggregate counterfactual mortality rates into an estimate of period life expectancy at age 40, following Chetty et al. 2016.[24]

We can now separate out the effect of the differential accumulation of chronic conditions on life-expectancy. The counterfactual baseline starts from each group's observed mean CDI at age 20 and then cumulates the estimated aging, like in Figure 8A.[25] Applying the same mortality mapping and aggregation to these new simulated CDI paths yields life expectancies of 81.2 and 85.3 years for below-median and above-median income respectively, and hence a 4.1-year gap (Table 3). We can then test three counterfactual scenarios to replace the low-income aging component with the high-income component from age 20, 40, or 60 onward.

23. We thus preserve the common CDI scale to aggregate the various chronic conditions at all ages, following the evidence in Section 4.3, but then allow the CDI's relationship with mortality to vary flexibly with age.

24. Between ages 40 and 78, we use the estimated mortality rates. For ages 79 to 90, we use a Gompertz extrapolation. For ages 91 to 110, we revert to observed mortality rates for the full sample. See Appendix H for further details. The aggregation implies a life expectancy of 80.8 years for low-income individuals and 85.2 years for high-income individuals, a gap of 4.4 years (Appendix Table H.1).

25. We retain mortality attrition because it would be implausible to simulate the later-life path of individuals who, under the relevant conditions, would already have died. We note that the greater attrition of people with high CDI may reflect differential aging, health-based sorting, and cohort effects, which cannot be cleanly separated here. Appendix Figure G.2 illustrates the effect of adding mortality attrition to the CDI paths.

Starting at age 20 raises low-income life expectancy by 1.3 years and closes 31 percent of that gap. This is an important improvement, but still far from fully closing the gap, since chronic conditions only explain part of the mortality differences. Importantly, starting at age 40 raises life expectancy to 82.3 years and closes 27 percent. However, waiting until age 60 closes only 8 percent. As expected, waiting until this age has missed much of the relevant disease accumulation.

We can also consider the impact of the counterfactual on the 'biological age' of the low-income group relative to the high-income group, to illustrate its implications for health over the life course rather than for life expectancy alone. Intervening from age 20 avoids the otherwise steady increase in the biological-age gap in early adulthood, as shown in Table 3 . Intervening from age 40 misses this early accumulation, but still has enough time to close much of the gap by age 70 (see also Appendix Figure C.9). The contrast between the age-20 and age-40 counterfactuals is therefore mainly about the years lived with excess chronic disease burden, while the contrast with the age-60 counterfactual is about both accumulated morbidity and lost life-expectancy gains.

A similar mapping and aggregation for healthcare spending as for the mortality rates then allows us to also estimate the gap in lifetime healthcare costs. The exercise therefore combines two offsetting forces: lower chronic-disease burden reduces annual healthcare spending, while improved survival raises the number of years over which costs are incurred. Table 3 reports that lifetime healthcare costs are only 1.2 percent higher for low-income individuals than for high-income individuals (€157,700 and €155,900 respectively). While costs are substantially higher for low-income individuals at any given age, the lifetime difference is muted, simply because of their shorter life expectancy. The same countervailing effects appear in the counterfactuals: equalizing the aging process produces cost savings of less than 3 percent even when starting at age 20.

Taken together, the counterfactual scenarios show that earlier reductions in chronic disease accumulation are associated with larger life-expectancy gains than later reductions. Starting earlier also matters because it avoids many years of excess biological age, even when middle-age intervention captures a substantial share of the eventual life-expectancy gain. The savings on lifetime healthcare expenditures, however, are more limited.

6 Conclusion

Health inequality forms over many years, even though most mortality differences appear only late in life. Mackenbach (2019, p. 178) summarizes the broader challenge: "We know that the explanation of health inequalities involves three basic mechanisms: direct causation, reverse causation, and confounding (due to selection on personal characteristics during social mobility). This was already known when I started to work in this area in the late 1980s, but after decades of research we still do not know what the relative importance of each of these mechanisms is." Our analysis does not fully resolve this causal question, but it makes progress on the empirical problem that underlies it. By measuring income, chronic disease, healthcare use, and mortality for an entire population in the same setting, we show when chronic disease inequality forms and separate its growth into differential aging and health-based sorting.

By measuring mortality-relevant chronic disease before death, the CDI reveals a central lifecycle fact that mortality gradients obscure: the gap is already substantial by midlife. Mortality *after age 70* accounts for about 80% of life expectancy gaps across income groups. In contrast, nearly half of the old-age gap in chronic disease burden is formed *before age 40*. This shifts the empirical focus from when mortality differences are realized to when the underlying chronic disease gap is formed.

Both differential aging and health-based sorting contribute substantially to the growth of this gap, but the contribution of differential aging is larger. The relative importance of the two components is strikingly different across income ranks and chronic conditions. The cumulative aging effect declines relatively smoothly with income. In contrast, individuals in poor health do not simply experience downward income mobility, but disproportionately move to the bottom of the income distribution. At the condition level, psychoses and psychological disorders contribute most to health-based sorting, while cardiovascular disease and diabetes contribute most to differential aging. Taken together, these results provide a unified account of two central features of health inequality: differential aging accounts for the broad socioeconomic gradient, while health-based sorting accounts for its sharp concentration at the bottom.

The counterfactual analysis quantified how much the disparities in life expectancy relate to differences in aging. While descriptive, it illustrates how strongly these model-implied life expectancy gains depend on the timing of intervention. These results also reinforce the importance of two complementary research agendas concerned with the upstream formation of the

income-health gradient. The first is to identify the early- and mid-life exposures, treatments, and institutions that account for differences in chronic disease incidence and progression across socioeconomic groups. The second is to determine how labour-market institutions can limit the extent to which deteriorating health translates into lower income and whether social insurance provides sufficient protections. Because the CDI can be measured repeatedly in administrative panels, it provides a way to quantify the importance of these processes and identify when they operate—before their effects appear in mortality.

References

Abdalla, Salma M, Samuel B Rosenberg, Nason Maani, Catalina Melendez Contreras, Shui Yu, and Sandro Galea. 2025. "Income, education, and the clustering of risk in cardiovascular disease in the US, 1999–2018: an observational study." *The Lancet Regional Health–Americas* 44.

Adda, Jérôme, James Banks, and Hans-Martin von Gaudecker. 2009. "The Impact of Income Shocks on Health: Evidence from Cohort Data." *Journal of the European Economic Association* 7 (6): 1361–1399. ISSN: 1542-4766, accessed December 18, 2023. JSTOR: 40601206.

Avendano, Mauricio, M Maria Glymour, James Banks, and Johan P Mackenbach. 2009. "Health disadvantage in US adults aged 50 to 74 years: a comparison of the health of rich and poor Americans with that of Europeans." *American journal of public health* 99 (3): 540–548.

Banks, James, Alastair Muriel, and James P Smith. 2010. "Disease prevalence, disease incidence, and mortality in the United States and in England." *Demography* 47 (Suppl 1): S211–S231.

Bauer, Ursula E, Peter A Briss, Richard A Goodman, and Barbara A Bowman. 2014. "Prevention of Chronic Disease in the 21st Century: Elimination of the Leading Preventable Causes of Premature Death and Disability in the USA." *The Lancet* 384, no. 9937 (July): 45–52. ISSN: 0140-6736, accessed May 9, 2024. https://doi.org/10.1016/S0140-6736(14)60648-6.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28, no. 2 (May): 29–50. ISSN: 0895-3309, accessed December 18, 2023. https://doi.org/10.1257/jep.28.2.29.

Black, Sandra E, Neil Duzett, Adriana Lleras-Muney, Nolan Pope, and Joseph Price. 2024. "Intergenerational Transmission of Lifespan in the US." *NBER Working Paper* 31034 (March).

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2015. "Losing Heart? The Effect of Job Displacement on Health." *ILR Review* 68, no. 4 (August): 833–861. ISSN: 0019-7939, accessed April 22, 2024. https://doi.org/10.1177/0019793915586381.

Bleich, Sara N, Marian P Jarlenski, Caryn N Bell, and Thomas A LaVeist. 2012. "Health inequalities: trends, progress, and policy." *Annual review of public health* 33 (1): 7–40.

Blundell, Richard, Jack Britton, Monica Costa Dias, and Eric French. 2021. "The Impact of Health on Labor Supply Near Retirement." *Journal of Human Resources* (January). ISSN: 0022-166X, 1548-8004, accessed April 22, 2024. https://doi.org/10.3368/jhr.58.3.1217-9240R4.

Bolt, Uta. 2022. *What Is the Source of the Health Gradient? The Case of Obesity.* Working Paper. Accessed April 22, 2024.

Borella, Margherita, Francisco A Bullano, Mariacristina De Nardi, Benjamin Krueger, and Elena Manresa. 2024. *Health inequality and health types.* Technical report. National Bureau of Economic Research.

Buttorff, Christine, Teague Ruder, and Melissa Bauman. 2017. *Multiple Chronic Conditions in the United States.* RAND Tool TL-221-PFCD. Santa Monica, CA: RAND Corporation. https://doi.org/10.7249/TL221. https://www.rand.org/pubs/tools/TL221.html.

Campbell, Emma, Ellie Macey, Chris Shine, Vahé Nafilyan, Nathan Cadogan Clark, Piotr Pawelek, Isobel Ward, Andrew Hughes, Veena Raleigh, Amitava Banerjee, et al. 2023. "Sociodemographic and health-related differences in undiagnosed hypertension in the health survey for England 2015–2019: a cross-sectional cohort study." *EClinicalMedicine* 65.

Case, Anne, and Angus S. Deaton. 2005. "Broken Down by Work and Sex: How Our Health Declines." In *Analyses in the Economics of Aging,* 185–212. University of Chicago Press, August. Accessed May 9, 2024.

Case, Anne, Angela Fertig, and Christina Paxson. 2005. "The Lasting Impact of Childhood Health and Circumstance." *Journal of Health Economics* 24, no. 2 (March): 365–389. ISSN: 0167-6296, accessed December 18, 2023. https://doi.org/10.1016/j.jhealeco.2004.09.008.

Case, Anne, Darren Lubotsky, and Christina Paxson. 2002. "Economic status and health in childhood: The origins of the gradient." *American economic review* 92 (5): 1308–1334.

Centers for Disease Control and Prevention. 2026a. *Fast Facts: Health and Economic Costs of Chronic Conditions,* May 26, 2026. Accessed July 2, 2026. https://www.cdc.gov/chronic-disease/data-research/facts-stats/index.html.

———. 2026b. *National Diabetes Statistics Report,* January 21, 2026. Accessed July 2, 2026. https://www.cdc.gov/diabetes/php/data-research/index.html.

Chen, Yiqun, Petra Persson, and Maria Polyakova. 2022. "The roots of health inequality and the value of intrafamily expertise." *American Economic Journal: Applied Economics* 14 (3): 185–223.

Chetty, Raj, Janet Currie, John N. Friedman, Ines Guix Sauquet, Nathaniel Hendren, Michael Stepner, Harvey Barnhard, Dhruv Gaur, Tyler Jacobson, Emma Lee, et al. 2025. *Growing Class Gaps and Shrinking Race Gaps in Life Expectancy, 2001-2019.* Work in Progress, presented at the National Bureau of Economic Research.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016. "The Association Between Income and Life Expectancy in the United States, 2001-2014." *JAMA* 315, no. 16 (April): 1750–1766. ISSN: 0098-7484, accessed December 18, 2023. https://doi.org/10.1001/jama.2016.4226.

Currie, Janet. 2009. "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development." *Journal of Economic Literature* 47, no. 1 (March): 87–122. https://doi.org/10.1257/jel.47.1.87. https://www.aeaweb.org/articles?id=10.1257/jel.47.1.87.

Currie, Janet, and Mark Stabile. 2003. "Socioeconomic status and child health: why is the relationship stronger for older children?" *American Economic Review* 93 (5): 1813–1823.

Cutler, David M., Adriana Lleras-Muney, and Tom Vogl. 2011. "Socioeconomic Status and Health: Dimensions and Mechanisms." In *The Oxford Handbook of Health Economics,* edited by Sherry Glied and Peter C. Smith. Oxford University Press, April. ISBN: 978-0-19-923882-8, accessed December 18, 2023. https://doi.org/10.1093/oxfordhb/9780199238828.013.0007.

Danesh, Kaveh, William Parker, Mieke Aarts, Jonathan Kolstad, and Johannes Spinnewijn. 2026. "Socioeconomic differences in cancer incidence, stage, and survival in the Netherlands, 2006-2021." Working Paper.

Darden, Michael, Donna B. Gilleskie, and Koleman Strumpf. 2018. "Smoking and Mortality: New Evidence from a Long Panel." *International Economic Review* 59 (3): 1571–1619. ISSN: 1468-2354, accessed May 22, 2024. https://doi.org/10.1111/iere.12314.

De Nardi, Mariacristina, Svetlana Pashchenko, and Ponpoje Porapakkarm. 2023. "The Lifetime Costs of Bad Health." *The Review of Economic Studies, accepted.*

Deaton, Angus S., and Christina Paxson. 2004. "Mortality, Income, and Income Inequality over Time in Britain and the United States." *NBER Chapters,* 247–286. Accessed December 18, 2023.

Dharmayat, Kanika, Maria Woringer, Nikolaos Mastellos, Della Cole, Josip Car, Sumantra Ray, Kamlesh Khunti, Azeem Majeed, Kausik K Ray, and Sreenivasa Rao Kondapally Seshasai. 2020. "Investigation of Cardiovascular Health and Risk Factors Among the Diverse and Contemporary Population in London (the TOGETHER Study): Protocol for Linking Longitudinal Medical Records." *JMIR Res Protoc* 9, no. 10 (August): e17548. ISSN: 1929-0748. https://doi.org/10.2196/17548. http://www.ncbi.nlm.nih.gov/pubmed/33006568.

Dobkin, Carlos, Amy Finkelstein, Raymond Kluender, and Matthew J. Notowidigdo. 2018. "The Economic Consequences of Hospital Admissions." *American Economic Review* 108, no. 2 (February): 308–352. ISSN: 0002-8282, accessed December 18, 2023. https://doi.org/10.1257/aer.20161038.

Eurostat. 2023. *EU Statistics on Income and Living Conditions Microdata, Release 2 in 2023, Data 2004-2022 Version 1.* Accessed April 24, 2024. https://doi.org/10.2907/EUSILC2004-2022V1.

Executive Office of the President. 2025. *Executive Order 14212: Establishing the President's Make America Healthy Again Commission.* Federal Register, 90 FR 9833–9836, February 19, 2025. Accessed July 2, 2026. https://www.federalregister.gov/documents/2025/02/19/2025-02871/establishing-the-presidents-make-america-healthy-again-commission.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2021. "Place-Based Drivers of Mortality: Evidence from Migration." *American Economic Review* 111, no. 8 (August): 2697–2735. ISSN: 0002-8282, accessed December 18, 2023. https://doi.org/10.1257/aer.20190825.

Finkelstein, Amy, and Robin McKnight. 2008. "What Did Medicare Do? The Initial Impact of Medicare on Mortality and out of Pocket Medical Spending." *Journal of Public Economics* 92, no. 7 (July): 1644–1668. ISSN: 0047-2727, accessed December 18, 2023. https://doi.org/10.1016/j.jpubeco.2007.10.005.

Galama, Titus J, and Hans van Kippersluis. 2019. "A Theory of Socio-economic Disparities in Health over the Life Cycle." *The Economic Journal* 129, no. 617 (January): 338–374. ISSN: 0013-0133, accessed May 9, 2024. https://doi.org/10.1111/ecoj.12577.

Godøy, Anna, and Ingrid Huitfeldt. 2020. "Regional variation in health care utilization and mortality." *Journal of Health Economics* 71:102254.

Grossman, Michael. 1972. "On the Concept of Health Capital and the Demand for Health." *Journal of Political Economy* 80 (2): 223–255. ISSN: 0022-3808, accessed December 18, 2023. JSTOR: 1830580.

Harteloh, Peter, Kim de Bruin, and Jan Kardaun. 2010. "The Reliability of Cause-of-Death Coding in The Netherlands." *European Journal of Epidemiology* 25, no. 8 (August): 531–538. ISSN: 1573-7284, accessed May 21, 2024. https://doi.org/10.1007/s10654-010-9445-5.

Hosseini, Roozbeh, Karen Kopecky, and Kai Zhao. 2025. "How Important Is Health Inequality for Lifetime Earnings Inequality?" Rdaf030, *The Review of Economic Studies,* https://doi.org/10.1093/restud/rdaf030. https://doi.org/10.1093/restud/rdaf030.

Hosseini, Roozbeh, Karen A. Kopecky, and Kai Zhao. 2022. "The Evolution of Health over the Life Cycle." *Review of Economic Dynamics* 45 (July): 237–263. ISSN: 1094-2025, accessed December 18, 2023. https://doi.org/10.1016/j.red.2021.07.001.

Huber, Carola A., Thomas D. Szucs, Roland Rapold, and Oliver Reich. 2013. "Identifying Patients with Chronic Conditions Using Pharmacy Data in Switzerland: An Updated Mapping Approach to the Classification of Medications." *BMC Public Health* 13, no. 1 (October): 1030. ISSN: 1471-2458, accessed December 18, 2023. https://doi.org/10.1186/1471-2458-13-1030.

Karahan, Fatih, Serdar Ozkan, and Jae Song. 2019. *Anatomy of Lifetime Earnings Inequality: Heterogeneity in Job Ladder Risk vs. Human Capital.* Staff Reports 908. Federal Reserve Bank of New York.

Kennedy Moulton, Kate, Sarah Miller, Petra Persson, Maya Rossin Slater, Laura Wherry, and Gloria Aldana. 2022. *Maternal and Infant Health Inequality: New Evidence from Linked Administrative Data.* Working Paper 30693. NBER, November. Accessed December 18, 2023.

Khanolkar, Amal R, Nishi Chaturvedi, Valerie Kuan, Daniel Davis, Alun Hughes, Marcus Richards, David Bann, and Praveetha Patalay. 2021. "Socioeconomic inequalities in prevalence and development of multimorbidity across adulthood: a longitudinal analysis of the MRC 1946 national survey of health and development in the UK." *PLoS medicine* 18 (9): e1003775.

Kinge, Jonas Minet, Jørgen Heibø Modalsli, Simon Øverland, Håkon Kristian Gjessing, Mette Christophersen Tollånes, Ann Kristin Knudsen, Vegard Skirbekk, Bjørn Heine Strand, Siri Eldevik Håberg, and Stein Emil Vollset. 2019. "Association of Household Income With Life Expectancy and Cause-Specific Mortality in Norway, 2005-2015." *JAMA* 321, no. 19 (May): 1916–1925. ISSN: 0098-7484, accessed December 18, 2023. https://doi.org/10.1001/jama.2019.4329.

Kolsrud, Jonas, Camille Landais, Daniel Reck, and Johannes Spinnewijn. 2024. "Retirement Consumption and Pension Design." *American Economic Review* 114, no. 1 (January): 89–133. ISSN: 0002-8282, accessed April 22, 2024. https://doi.org/10.1257/aer.20221426.

Lamers, Leida M., and René C. J. A. van Vliet. 2004. "The Pharmacy-based Cost Group Model: Validating and Adjusting the Classification of Medications for Chronic Conditions to the Dutch Situation." *Health Policy* 68, no. 1 (April): 113–121. ISSN: 0168-8510, accessed January 8, 2024. https://doi.org/10.1016/j.healthpol.2003.09.001.

Lleras-Muney, Adriana, Hannes Schwandt, and Laura Wherry. 2024. *Poverty and Health.* Working Paper, Working Paper Series 32866. National Bureau of Economic Research, August. https://doi.org/10.3386/w32866. http://www.nber.org/papers/w32866.

Loucks, Eric B, Kristjan T Magnusson, Stephen Cook, David H Rehkopf, Earl S Ford, and Lisa F Berkman. 2007. "Socioeconomic position and the metabolic syndrome in early, middle, and late life: evidence from NHANES 1999–2002." *Annals of epidemiology* 17 (10): 782–790.

Lu, Yuan, Xin Xin, Chungsoo Kim, Jordan Asher, Harlan Krumholz, and John Brush. 2024. "Leveraging Electronic Health Records to Assess Neighborhood Advantages and Risk of Cardiovascular Outcomes Among Hypertensive Patients." *Hypertension* 81 (Suppl_1): AMP20–AMP20.

Mackenbach, Johan P. 2019. *Health Inequalities: Persistence and Change in European Welfare States.* Oxford University Press, August. ISBN: 978-0-19-186911-2, accessed January 9, 2024. https://doi.org/10.1093/oso/9780198831419.001.0001.

Marmot, M. G., S. Stansfeld, C. Patel, F. North, J. Head, I. White, E. Brunner, A. Feeney, M. G. Marmot, and G. Davey Smith. 1991. "Health Inequalities among British Civil Servants: The Whitehall II Study." *The Lancet,* Originally Published as Volume 1, Issue 8754, 337, no. 8754 (June): 1387–1393. ISSN: 0140-6736, accessed December 18, 2023. https://doi.org/10.1016/0140-6736(91)93068-K.

Marmot, Michael. 2015. *The Health Gap: The Challenge of an Unequal World.* 1st edition. New York, NY London Oxford New Delhi Sydney: Bloomsbury Press, November. ISBN: 978-1-63286-078-1.

Martinson, Melissa L. 2012. "Income inequality in health at all ages: a comparison of the United States and England." *American Journal of Public Health* 102 (11): 2049–2056.

Mortensen, Laust H., Johan Rehnberg, Espen Dahl, Finn Diderichsen, Jon Ivar Elstad, Pekka Martikainen, David Rehkopf, Lasse Tarkiainen, and Johan Fritzell. 2016. "Shape of the Association between Income and Mortality: A Cohort Study of Denmark, Finland, Norway and Sweden in 1995 and 2003." *BMJ Open* 6, no. 12 (December): e010974. ISSN: 2044-6055, 2044-6055, accessed December 18, 2023. https://doi.org/10.1136/bmjopen-2015-010974.

Murray, Christopher J. L., Aleksandr Y. Aravkin, Peng Zheng, Cristiana Abbafati, Kaja M. Abbas, Mohsen Abbasi-Kangevari, Foad Abd-Allah, Ahmed Abdelalim, Mohammad Abdollahi, Ibrahim Abdollahpour, et al. 2020. "Global Burden of 87 Risk Factors in 204 Countries and Territories, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019." *The Lancet* 396, no. 10258 (October): 1223–1249. ISSN: 0140-6736, 1474-547X, accessed January 9, 2024. https://doi.org/10.1016/S0140-6736(20)30752-2.

National Center for Health Statistics. 2022. *Health, United States, Annual Perspective, 2020-2021.* Technical report. National Center for Health Statistics (U.S.) Accessed April 24, 2024. https://doi.org/10.15620/cdc:122044.

Nesson, Erik, and Joshua J Robinson. 2017. "The measurement of health and the connection between health inequality and income." *Available at SSRN 2994011.*

O'Donnell, Owen, Eddy Van Doorslaer, and Tom Van Ourti. 2015. "Chapter 17 - Health and Inequality." In *Handbook of Income Distribution*, edited by Anthony B. Atkinson and François Bourguignon, 2:1419–1533. Handbook of Income Distribution. Elsevier, January. Accessed May 21, 2024. https://doi.org/10.1016/B978-0-444-59429-7.00018-2.

Obozinski, Guillaume, Martin J. Wainwright, and Michael I. Jordan. 2011. "Support Union Recovery in High-Dimensional Multivariate Regression." *The Annals of Statistics* 39 (1): 1–47. ISSN: 0090-5364, accessed April 22, 2024. JSTOR: 29783630.

Pais, Jeremy. 2014. "Cumulative structural disadvantage and racial health disparities: The pathways of childhood socioeconomic influence." *Demography* 51:1729–1753.

Pampel, Fred C., Patrick M. Krueger, and Justin T. Denney. 2010. "Socioeconomic Disparities in Health Behaviors." *Annual Review of Sociology* 36 (1): 349–370. Accessed December 18, 2023. https://doi.org/10.1146/annurev.soc.012809.102529.

Poulsen, Melissa N, Annemarie G Hirsch, Lorraine Dean, Jonathan Pollak, Joseph DeWalle, Katherine Moon, Meghann Reeder, Karen Bandeen-Roche, and Brian S Schwartz. 2024. "Community credit scores and community socioeconomic deprivation in association with type 2 diabetes across an urban to rural spectrum in Pennsylvania: a case–control study." *BMJ Public Health* 2 (1).

Russo, Nicolo, Rory McGee, Mariacristina De Nardi, Margherita Borella, and Ross Abram. 2024. *Health Inequality and Economic Disparities by Race, Ethnicity, and Gender.* Working Paper, Working Paper Series 32971. National Bureau of Economic Research, September. https://doi.org/10.3386/w32971. http://www.nber.org/papers/w32971.

Sapolsky, Robert M. 2005. "The Influence of Social Hierarchy on Primate Health." *Science* 308, no. 5722 (April): 648–652. Accessed December 18, 2023. https://doi.org/10.1126/science.1106477.

Seeman, Teresa, Sharon S Merkin, Eileen Crimmins, Brandon Koretz, Susan Charette, and Arun Karlamangla. 2008. "Education, income and ethnic differences in cumulative biological risk profiles in a national sample of US adults: NHANES III (1988–1994)." *Social science & medicine* 66 (1): 72–87.

Shui, Ailun, Gerard J van den Berg, Jochen O Mierau, and Laura Viluma. 2025. "Lifetime Trajectories and Drivers of Socioeconomic Health Disparities: Evidence from Longitudinal Biomarkers in the Netherlands."

Singh, Gopal K, Gem P Daus, Michelle Allender, Christine T Ramey, Elijah K Martin, Chrisp Perry, Andrew A De Los Reyes, and Ivy P Vedamuthu. 2017. "Social determinants of health in the United States: addressing major health inequality trends for the nation, 1935-2016." *International Journal of MCH and AIDS* 6 (2): 139.

Smith, James P. 2004. "Unraveling the SES: health connection." *Population and development review* 30:108–132.

Statistics Netherlands. 2024. *Health Expenditure; Providers and Financing, 1998–2022.* StatLine dataset 85546ENG, December 10, 2024. Accessed July 30, 2026. https://www.cbs.nl/en-gb/figures/detail/85546ENG.

Stepner, Michael. 2019. *The Insurance Value of Redistributive Taxes and Transfers.* Working Paper. Accessed April 22, 2024.

Sullivan, Daniel, and Till von Wachter. 2009. "Job Displacement and Mortality: An Analysis Using Administrative Data*." *The Quarterly Journal of Economics* 124, no. 3 (August): 1265–1306. ISSN: 0033-5533, accessed December 18, 2023. https://doi.org/10.1162/qjec.2009.124.3.1265.

van den Berg, Gerard J., Maarten Lindeboom, and France Portrait. 2006. "Economic Conditions Early in Life and Individual Mortality." *American Economic Review* 96, no. 1 (March): 290–302. ISSN: 0002-8282, accessed April 22, 2024. https://doi.org/10.1257/000282806776157740.

van Ooijen, Raun, Rob J. M. Alessie, and Marike Knoef. 2015. *Health Status Over the Life Cycle.* SSRN Scholarly Paper, 2743110, Rochester, NY, October. Accessed January 8, 2024. https://doi.org/10.2139/ssrn.2743110.

Watson, Kathleen B., Jennifer L. Wiltz, Kunthea Nhim, Rachel B. Kaufmann, Craig W. Thomas, and Kurt J. Greenlund. 2025. "Trends in Multiple Chronic Conditions Among US Adults, By Life Stage, Behavioral Risk Factor Surveillance System, 2013–2023." *Preventing Chronic Disease* 22:240539. https://doi.org/10.5888/pcd22.240539. https://www.cdc.gov/pcd/issues/2025/24_0539.htm.

World Health Organization. 1985. "Targets for Health for All: Targets in Support of the European Regional Strategy for Health for All," accessed January 9, 2024.

———. 2008. *Closing the Gap in a Generation: Health Equity through Action on the Social Determinants of Health.* Technical report. Geneva: World Health Organization, August. Accessed May 9, 2024.

———. 2017. *National Health Inequality Monitoring: A Step-by-Step Manual.* World Health Organization. ISBN: 978-92-4-151218-3, accessed May 21, 2024.

Yuan, Ming, and Yi Lin. 2006. "Model Selection and Estimation in Regression with Grouped Variables." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 68, no. 1 (February): 49–67. ISSN: 1369-7412, accessed April 22, 2024. https://doi.org/10.1111/j.1467-9868.2005.00532.x.

Zaninotto, Paola, George David Batty, Sari Stenholm, Ichiro Kawachi, Martin Hyde, Marcel Goldberg, Hugo Westerlund, Jussi Vahtera, and Jenny Head. 2020. "Socioeconomic inequalities in disability-free life expectancy in older people from England and the United States: a cross-national population-based study." *The journals of gerontology: Series A* 75 (5): 906–913.

Zhu, Yinjie, Louise H Dekker, and Jochen O Mierau. 2023. "Socio-economic gradients in diagnosed and undiagnosed type 2 diabetes and its related health complications." *Nutrition, Metabolism and Cardiovascular Diseases* 33 (1): 90–94.

A Tables

Table 1: MAPPING PHARMACEUTICAL DATA TO CHRONIC DISEASE

Chronic Disease	ATC Code(s)	Medicine Description
Acid related disorders	A02	Drugs for acid related disorders
Bone diseases (osteoporosis)	M05	Drugs for treatment of bone diseases
Cancer*	L01	Antineoplastic agents
Cardiovascular diseases (inc. hypertension)	B01A, C01, C04A, C02, C07, C08, C09	Antithrombotic agents, cardiac therapy, peripheral vasodilators, antihypertensives, beta blocking agents, calcium channel blockers, agents acting on the renin-angiotensin system
Dementia	N06D	Anti-dementia drugs
Diabetes (mellitus)	A10A, A10B, A10X	Insulins and analogues, Blood glucose lowering drugs (excl. insulins), other drugs used in diabetes
Epilepsy	N03	Antiepileptics
Glaucoma	S01E	Antiglaucoma preparations and miotics
Gout (Hyperuricemia)	M04	Antigout preparations
HIV	J05A	Direct acting antivirals
Hyperlipidemia	C10	Lipid modifying agents
Intestinal (inflammatory) diseases	A07E	Intestinal antiinflammatory agents
(Iron deficiency) anemia	B03A	Iron preparations
Migraines	N02C	Antimigraine preparations
Pain	N02A, N02B	Opioids, other analgesics and antipyretics
Parkinson's disease	N04, N05B, N05C	Anxiolytics, hypnotics and sedatives
Psychological disorders	N06A	Antidepressants
Psychoses	N05A	Antipsychotics
Respiratory illnesses	R03	Drugs for obstructive airway diseases
Rheumatological conditions	L04A	Immunosuppressants
Thyroid disorders	H03	Thyroid therapy
Tuberculosis	J04A	Drugs for treatment of tuberculosis

Note: The table reports the mapping from three-digit ATC codes to chronic diseases. The mapping is adapted from Huber et al. (2013), with specific refinements described in Appendix D. *Cancer here refers to cancers treated with pharmacy-dispensed medications, which is around 5% of all cancer diagnoses. Digestive tract and skin cancers dominate this measure: they account for over 60% of the diagnoses.

Table 2: LIFECYCLE DECOMPOSITION OF THE CDI GAP, $\times 1000$

Age window	11-20	21-30	31-40	41-50	51-60	61-70	11-70
1. Differential Aging	0.04	1.01	1.43	2.98	3.74	4.69	13.88
2. Health-Based Sorting	0.00	1.91	1.94	2.26	3.63	0.40	10.14
3. Compositional Effects							
a. Attrition due to death	0.01	-0.03	-0.18	-0.50	-1.62	-3.50	-5.82
b. Cohort effects	-0.04	-0.17	-0.57	-1.39	-1.63	-2.32	-6.10
Total Change	0.01	2.72	2.62	3.35	4.12	-0.73	12.09

Note: The table attributes changes in the low-minus-high income CDI gap to differential aging, health-based sorting, mortality attrition, and cohort entry and exit effects, by ten-year age bin and over the full lifecycle. Values are multiplied by 1000; positive values widen the gap and negative values narrow it. Over the lifecycle, the positive aging and sorting contributions are partly offset by mortality attrition and cohort effects, explaining why their sum exceeds the observed change in the CDI gap. Appendix G provides the full decomposition; Figure 8 and Appendix Figure G.3 report the corresponding patterns across finer income groups.

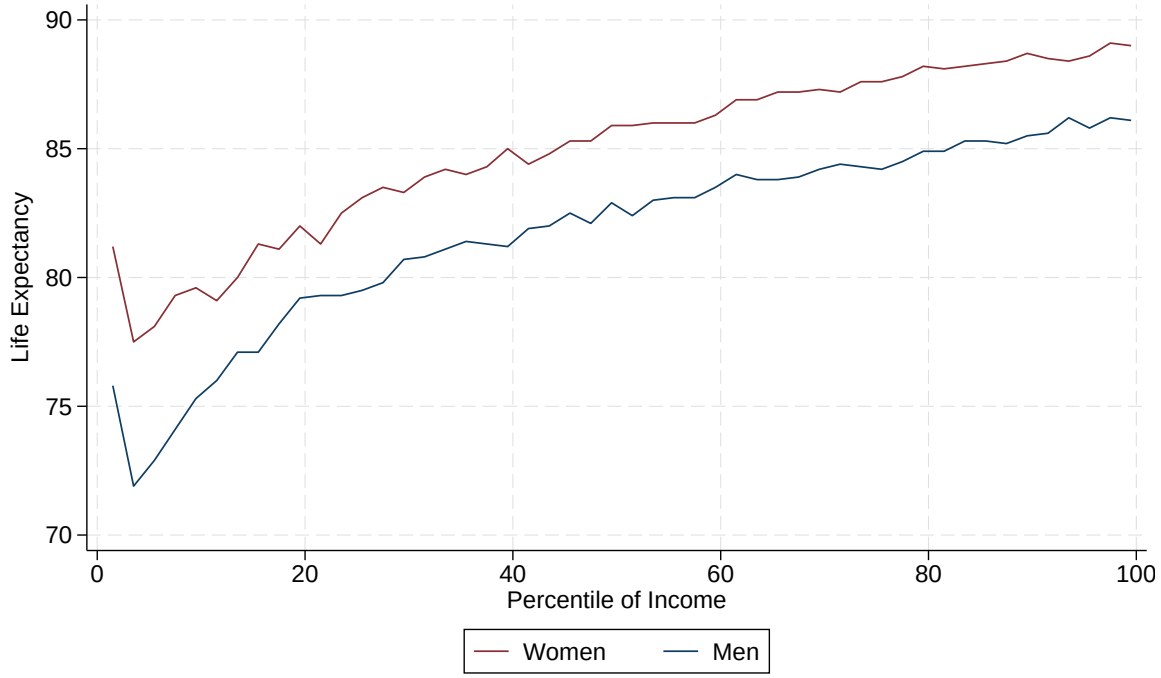
Table 3: COUNTERFACTUAL ANALYSIS

	High Income Y_H	Low Income Y_L			
	Baseline	Baseline	Y_H Ageing		
			From 60	*From 40*	*From 20*
1. Life Expectancy (at age 40)	85.3	81.2	81.6	82.3	82.5
2. Biological Age (relative to Y_H)					
a. at 70	70.0	75.4	73.7	71.2	70.3
b. at 40	40.0	49.3	49.3	49.3	43.3
3. Lifetime Healthcare Costs (€'000s)					
a. Net Effect	155.9	157.7	157.6	155.6	153.2
b. Keeping Survival Unchanged	155.9	157.7	155.5	148.9	145.6

Note: The table reports simulated life expectancy, biological ages, and lifetime cost. In the baseline calculations, we apply the high- and low-income aging and mortality-attrition effects to their respective income groups starting from age 20. In the counterfactual scenarios, the high-income aging effects are applied to the low-income baseline from different ages onwards. The reported biological-age figures exclude mortality attrition and are visualized in Appendix Figure C.9; Appendix Figure G.2 gives the corresponding attrition-adjusted profiles. Life expectancy and expected lifetime costs are calculated conditional on survival to age 40; healthcare costs are estimated directly through age 70 and extrapolated thereafter. More detail is provided in Appendix H.

B Figures

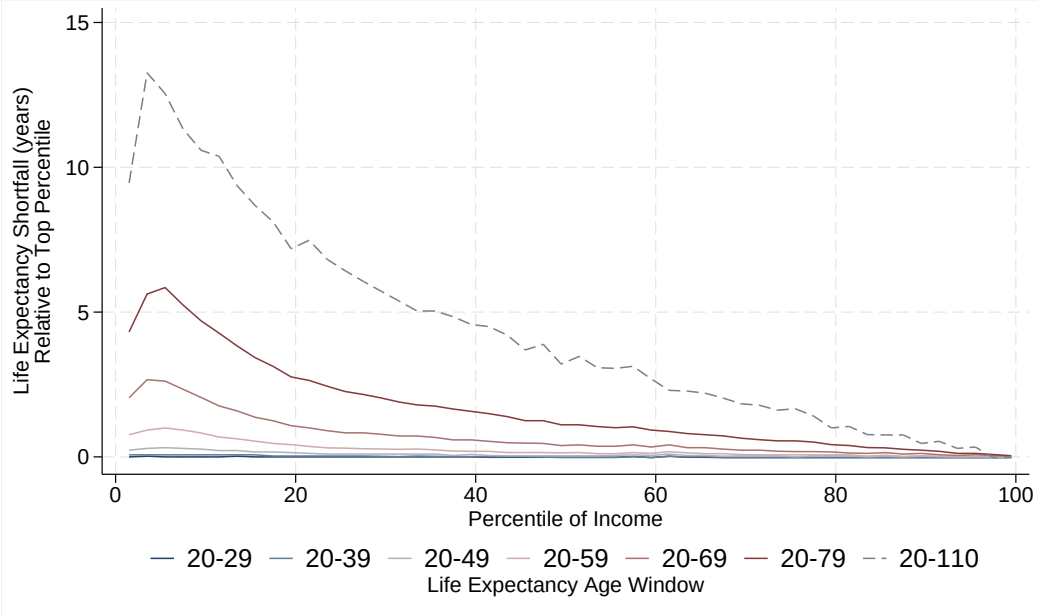
Figure 1: LIFE EXPECTANCY BY INCOME PERCENTILE



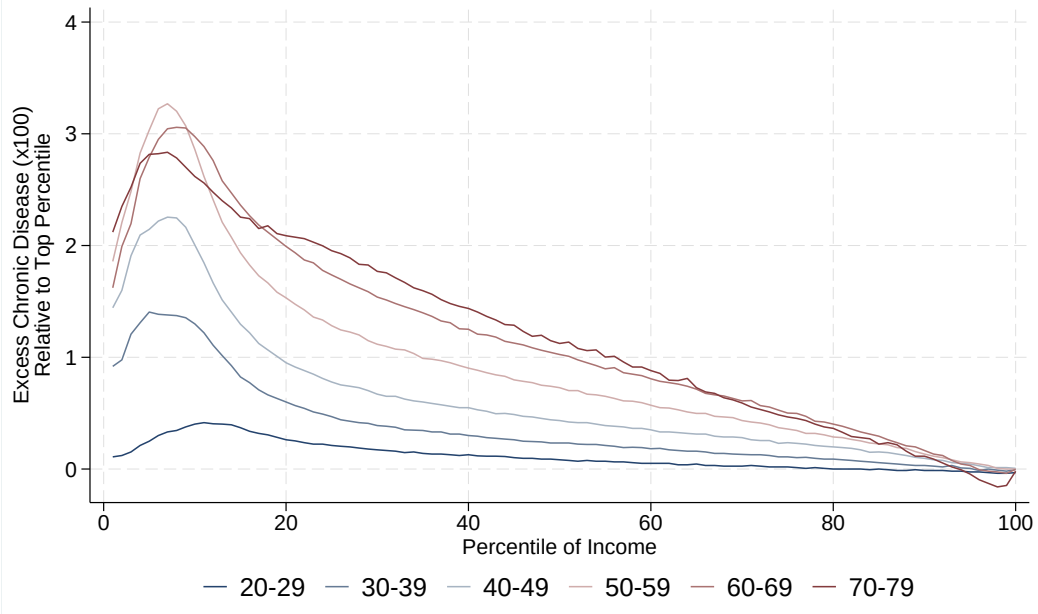
Note: This figure plots period life expectancy at age 40 by income percentile separately for women and men, following the approach in Chetty et al. (2016). For each gender and income percentile, we construct age-specific one-year mortality rates and use them as a period life table: observed mortality rates define survival probabilities at each age, older-age mortality hazards are completed using a Gompertz extrapolation where direct estimates become sparse, and the resulting survival curve is extended to age 110. Life expectancy is then computed by aggregating the implied age-at-death distribution, conditional on survival to age 40. For exposition, the income distribution is represented using 50 quantiles, plotted at each bin's midpoint.

Figure 2: MORTALITY VERSUS HEALTH GAPS OVER THE LIFECYCLE

A. Life Expectancy Shortfall



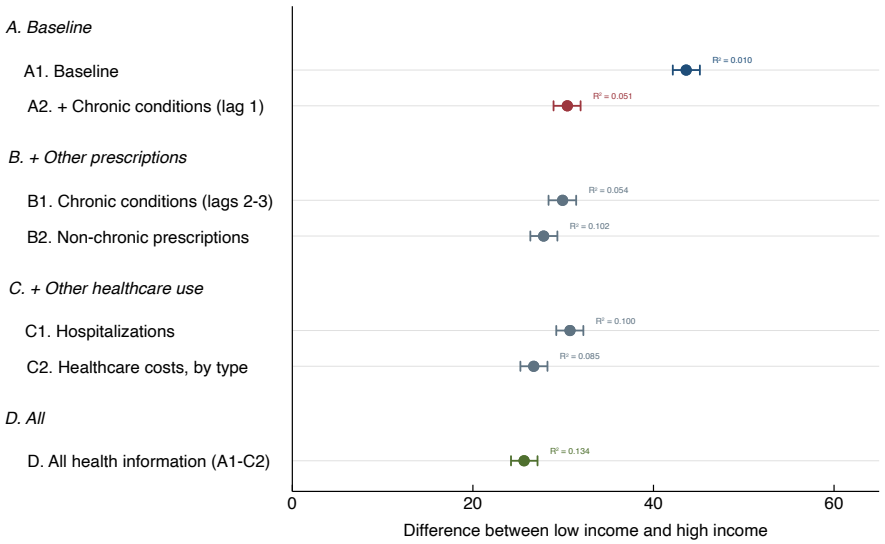
B. Excess CDI



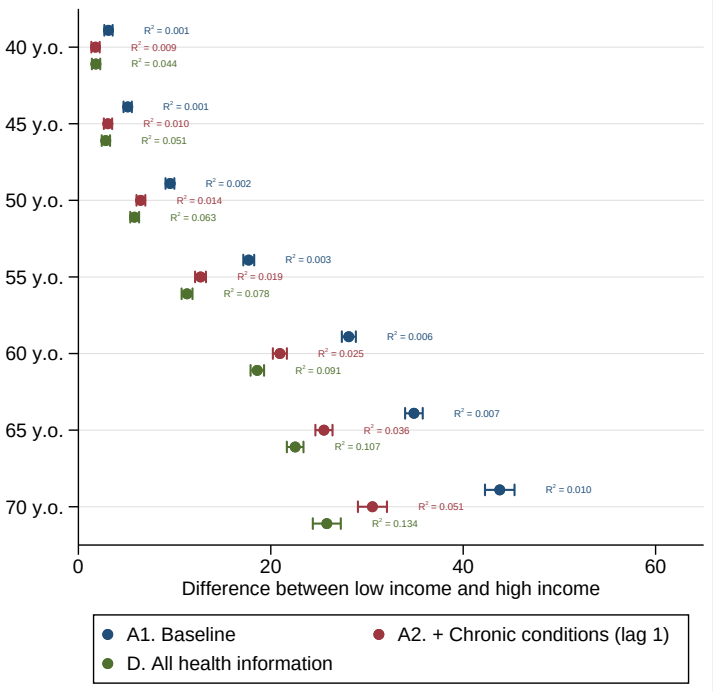
Note: Panel A shows the life expectancy shortfall relative to the top income percentile for seven cumulative age windows: ages 20–29, 20–39, 20–49, 20–59, 20–69, 20–79, and 20–110. The life expectancy contribution of window $[a_0, a_1]$ is computed as $\sum_{a=a_0}^{a_1} \Pr(A = a) \cdot a$, where A denotes age at death and the age-at-death probabilities are obtained from the income-percentile-specific survival curve used in Figure 1. The life expectancy windows are cumulative because mortality risk is a flow. For exposition, the income distribution in panel A is represented using 50 quantiles, plotted at each bin's midpoint. By contrast, Panel B reports decade-specific CDI gaps since the CDI is a stock measure of accumulated chronic disease burden observed at a given age. Panel B plots excess chronic disease burden, measured by the Chronic Disease Index (CDI), relative to the top income percentile for each age window. CDI gaps are pooled across gender and expressed relative to the top income percentile.

Figure 3: CHRONIC CONDITIONS AND THE MORTALITY GAP

A. Gap in Five-year Mortality Rate (per 1000) at 70 Years of Age



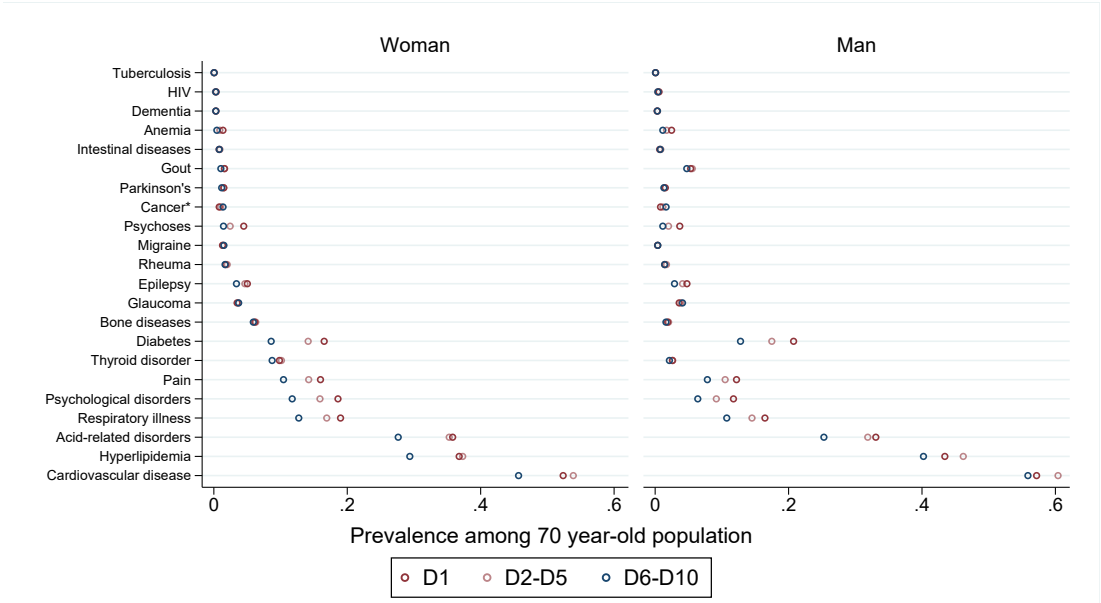
B. Gap in Five-year Mortality Rate (per 1000) at Various Ages



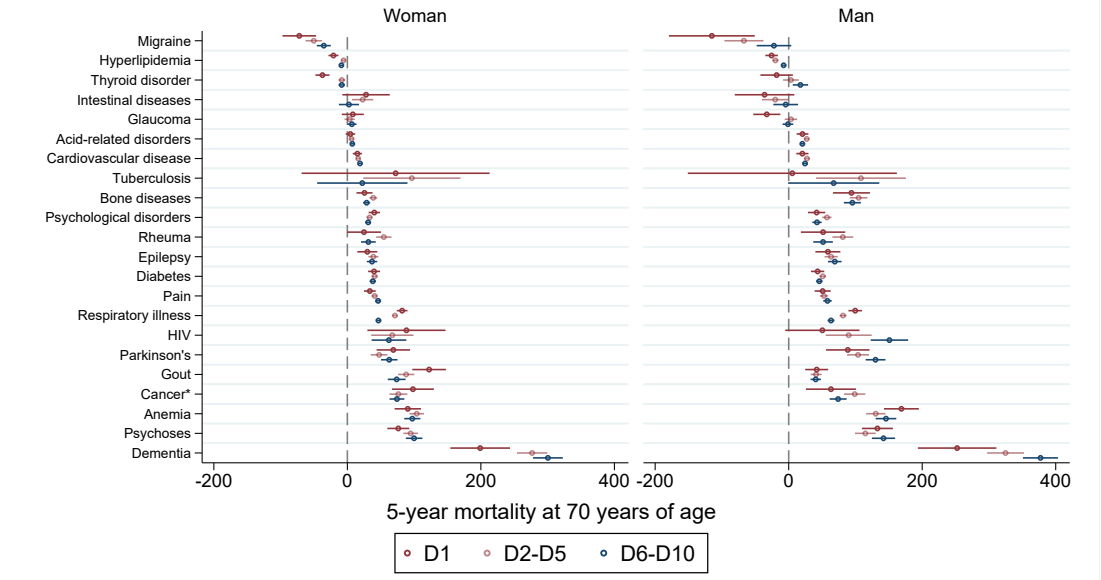
Note: In Panel A, each row corresponds to a different regression of five-year mortality (per 1,000) on income (defined as low- vs. high-income) and a series of controls identified by the row label on the left. For each specification, the plot shows the estimated coefficient on income. Specifications reported after A2 include all the chronic condition controls used in A2, as well as those listed in the left column. Row D includes all health-related variables jointly. Panel B reports the mortality gap estimates from specifications A1, A2, and D at different ages. We pool the observations for years between 2013 and 2016. For more information, refer to Appendix Section E.

Figure 4: PREVALENCE AND TREATMENT EFFECTS OF CHRONIC CONDITIONS

A. Prevalence of Chronic Conditions at Age 70

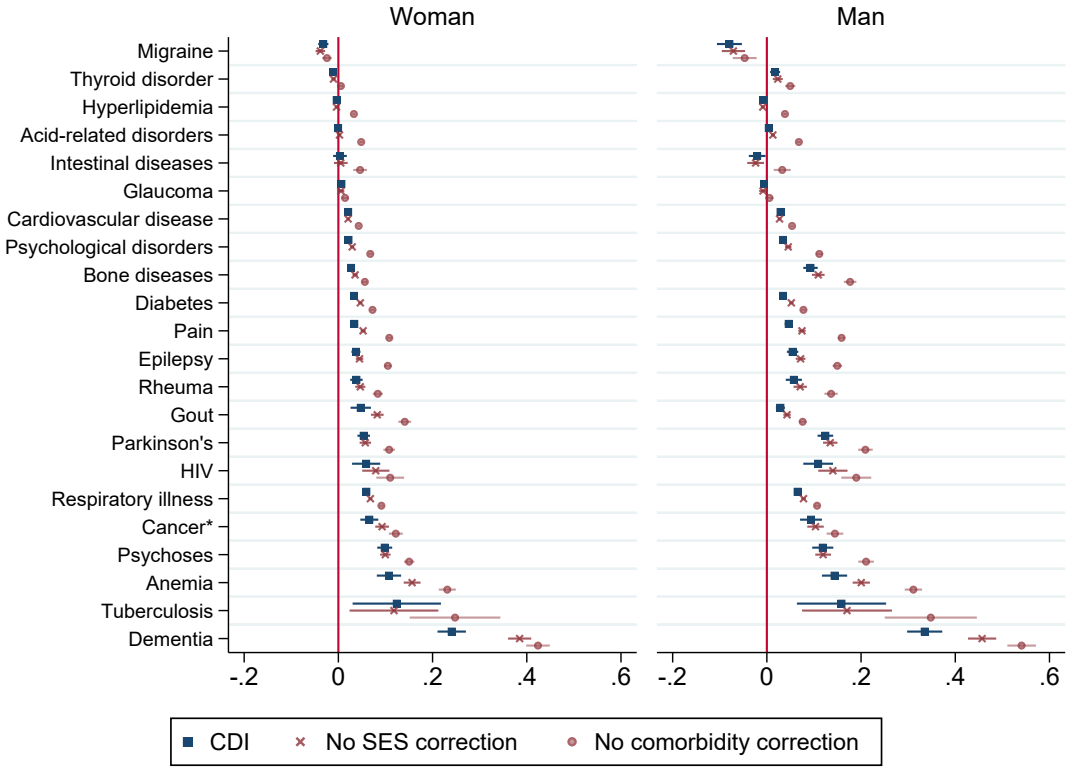


B. Effect of Chronic Conditions on Five-Year Mortality at Age 70 (per 1000)



Note: Panel A reports chronic-condition prevalence at age 70 by gender and income group: D1, D2-D5, and D6-D10. Panel B reports coefficient estimates from regressions of five-year mortality on all chronic conditions, separately by income group and gender. *Cancer refers to cancers treated with pharmacy-dispensed medications, around 5% of all cancer diagnoses.

Figure 5: CHRONIC-CONDITION CDI WEIGHTS AND REGRESSION COEFFICIENTS



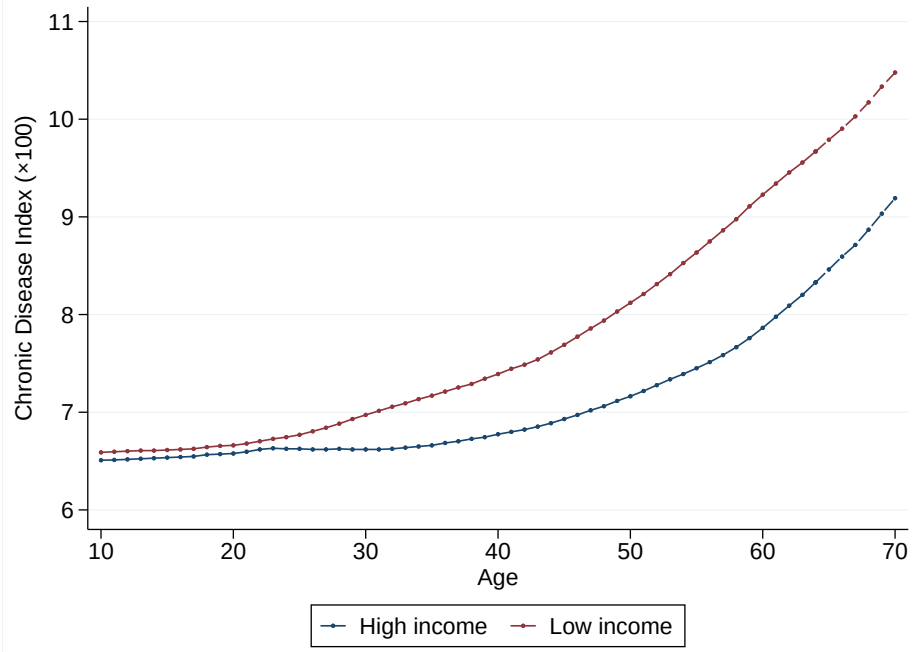
Note: This figure presents condition-specific CDI weights, by gender. For each condition j, the weight is defined as:

$\hat{\beta}^j_{70} = E[CDI_{70} \mid c^j_{69} = c^j_{68} = c^j_{67} = 1, c^{-j} = \overline{c^{-j}}] - E[CDI_{70} \mid c^j_{69} = c^j_{68} = c^j_{67} = 0, c^{-j} = \overline{c^{-j}}] \quad \forall j = 1, ..., 22.$

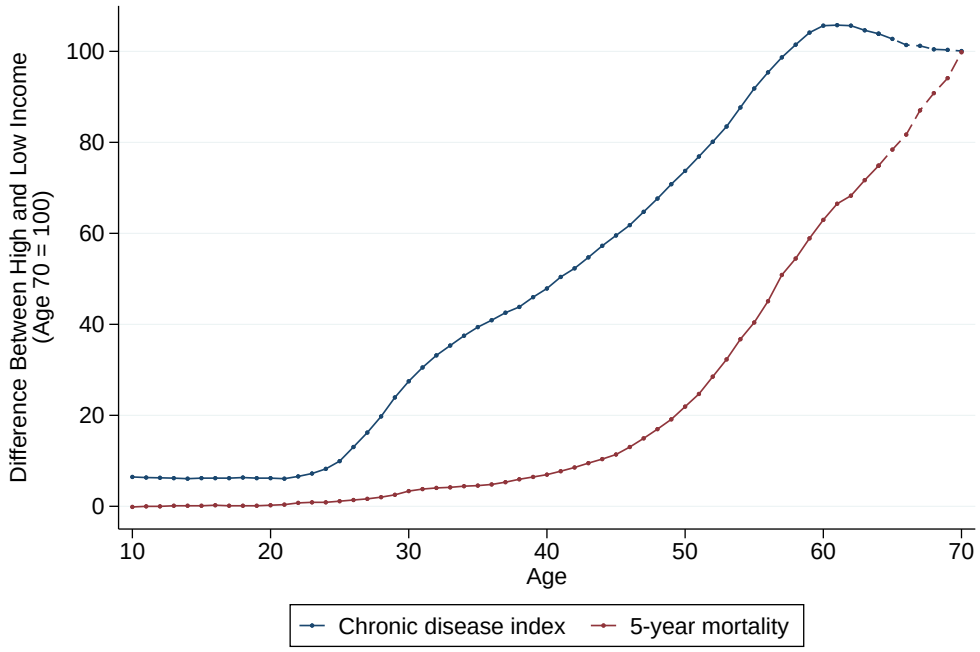
The figure also presents coefficients from multivariate and univariate regressions of predicted CDI on lagged chronic conditions. Multivariate regressions estimate coefficients for all chronic conditions simultaneously. For comparability with the CDI-weight definition, the displayed coefficient for a given chronic condition is the sum of the coefficients for its three lags. *Cancer here refers to cancers treated with pharmacy-dispensed medications, which is around 5% of all cancer diagnoses. Digestive tract and skin cancers dominate this measure: they account for over 60% of the diagnoses.

Figure 6: HEALTH GAP OVER THE LIFECYCLE

A. Observed CDI, by Income Group



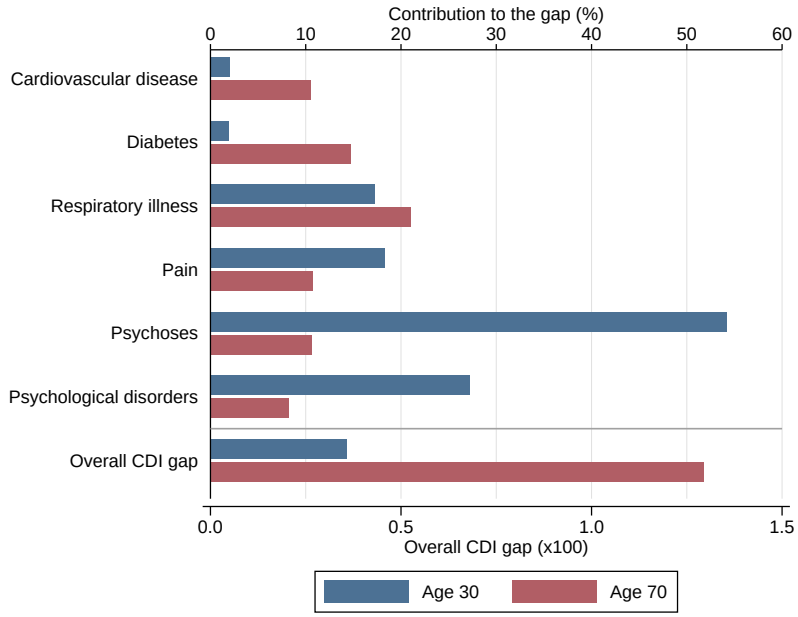
B. CDI Gap versus Mortality Gap



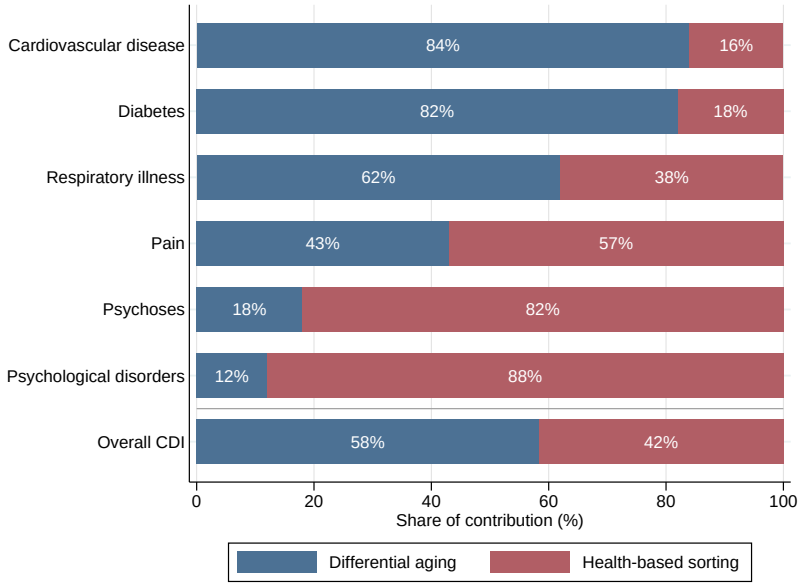
Note: Panel A plots the average chronic disease index by income group over the lifecycle. That is, at each age, the average CDI is shown for the relevant income group. Individuals are ranked on the mean of $(Y_{t-4}, Y_{t-3}, Y_{t-2})$ within year, age, and gender. High income is defined as above median income, and low income as below median. From age 65 onwards, we fix income as the mean of (Y_{60}, Y_{61}, Y_{62}); this is represented by the dashed lines in Panel A. Panel B shows the difference in CDI between income groups, along with the difference in observed five-year mortality. Both gaps in Panel B are shown relative to age 70, which is set to 100. We pool all observations between 2009 and 2021 in both panels.

Figure 7: CONDITION-SPECIFIC CONTRIBUTIONS TO THE CDI GAP

A. Contribution to the CDI gap at ages 30 and 70

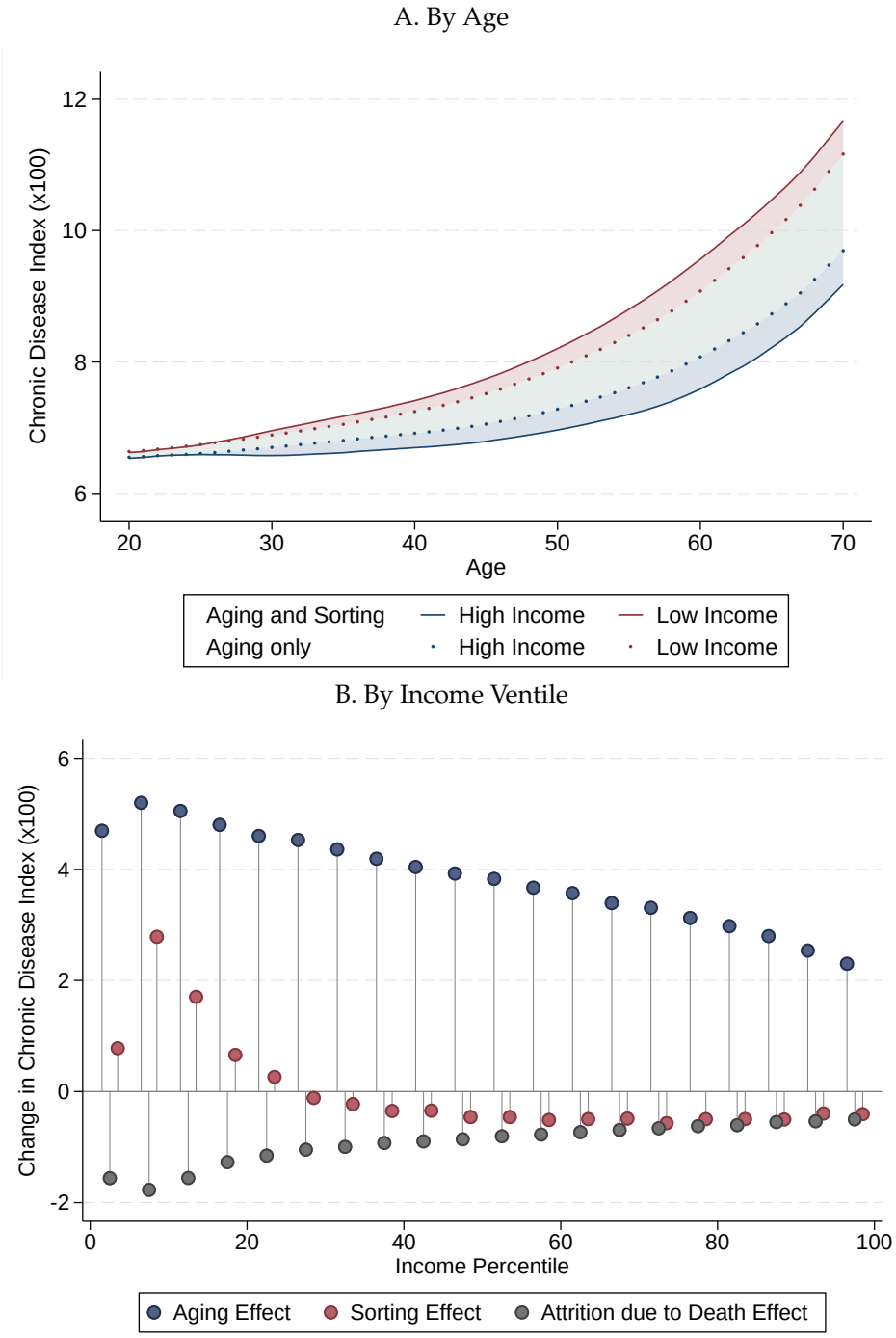


B. Aging versus sorting over lifecycle



Note: Panel A compares each selected condition's contribution to the CDI gap at ages 30 and 70, computed as $\kappa_a^j = (CC_{L,a}^j - CC_{H,a}^j) \cdot \hat{\beta}_{70}^j$, where $CC_{L,a}^j$ and $CC_{H,a}^j$ are the shares of low- and high-income individuals with condition j, and $\hat{\beta}_{70}^j$ is the condition's CDI weight defined in Figure 5. The final row reports the overall CDI gap at each age, using the separate CDI-gap (x100) scale at the bottom. Panel B decomposes the selected-condition contribution into the share attributable to differential aging versus health-based sorting, summed over ages 10–70. Both panels use the same six conditions, selected because they make the largest contributions to the CDI gap over the lifecycle.

Figure 8: DECOMPOSING SORTING VERSUS AGING EFFECTS



Note: Panel A shows the evolution of the CDI when it is simulated using either (a) aging effects only, or (b) aging and health-based sorting effects. The simulated CDI paths start from the observed CDI at age 10 and use the components defined in equation 7 to simulate the CDI at later ages. The teal shaded area represents the health gap due to differential aging. The blue and red areas are the gaps due to positive and negative sorting for high- and low-income individuals, respectively. Panel B decomposes cumulative aging, health-based sorting, and mortality attrition effects across income ventiles, aggregated over the lifecycle. The corresponding quintile-by-age decomposition is reported in Panel B of Figure G.3. All decompositions pool observations from 2009 to 2021.

C Appendix: Additional Tables and Figures

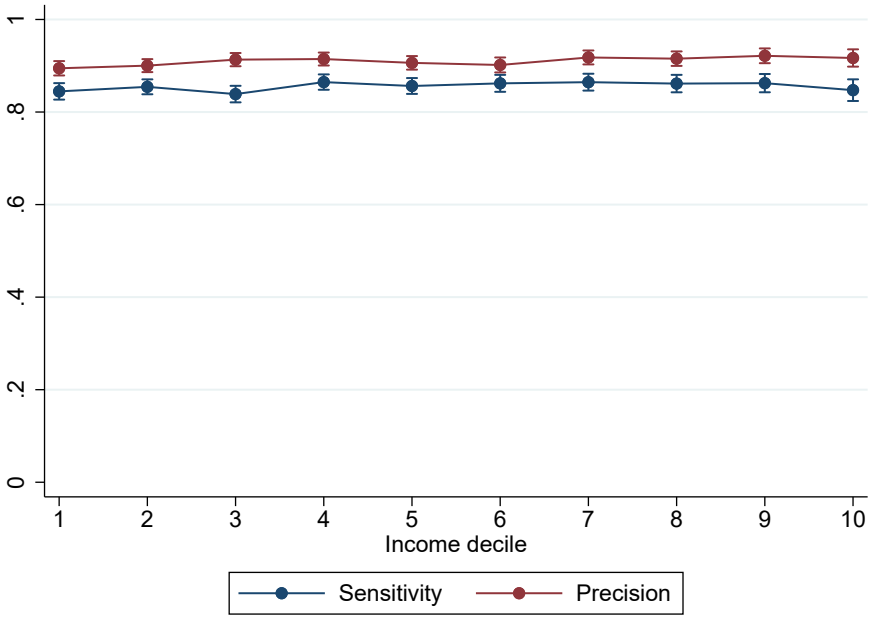
Table C.1: SAMPLE DESCRIPTIVE STATISTICS

	Full sample	2006	2016	Aged 40	Aged 70
A. Demographics					
Age	41.12	39.31	41.95	40	70
Foreign-born	0.12	0.10	0.12	0.19	0.09
Male	0.50	0.49	0.50	0.50	0.49
Self-employed	0.06	0.05	0.06	0.11	0.02
With partner	0.72	0.74	0.72	0.77	0.73
With kids	0.55	0.56	0.54	0.72	0.07
B. Education					
Less than High School	0.21	0.21	0.22	0.04	0.05
High School	0.26	0.19	0.29	0.29	0.11
College	0.08	0.05	0.09	0.16	0.03
Further Studies	0.04	0.03	0.05	0.10	0.01
Education missing	0.40	0.51	0.35	0.41	0.79
C. Income and Wealth					
Household Income (€)	27,608	21,999	29,736	27,397	27,088
Household Wealth (€)	181,684	-	168,090	100,844	296,894
D. Health and Healthcare					
Chronic cond. count	0.85	0.77	0.87	0.50	2.03
Has chronic conditions	0.39	0.38	0.39	0.31	0.76
Cardiovascular disease	0.18	0.15	0.19	0.05	0.53
Diabetes	0.05	0.04	0.05	0.01	0.14
Respiratory illness	0.09	0.08	0.09	0.07	0.14
Pain	0.06	0.05	0.07	0.05	0.11
Psychological disorders	0.08	0.13	0.08	0.09	0.13
Psychoses	0.02	0.01	0.02	0.02	0.02
Medicines taken	2.67	2.56	2.70	1.97	5.11
Takes medicines	0.67	0.68	0.67	0.65	0.88
Total healthcare spending	2,261	-	2,387	1,539	4,140
Hospitalised	0.11	-	0.11	0.08	0.20
5-year mortality	50.67	48.51	54.41	6.03	112.60
Observations	272,889,744	16,499,473	17,187,337	3,650,513	2,653,596
Individuals	21,159,899	16,499,473	17,187,337	3,650,513	2,653,596

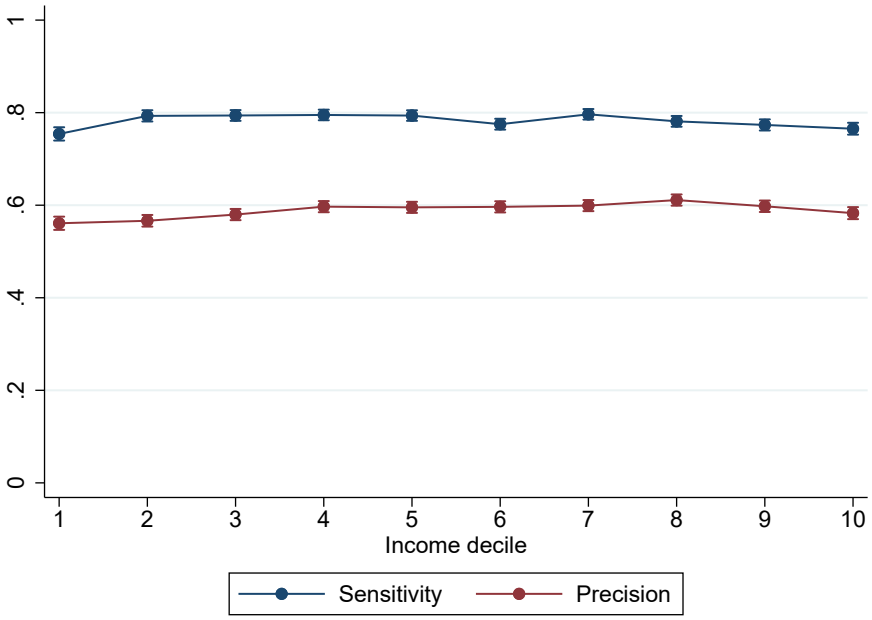
Note: This table provides descriptive statistics for our analysis sample, at selected ages and in selected calendar years.

Figure C.1: COMPARING PHARMACY DATA TO SELF-REPORTED SURVEY RESPONSES BY INCOME

A. Diabetes

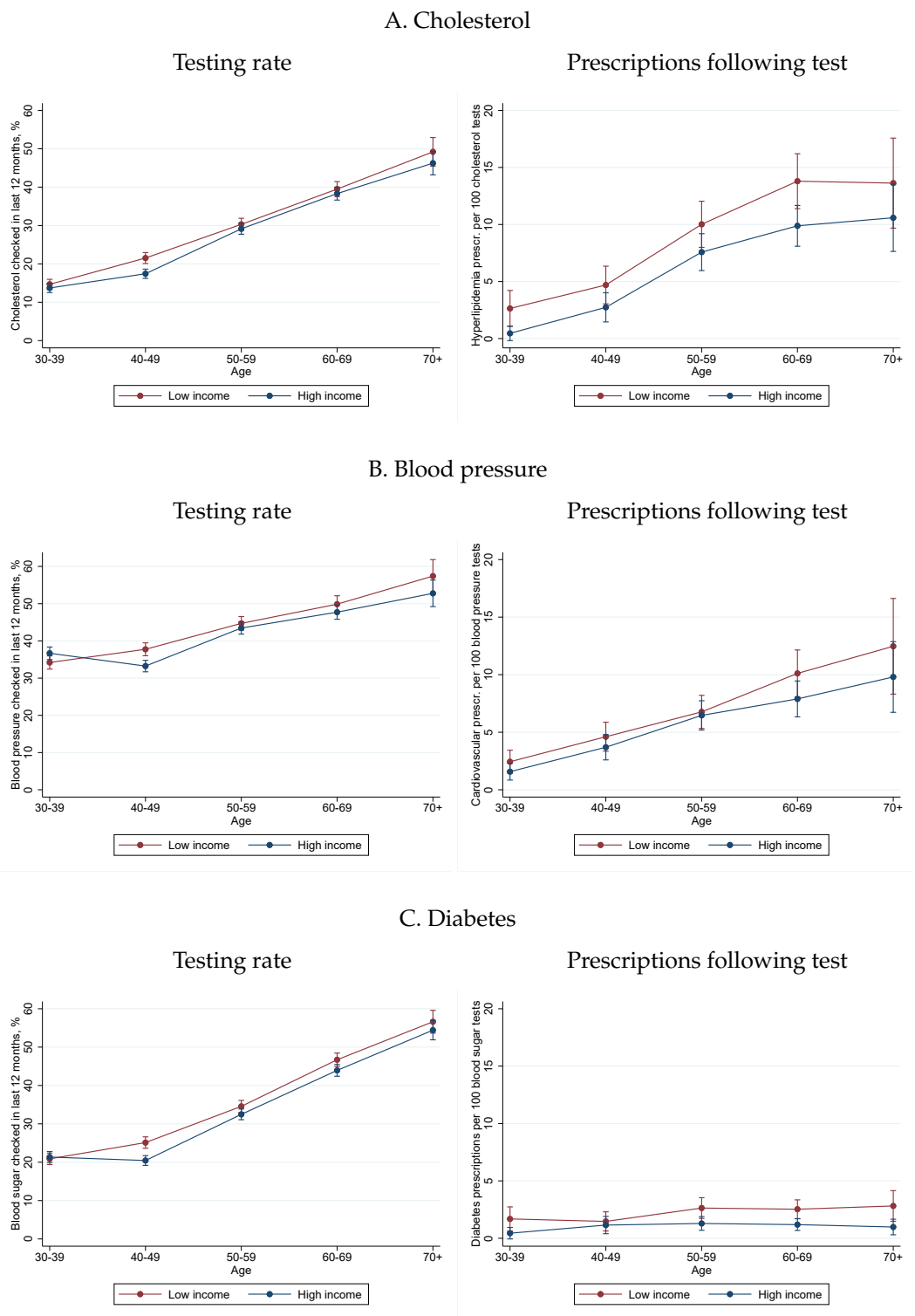


B. Cardiovascular Disease (incl. Hypertension)



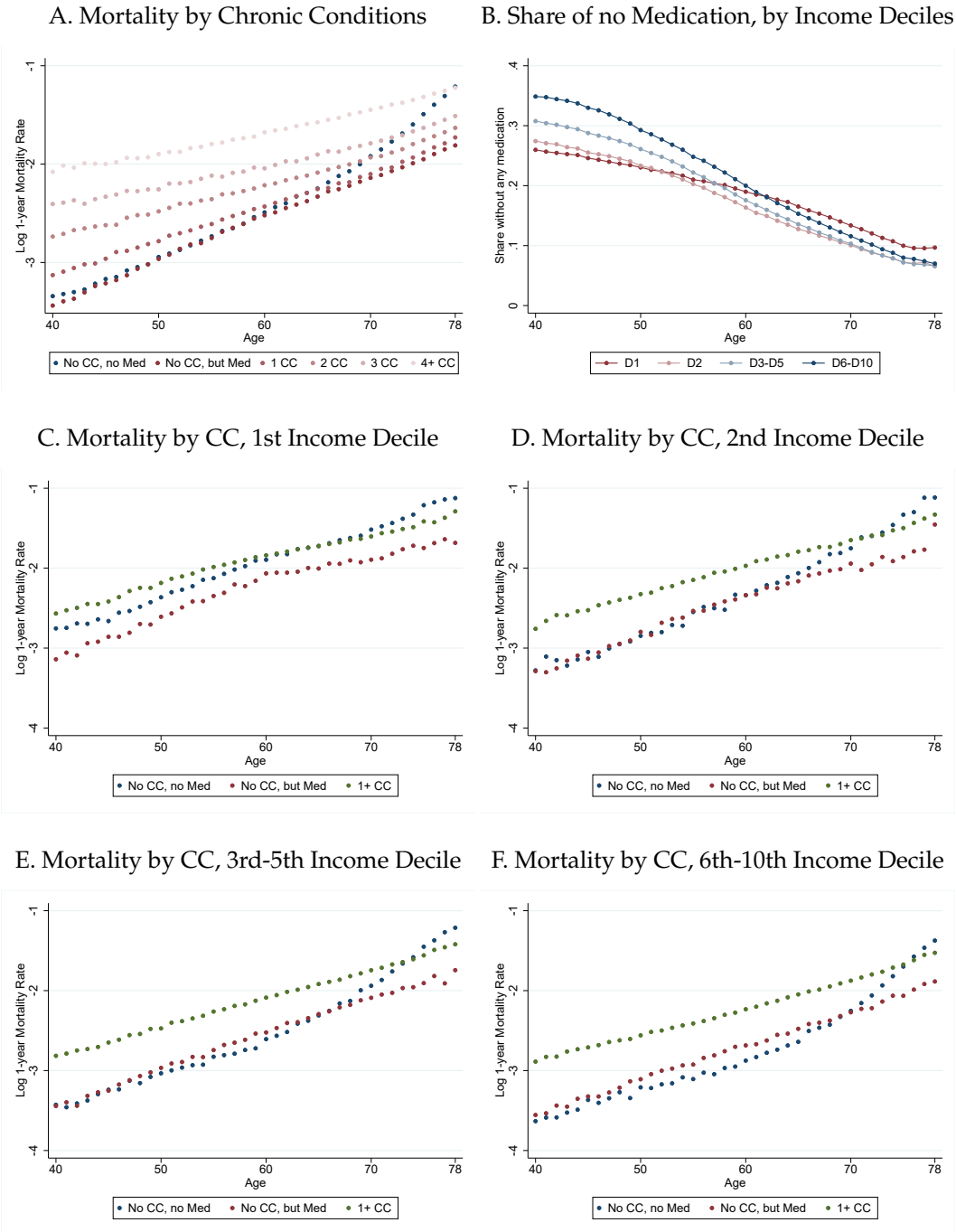
Note: This figure compares the detection of diabetes and cardiovascular disease using our ATC-to-chronic-condition mapping with self-reported conditions in the *Gezondheidsmonitor* survey data, by income decile. Sensitivity is defined as Pr(condition detected and reported | condition reported in survey). Precision is defined as Pr(condition detected and reported | condition detected in pharmacy data).

Figure C.2: TESTING AND PRESCRIPTIONS FOR SPECIFIC CONDITIONS



Note: This figure compares the rates of screening and subsequent prescription across incomes in the *Gezondheidsenquete* survey data. In the left-hand panels, we focus on the subset of people who are not currently prescribed the relevant medication, and report responses to the question "*Have you had a* [Cholesterol/Blood pressure/Blood sugar] *test in the past 12 months?*". The right-hand panels then plot the share of those who report being tested in the past 12 months that were newly prescribed with the relevant medication in the following year.

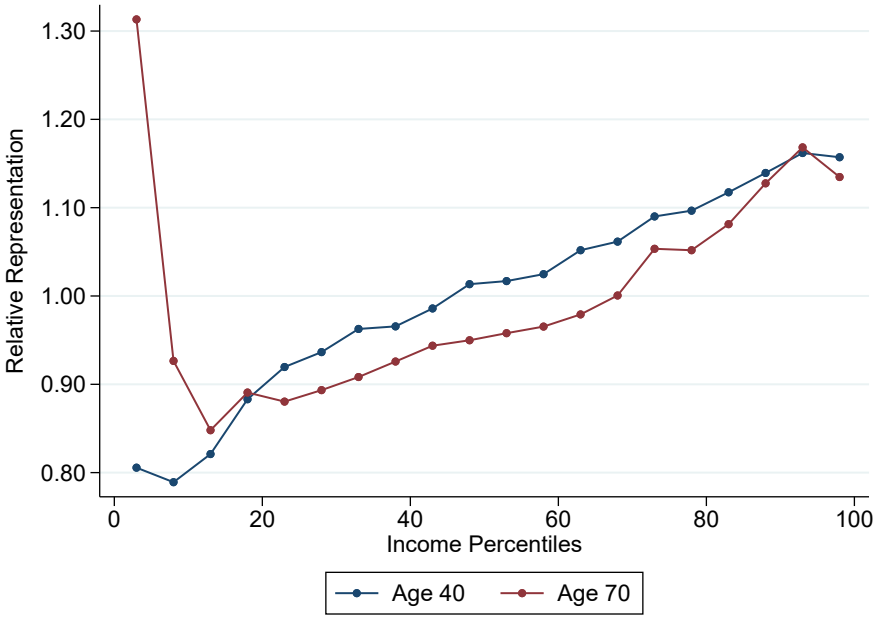
Figure C.3: ASSESSING UNDER-DIAGNOSIS BY INCOME DECILES



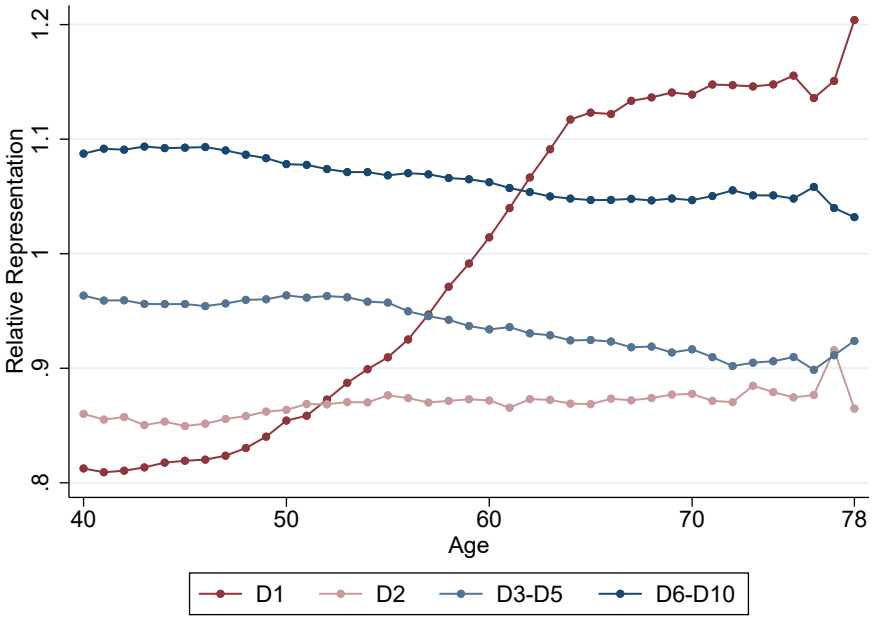
Note: Panel A reports one-year mortality by number of chronic conditions on a logarithmic scale. The category without chronic conditions is divided into individuals taking some prescription medication for non-chronic illnesses and individuals taking no prescription medication at all. Panel A pools all observations for which prescription data are available, covering 2006 to 2021. Panel B plots the share of individuals who do not take any prescription medication by income group. The income groups are decile 1, decile 2, deciles 3-5, and deciles 6-10, with income deciles defined within birth cohort and gender. Panels C-F show one-year mortality rates for individuals with different medication status within each of these income groups. Panels B-F pool all individuals for which our income measure is defined, covering 2007 to 2021.

Figure C.4: UNDER-DIAGNOSIS IN FIRST INCOME DECILE

A. Relative Representation in No Medication Group

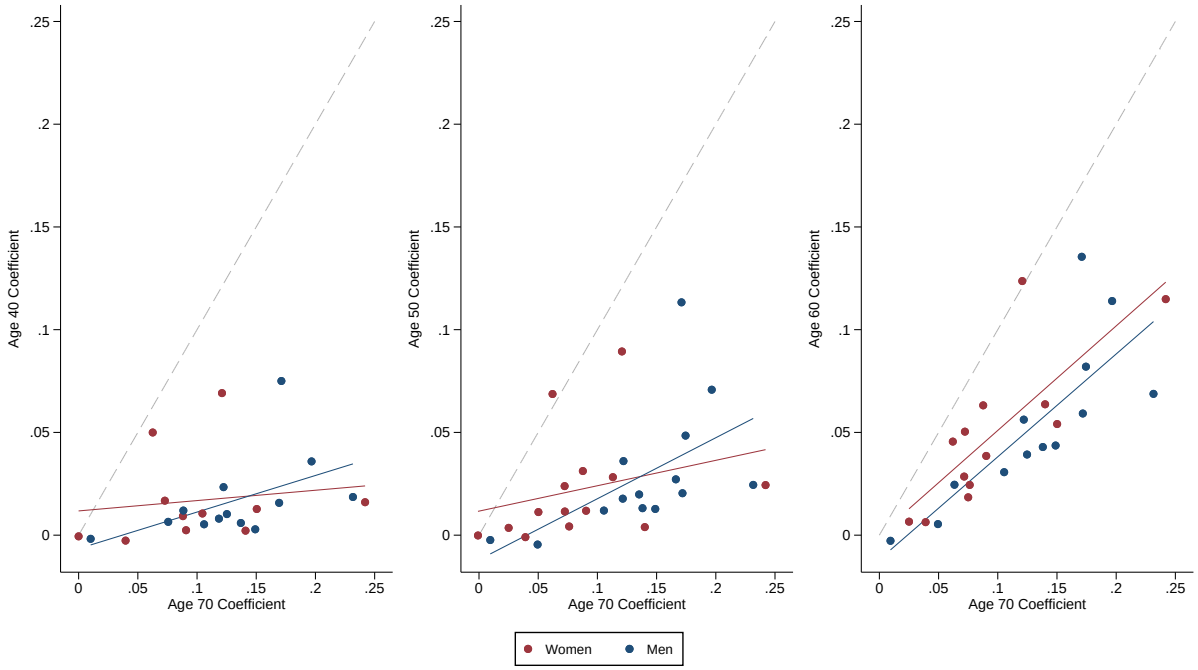


B. Representation in No Medication Sample



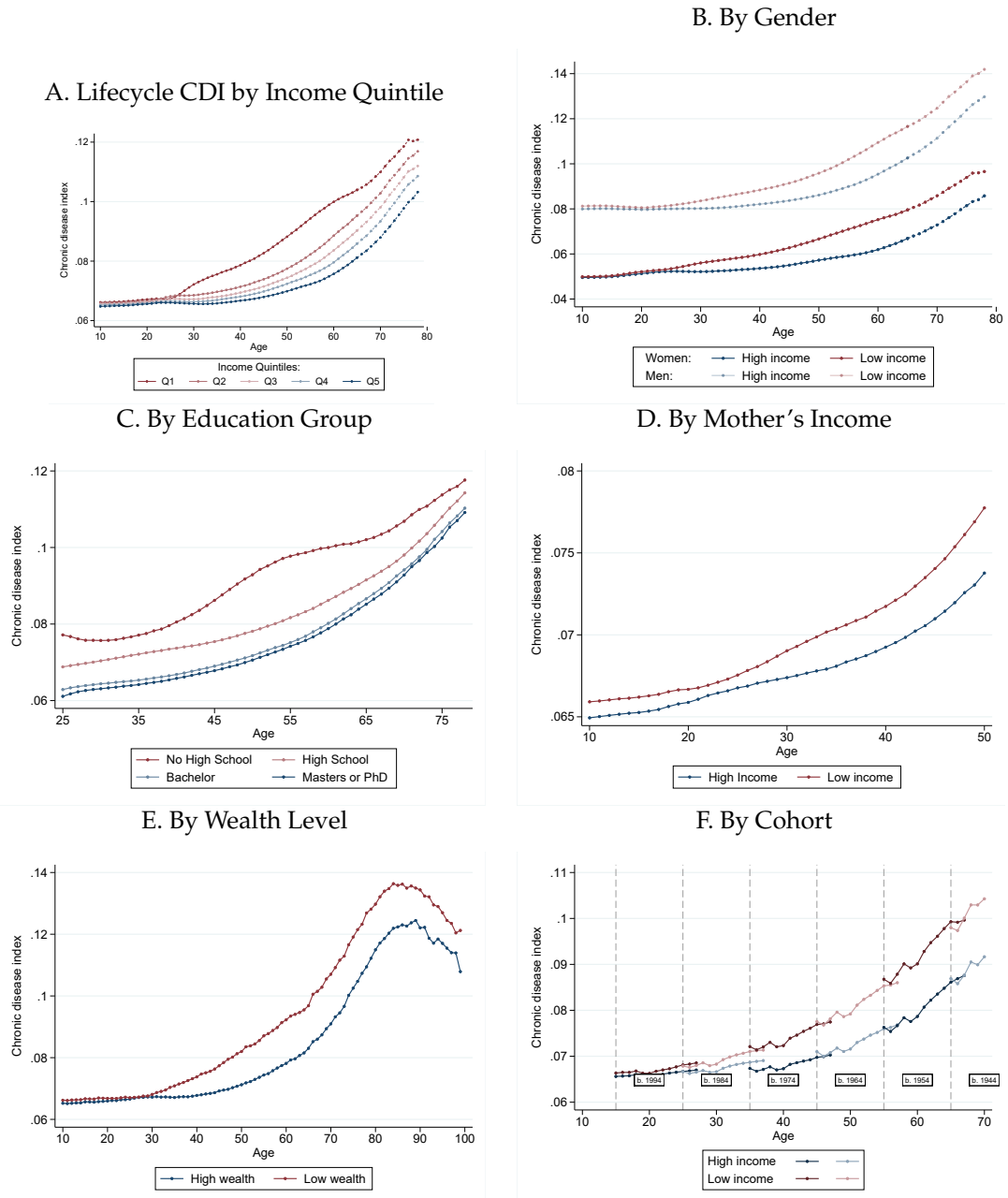
Note: This figure presents evidence for under-diagnosis among very low incomes. Panel A reports relative representation in the sample of people without any prescription medication by income ventile at age 40 and age 70. Relative representation is defined as the share of people within the no prescription medication sample who also belong to a certain income group, relative to the share of this income group in the full sample. Panel B shows the relative representation in the sample of people not taking any prescription medication at ages 40 to 78. The income groups considered consist of individuals situated in decile 1, 2, 3-5 and 6-10, respectively.

Figure C.5: AGE-SPECIFIC MORTALITY COEFFICIENTS RELATIVE TO AGE-70 COEFFICIENTS



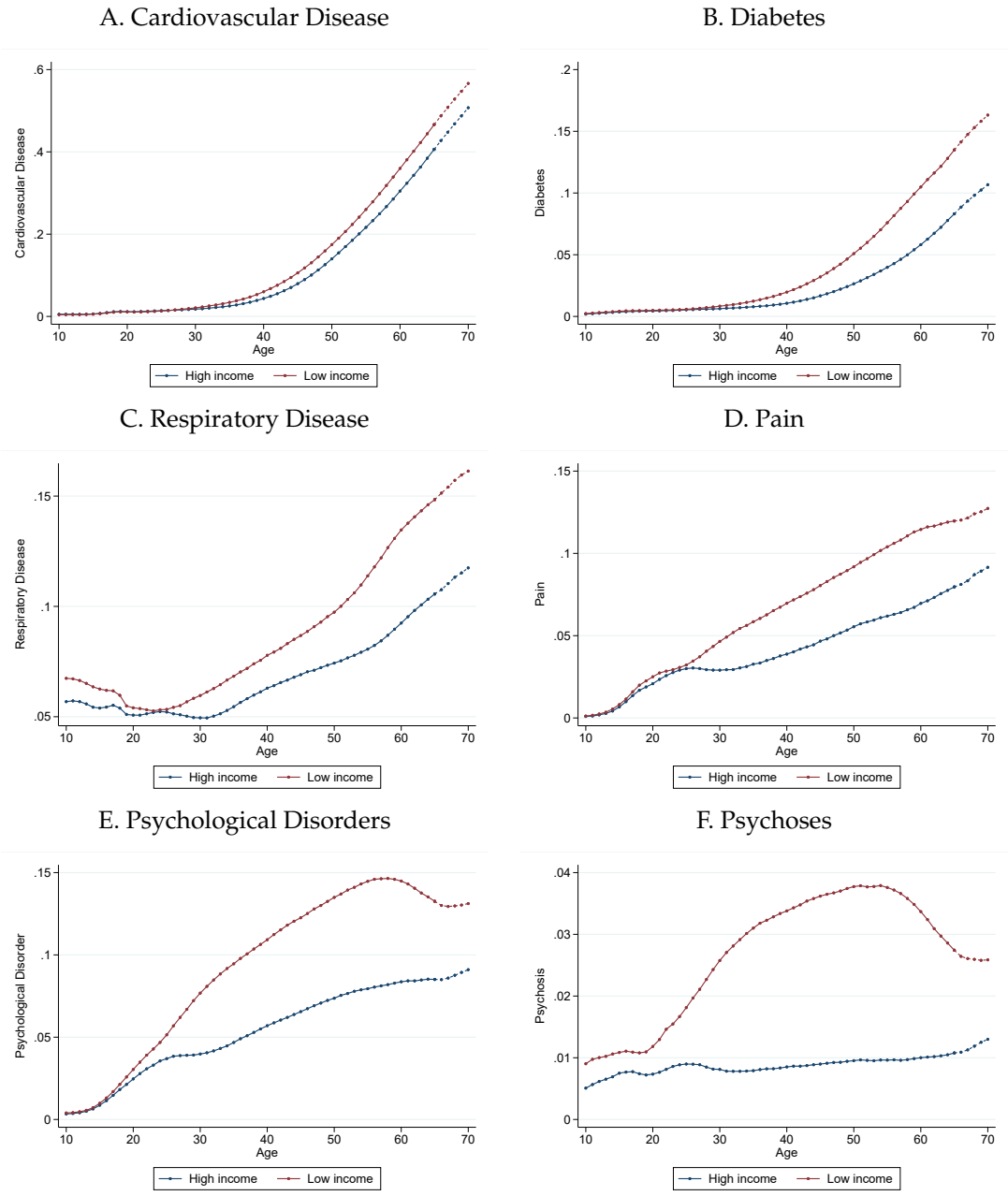
Note: The figure compares condition-specific five-year mortality coefficients estimated at ages 40, 50, and 60 with the corresponding age-70 coefficients, separately by gender. For each age a and gender g, we estimate the linear probability model $M_{i,a} = \alpha_{a,g} + \sum_{j=1}^{22} \beta^j_{a,g} C^j_{i,a-1} + \lambda_{p,g} + \varepsilon_{i,a}$, where $M_{i,a}$ is an indicator for dying within five years, $C^j_{i,a-1}$ is an indicator for chronic condition j measured in the previous year, and $\lambda_{p,g}$ denotes income-percentile fixed effects. Each point corresponds to one chronic condition: the horizontal axis reports $\beta^j_{70,g}$ and the vertical axis reports $\beta^j_{a,g}$ for $a \in \{40, 50, 60\}$. Dementia is excluded as a relative outlier for exposition purposes.

Figure C.6: LIFECYCLE CDI ACROSS SUBGROUPS



Note: This figure shows the evolution of the CDI across different subgroups and socioeconomic outcomes, analogous to Figure 6. At each age, the average CDI is shown for the relevant subgroup. Panel A shows CDI by income quintile. Panel B splits by gender and income group. Panel C splits by obtained level of education, and Panel D splits by income group of the individual's mother. Panel E reports average CDI by above- and below-median household net wealth. Panels A-E pool all observations from 2009 to 2021. Panel F reports how average CDI evolves for selected birth cohorts. For each cohort, average CDI is shown for 13 consecutive years, from its value in 2009 to its value in 2021. For this cohort analysis, income groups are defined in 2009 and kept constant until 2021.

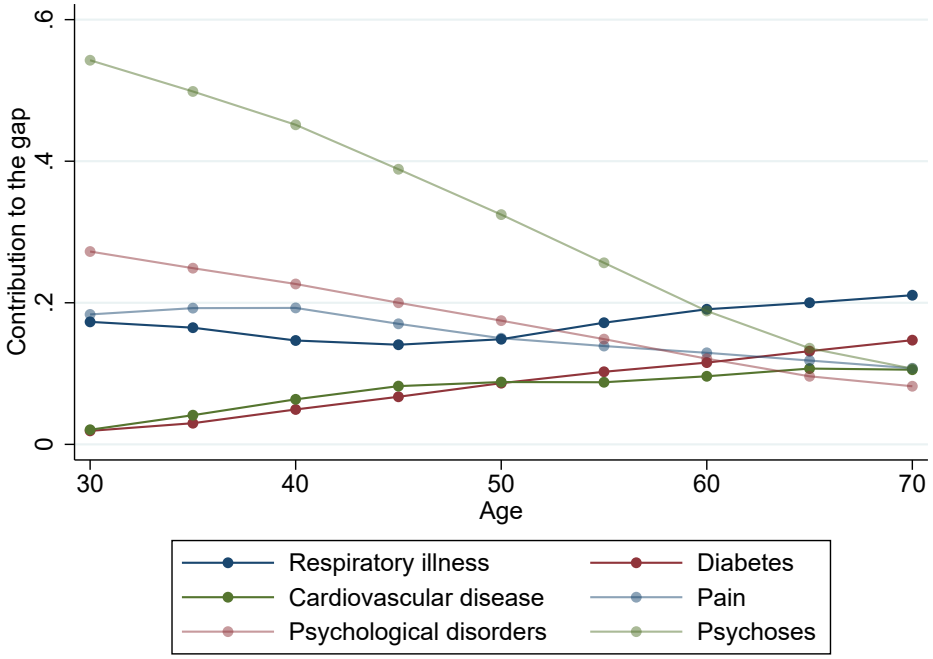
Figure C.7: LIFECYCLE PREVALENCE OF SPECIFIC CONDITIONS



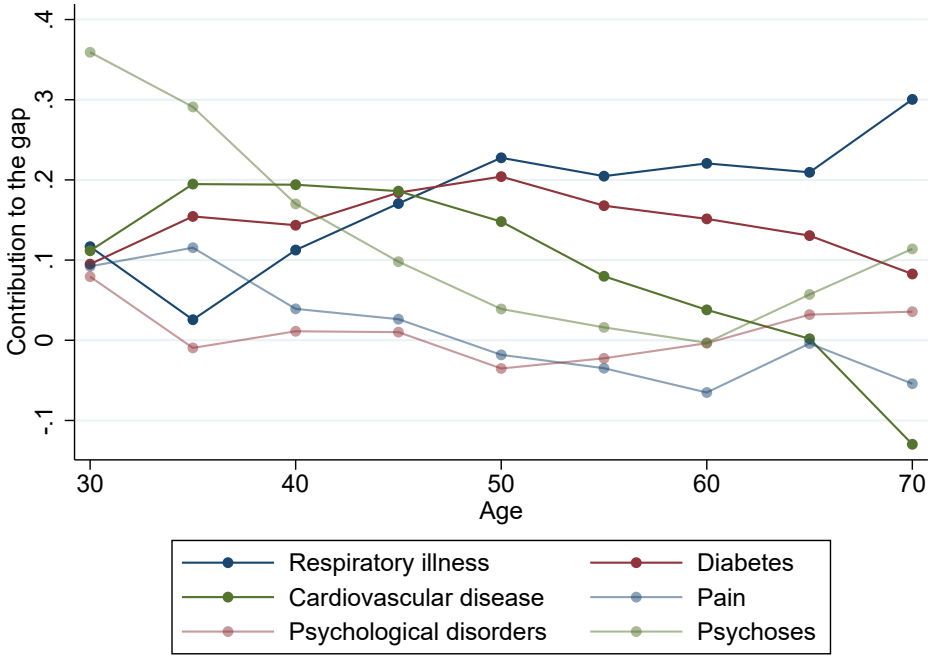
Note: All panels show the lifecycle prevalence of selected individual chronic diseases. At each age between 10 and 70, the figure reports the percentage of individuals in each income group who take medication identifying the specific chronic condition. Individuals are ranked on the mean of $(Y_{t-4}, Y_{t-3}, Y_{t-2})$ within year, age, and gender; high income is defined as above median income and low income as below median income. All panels pool observations from 2009 to 2021.

Figure C.8: CONTRIBUTION OF CHRONIC CONDITIONS TO THE CDI GAPS

A. Contribution to the gap in CDI levels

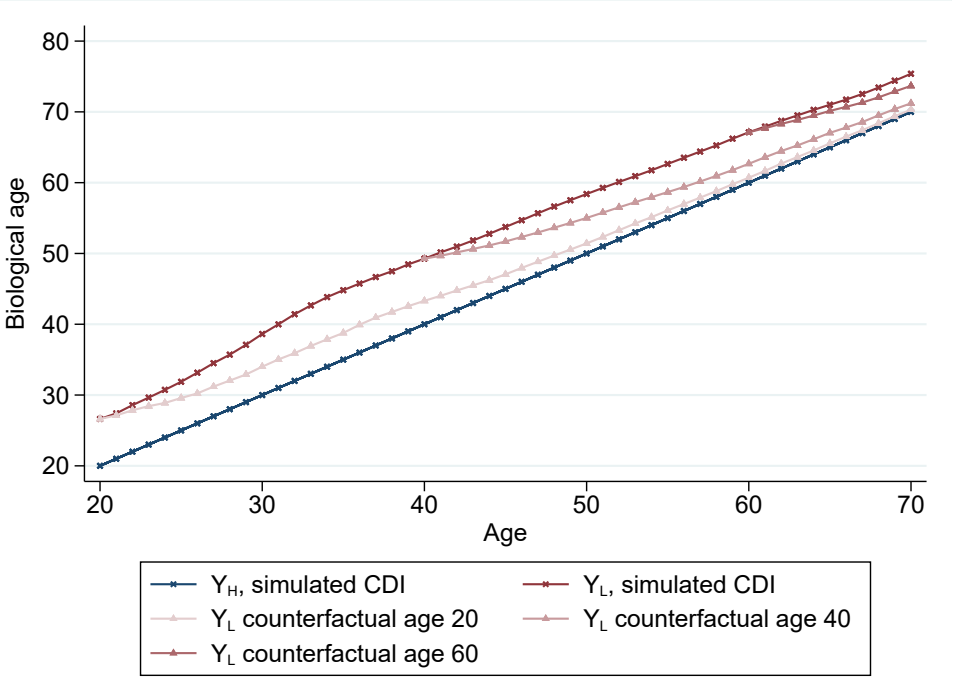


B. Contribution to the gap in five-year CDI growth



Note: Panel A shows the contribution of a selection of six chronic conditions to the chronic disease index over the lifecycle. The contribution is computed as $\kappa_a^j = (CC_{L,a}^j - CC_{H,a}^j) \cdot \hat{\beta}_{70}^j$, where $CC_{L,a}^j$ and $CC_{H,a}^j$ are the shares of low- and high-income individuals with condition j, and $\hat{\beta}_{70}^j$ is the condition's CDI weight defined in Figure 5. Panel B shows the contribution of those chronic conditions to the within-individual five-year change in the chronic disease index.

Figure C.9: BIOLOGICAL AGING



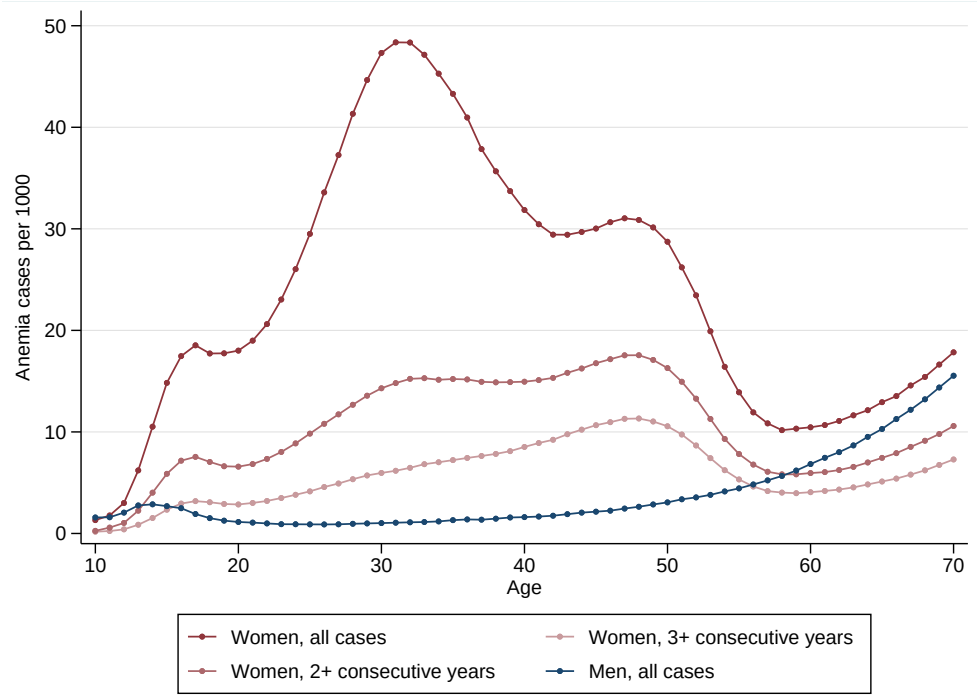
Note: The figure shows biological ages for different scenarios. These profiles intentionally exclude mortality attrition: they describe how chronic disease accumulates over the lifecycle, rather than who survives long enough to be observed. In the baseline scenario, the high- and low-income CDI paths are simulated using their respective estimated aging effects. In the counterfactual scenarios, the high-income aging effect is used to simulate the low-income CDI from different ages onwards. Table 3 reports the corresponding impacts of these counterfactuals on life expectancy and lifetime health expenditures; Appendix Figure G.2 shows the attrition-adjusted profiles.

D Refinements to the ATC-Chronic Condition mapping

We use Huber et al. (2013) as our basis to translate medication data into chronic disease indicators. We do however make a number of modifications, as described below.

- **Cardiovascular disease**: We use B01A (antithrombotic agents). To reduce false positives due to anti-blood-clot medication after an operation, we only consider that the person had cardiovascular disease if she/he took any of the medications in this group for at least two years in a row.
- **HIV**: We use J05A (direct acting antivirals). To reduce false positives, we require at least two years in a row.
- **Intestinal inflammatory diseases**: We use A07E (intestinal anti-inflammatory agents).
- **Iron deficiency anemia**: We use B03A (iron preparations). To reduce false positives due to pregnancy related anemia, we require at least three years in a row for women.
- **Rheumatic conditions**: We use L04A (immunosuppressants), but omit M01 and M02 as these are used predominantly for sport injuries at younger ages.

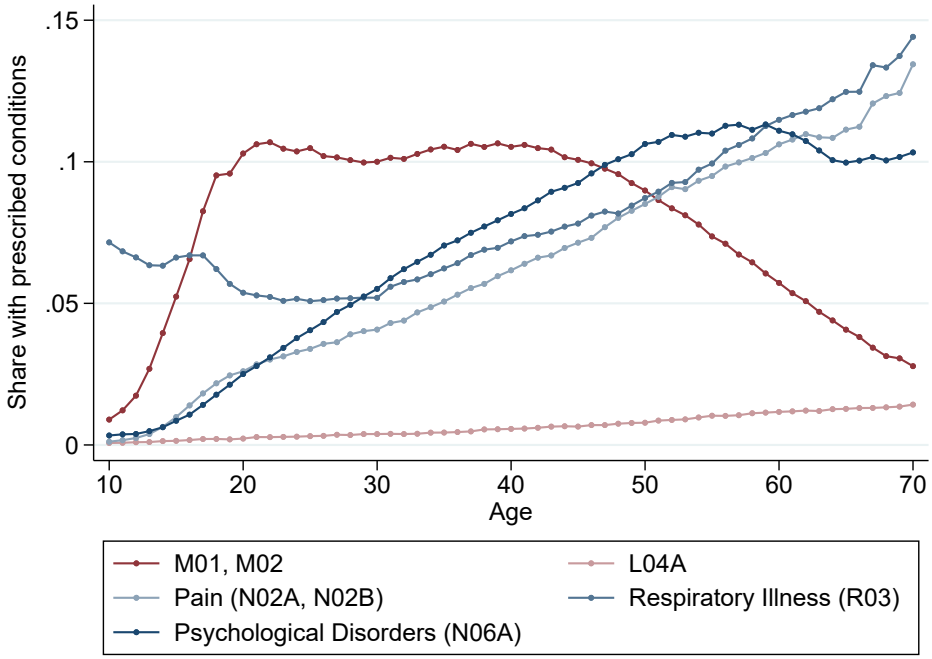
Figure D.1: ANEMIA PREVALENCE OVER THE LIFECYCLE



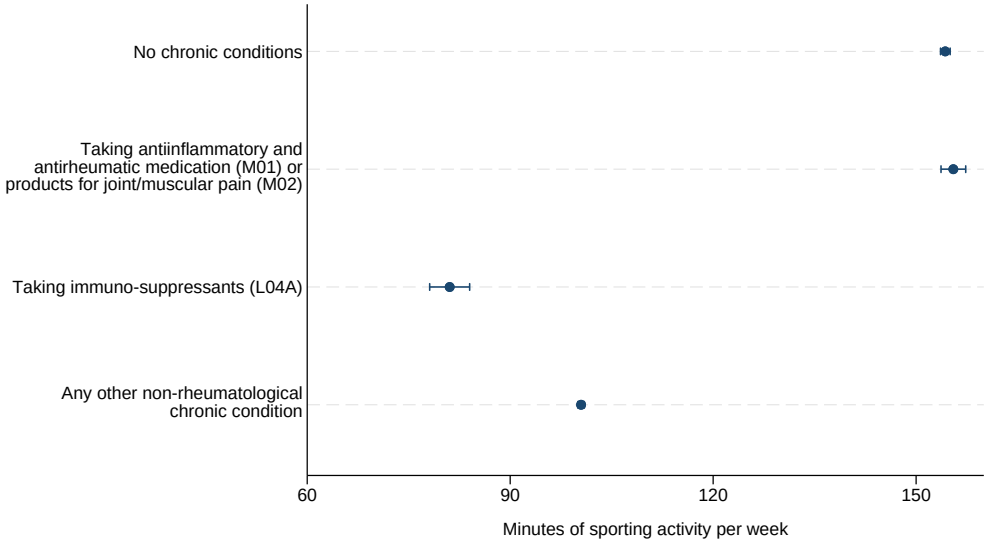
Note: This figure shows the number of anemia cases per 1,000 individuals for men and women between 2011 and 2021. For women, it also shows how the lifecycle profile changes when anemia is required to be observed for two or three consecutive years.

Figure D.2: IDENTIFYING MEDICINES FOR RHEUMATOLOGICAL CONDITIONS

A. Share of Prescribed Medicines by Age



B. Sporting Activity by Medication Groups



Note: Panel A shows the share of individuals taking different types of medicines for rheumatic conditions in 2012. Panel B displays minutes of sporting activity for different medication groups. L04A denotes immunosuppressants excluding corticosteroids. M01 denotes anti-inflammatory and antirheumatic products. M02 denotes topical products for joint and muscular pain. Both M01 and M02 were included in Huber et al. (2013), but are excluded from our analysis.

E Chronic Disease and the Mortality Gap

This appendix documents the samples and regression specifications underlying Figure 3, Appendix Figure E.2, and the Oaxaca–Blinder decompositions in Appendix Figures E.1 and E.3. These exercises support the evidence summarized in Section 2 and the prevalence-versus-treatment analysis in Section 4.3.

For each age a and outcome Z, we estimate

$$Z_{i,a,t} = \delta_a L_{i,a,t} + R'_{i,a,t}\gamma_{a,g} + \lambda_t + \eta_g + \varepsilon_{i,a,t}, \tag{12}$$

where $Z_{i,a,t}$ is either an indicator for death within five years or total annual healthcare costs, $L_{i,a,t}$ indicates income percentiles 1–50, and $R_{i,a,t}$ is the set of health controls included in a given row. Income percentiles 51–100 form the reference group, so δ_a is the low-minus-high income gap. Mortality coefficients are reported per 1,000 and costs in euros. All specifications include calendar-year effects λ_t and a gender indicator η_g, and allow the health-related coefficients $\gamma_{a,g}$ to vary by gender. Panel A of each figure fixes $a = 70$; Panel B estimates the models separately at ages 40, 45, 50, 55, 60, 65, and 70. The five-year-mortality regressions pool base years 2013–2016.[26] At age 70, the mortality panel contains 680,362 person-year observations and the healthcare-cost panel contains 1,271,497; the latter can use additional outcome years because it does not require five years of subsequent mortality data.

The rows in Figure 3 and Appendix Figure E.2 use the same sequence of covariates. Specification *A1 Baseline* includes the low-income indicator, calendar-year effects, and gender. Specification *A2 + Chronic conditions (lag 1)* adds indicators for the 22 chronic conditions observed in the previous year. Specification *B1 Chronic conditions (lags 2–3)* adds the second and third lags of those indicators to A2. Specification *B2 Non-chronic prescriptions* instead adds to A2 indicators for medications not included in the chronic-condition mapping. These medication indicators are selected using separate gender-specific Lasso regressions of five-year mortality: the penalty is set to retain 20 indicators for each gender, and their union contains 24 distinct indicators.

Specification *C1 Hospitalizations* adds to A2 indicators for previous-year primary-diagnosis ICD codes and fourth-degree polynomials in the previous-year number of hospitalizations,

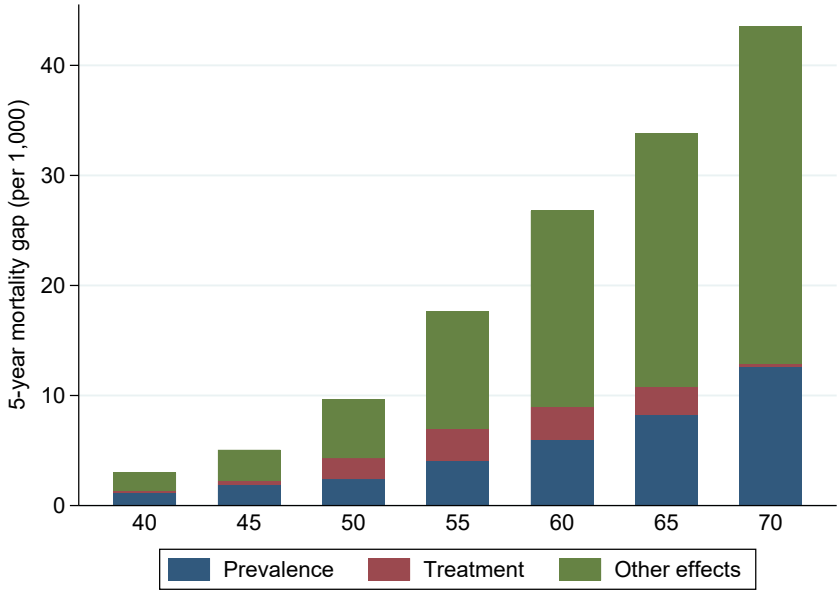
26. For individuals aged 65 and older, income is fixed at the average observed at ages 60–62. Income data begin in 2003, so the first cohort for which this pre-retirement measure is available reaches age 70 in 2013. Mortality is observed through 2021, making five-year mortality available through base year 2016.

hospitalized nights, and hospital costs. Specification *C2 Healthcare costs, by type* adds to A2 fourth-degree polynomials in previous-year general-practitioner, medication, mental-health, and other healthcare costs. Specification *D All health information (A1–C2)* includes the union of the controls in the preceding rows. The R^2 labels next to the estimates report the fit of the corresponding outcome regression. At age 70, the mortality R^2 rises from 0.010 in A1 to 0.051 in A2 and 0.134 in D; for healthcare costs, the corresponding values are 0.003, 0.079, and 0.281. The coefficient sequence and the R^2 annotations therefore distinguish how much each control set changes the estimated income gap from how much it improves overall predictive fit.

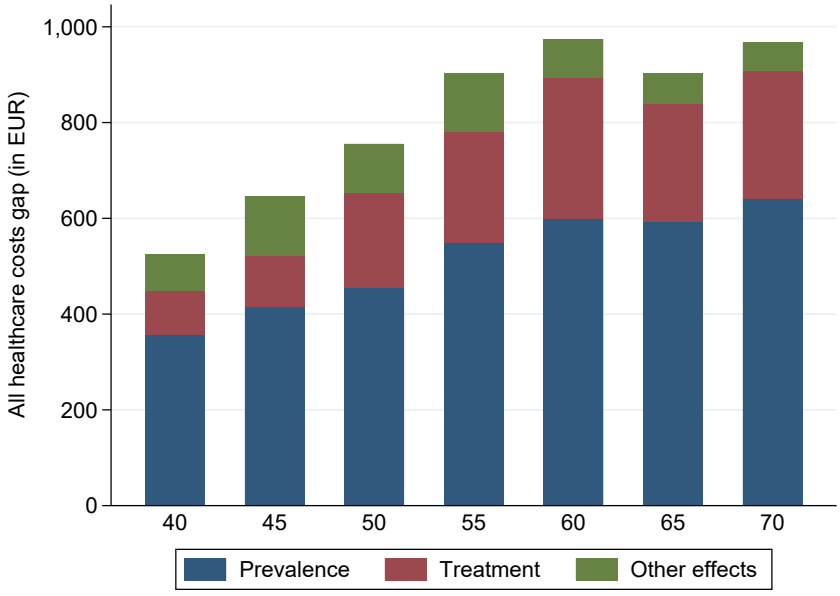
Appendix Figure E.1 implements the three-way decomposition in equation (2) separately at ages 40, 45, 50, 55, 60, 65, and 70. For each outcome and age, separate regressions are estimated for the low- and high-income groups using the 22 chronic-condition indicators observed in $t-1$. The "Prevalence" component reflects differences between groups in these condition indicators, the "Treatment" component reflects differences in their estimated coefficients excluding the constant, and "Other effects" reflects the difference in estimated constants. Appendix Figure E.3 assesses sensitivity to richer disease histories and comorbidity: it includes condition indicators from $t-1$, $t-2$, and $t-3$, identifies the ten most frequent two-way condition combinations from the $t-1$ distribution, and includes the corresponding interactions at each lag. Thus, the comparison between the two figures isolates whether the prevalence-versus-treatment conclusion depends on using only the most recent condition indicators.

Figure E.1: OAXACA-BLINDER DECOMPOSITION OF FIVE-YEAR MORTALITY AND HEALTHCARE COSTS

A. Five-year mortality



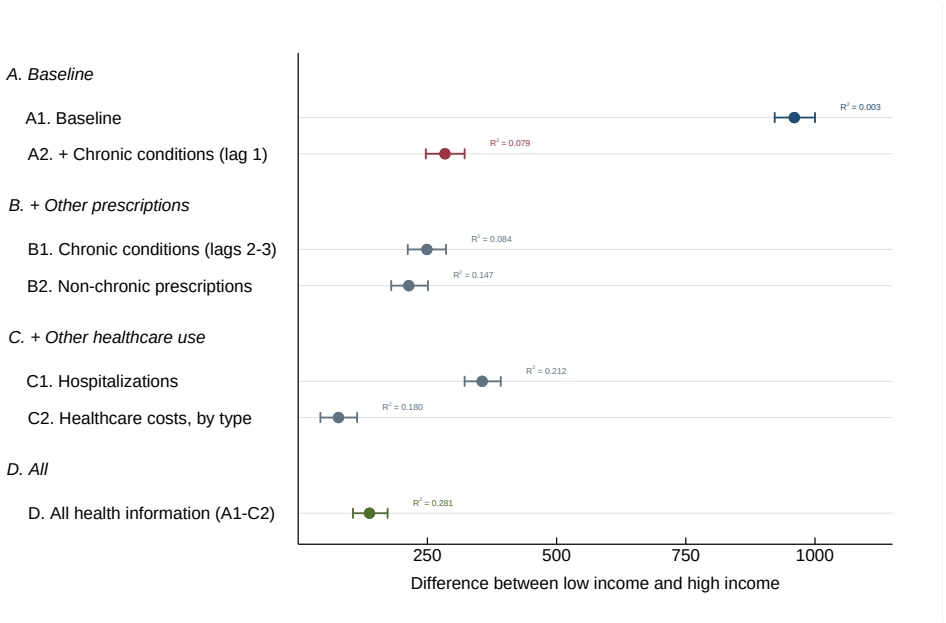
B. Healthcare costs



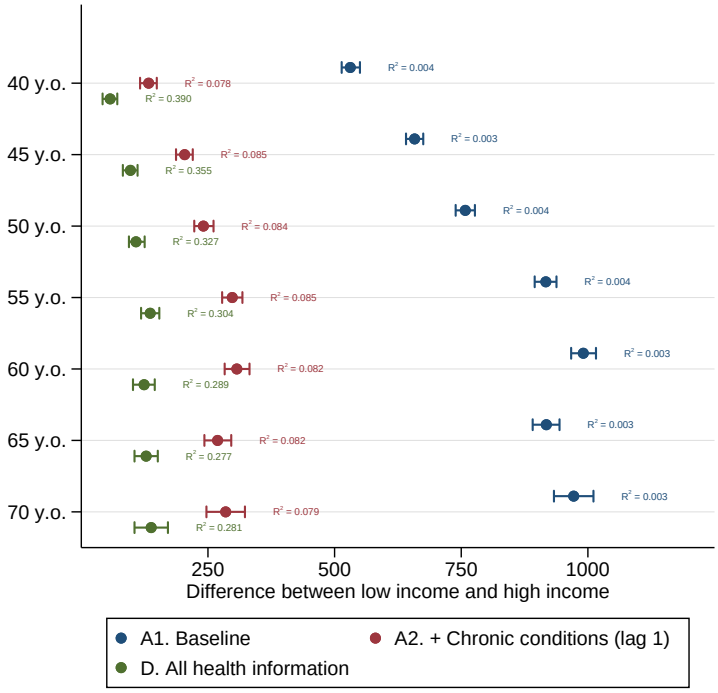
Note: The figure reports the results of a three-way Oaxaca-Blinder decomposition of five-year mortality in Panel A and total healthcare costs in Panel B, using lagged chronic condition indicators from the previous year as predictors. The two groups are low- and high-income individuals, using the median standardized household income as the threshold. The "Prevalence" component is the part of the mean difference explained by intergroup differences in chronic-condition endowments; the "Treatment" component is the part explained by intergroup differences in coefficients, excluding the constant term; and the "Other effects" component is the part explained by intergroup differences in the estimated constant term.

Figure E.2: GAP IN TOTAL HEALTHCARE COSTS BETWEEN HIGH- AND LOW-INCOME INDIVIDUALS

A. Gap in Healthcare Costs (€) at 70 Years of Age



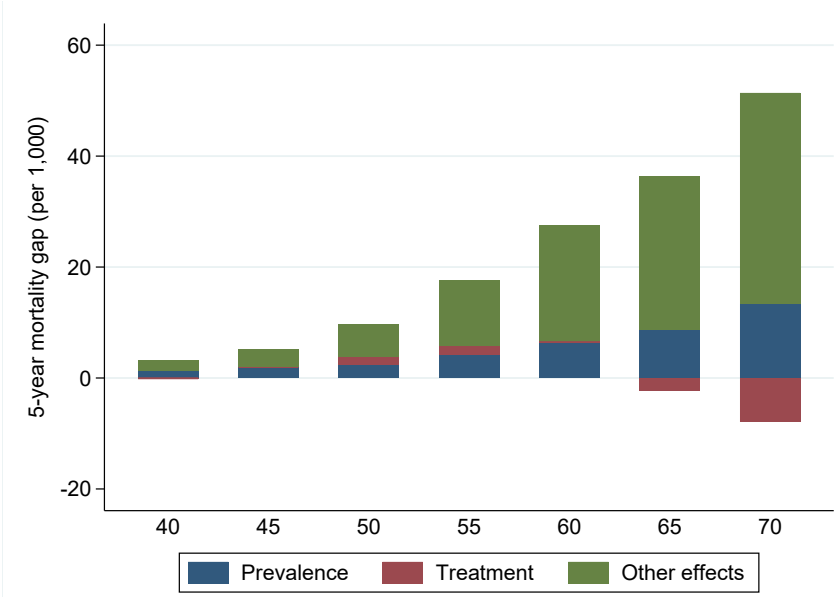
B. Gap in Healthcare Costs (€) at Different Ages



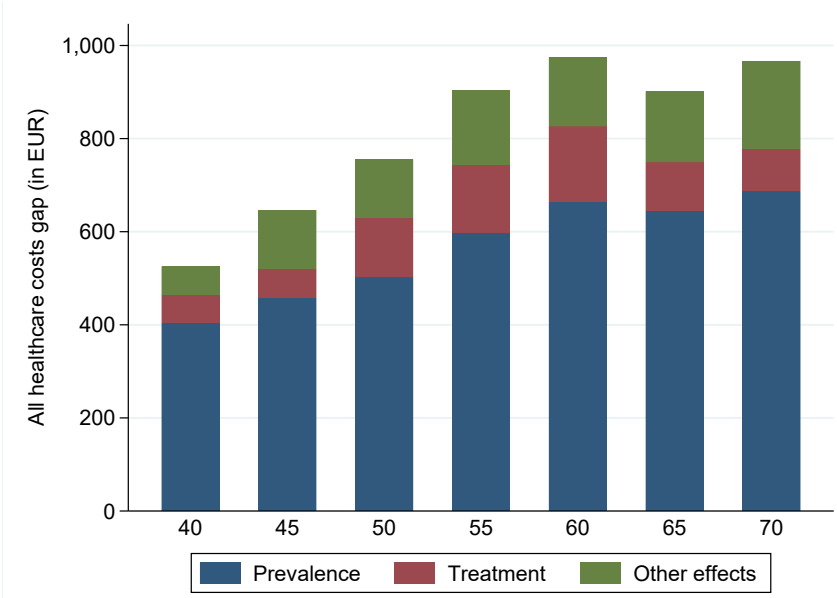
Note: In Panel A, each row corresponds to a different regression of total healthcare costs, in euros, on income, defined as low versus high income, and a set of controls identified by the row label. For each specification, the plot shows the estimated coefficient on income. Specifications reported after A2 include all the controls used in A2 as well as the additional controls listed on the left. Panel B reports healthcare-cost gap estimates from specifications A1, A2, and D at different ages.

Figure E.3: OAXACA-BLINDER DECOMPOSITION, USING MORE LAGGED AND INTERACTED CHRONIC CONDITIONS

A. Five-Year Mortality



B. Healthcare Costs



Note: The figure reports the results of a three-way Oaxaca-Blinder decomposition of five-year mortality in Panel A and total healthcare costs in Panel B, using lagged chronic condition indicators from the previous three years as predictors. The ten most frequent within-period chronic-condition interactions are also included. The two groups are low- and high-income individuals, using the median standardized household income as the threshold. The "Prevalence" component is the part of the mean difference explained by intergroup differences in chronic-condition endowments; the "Treatment" component is the part explained by intergroup differences in coefficients, excluding the constant term; and the "Other effects" component is the part explained by intergroup differences in the estimated constant term.

F Details of the Chronic Disease Index Model

This section provides more details on the CDI model and describes a series of considerations to understand the robustness of the CDI predictions to various modelling choices. Overall, the CDI is highly robust along these dimensions, as reported in Table F.1.

Lag Structure: One of the main modelling choices in constructing the CDI is selecting the lag sequence of chronic conditions. We choose the set $t = \{-1, -2, -3\}$: longer lags potentially contain more information, but would preclude certain cohorts from the sample. As shown in Figure F.1, the amount of additional information provided by the fourth and higher lags starts to taper off. This is despite those higher lags being highly predictive in themselves—due to the persistent nature of the conditions in question.

Linkage functions: For tractability, the CDI is estimated as a linear probability model. However, provided separability is maintained between the set of chronic conditions and the socioeconomic variables, other linkage functions may be used, for example a logistic or Gompertz function, which naturally bound the prediction range and have been used elsewhere in the literature. We have tested these two alternatives, but they do not increase predictive power; because the model is sufficiently regularised through variable selection, few predicted values fall outside the $[0, 1]$ bounds.

Linear model: As a benchmark, we construct $\widehat{M}_i^+$, including chronic conditions additively, without any interactions. This is shown in equation (14) below. The CDI outperforms this benchmark in terms of R^2 (+10%), while the AUC statistic is a marginal improvement.

$$M_i^+ = CC_i^{*\prime}\beta_{CC}^+ + X_i^{*\prime}\beta_X^+ + \zeta_i^+ \tag{13}$$

$$\widehat{M}_i^+ = CC_i^{*\prime}\widehat{\beta}_{CC}^+ + \overline{X}_i^{*\prime}\widehat{\beta}_X^+ \tag{14}$$

Target ages: As described in Section 4.4, we select age 70 as the reference age for the CDI estimation. This balances several considerations: selection among 70-year-olds through earlier mortality is still limited, under-diagnosis is not yet acute, and mortality risk is high enough to provide meaningful outcome variation. Alternative reference samples at age 65 and ages 65–75 do not qualitatively change the CDI estimates, although the explained share of the health gap is slightly smaller.

Variable selection: The advantage of a Lasso framework is that modelling decisions on spec-

ification and functional form can be data-driven rather than based on ad hoc choices. Some researcher decisions nevertheless remain. First, the candidate variable set cannot include every possible interaction among the predictors documented in Section 2 because of computational constraints. Since fewer than half of the candidate variables are selected, we judge the risk of material omission from this set to be limited. Second, the penalty parameter λ is chosen using the default `GLMNET` criterion: the most regularised model whose cross-validated error lies within one standard error of the minimum mean squared error. Choosing λ to minimise mean squared error selects more variables but does not materially alter model fit, CDI predictions, or the subsequent findings. Third, we conduct a separate Lasso selection for each chronic condition to identify relevant socioeconomic variables. A group Lasso could instead select $f(C_i)^*$ in one step, as in a seemingly unrelated regression framework (Yuan and Lin 2006). Obozinski, Wainwright, and Jordan (2011) show that separate Lasso steps can perform better when the dimensions of CC_i are highly correlated, supporting our approach given the correlation among chronic conditions.

Table F.1: SUMMARY STATISTICS ON ALTERNATIVE CDI SPECIFICATIONS

	Baseline CDI	Estimated at 65	Estimated on positive medication subsample	Logistic regression
A. Model performance (test sample)				
Test R squared	0.052	0.035	0.056	0.052
Test AUC	0.663	0.654	0.679	0.666
Estimation sample size	402,500	554,519	353,643	402,500
B. Predicted CDI distribution at 70 (whole population)				
10th percentile	0.048	0.034	0.047	0.045
Median	0.077	0.051	0.077	0.073
90th percentile	0.172	0.114	0.171	0.157
C. Explained gradient at 70 (whole population)				
Explained 5-year mortality gap	0.297	0.192	0.293	0.273
Explained healthcare costs gap	0.555	0.492	0.554	0.541

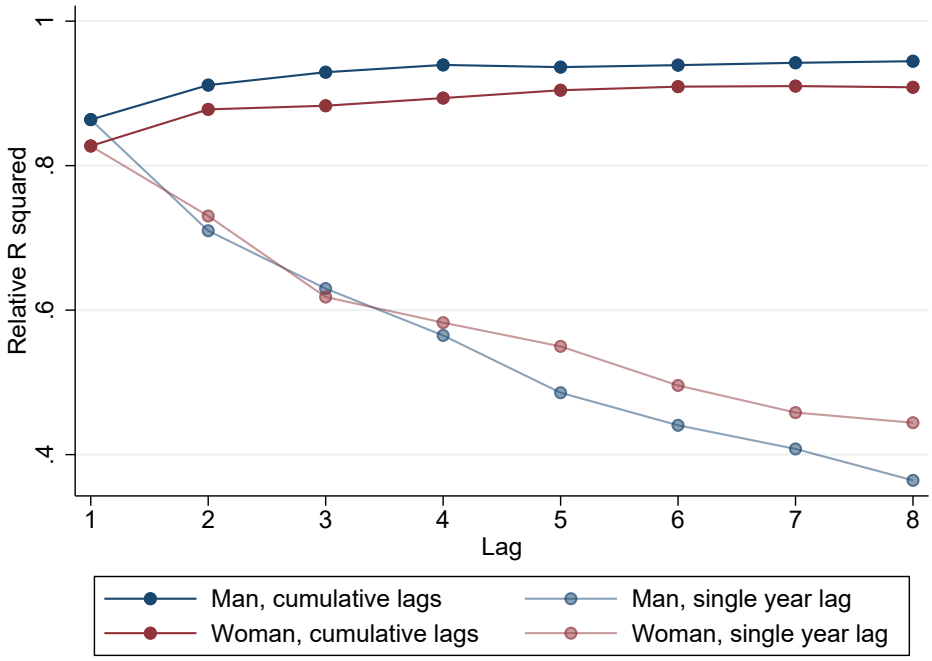
Note: The table displays three sets of statistics on the baseline CDI (the chronic disease index introduced in Section 4.4), and alternative versions estimated as robustness checks. One version was estimated using a sample of 65 year-olds (instead of 70 year-olds as is the case for the baseline CDI); another one used a sample limited to individuals reported to be taking at least one medication; finally, a version models the relationship between five-year mortality and its predictors using a logistic regression.

Table F.2: PREDICTORS INCLUDED IN THE CDI

Socioeconomic Status Predictors	
A. Individual Variables	
Gender	Household Composition*
Percentile of Household Disposable Income	Foreign Parents
Number of Household members with Income*	Household Main Source of Income*
Position in Household*	Work Status*
Percentile of Wealth Income**	Percentile of Household Assets**
Main source of household income*	Percentile of Household Savings
Calendar Year	House Owner*
Foreign Born	Percentile of Home value
Number of Household Members*	Percentile of Personal Primary Income
B. Interaction Terms	
Percentile of disposable Income x Main Income source	Percentile of Primary Income x House Owner
Percentile of Primary Income x Household Composition	Percentile of Personal Net Income x Work Status
Percentile of Personal Net Income x Percentile of Disposable Income	Percentile of Gross Income x Main Income Source
Percentile of Gross Income x Main Income Source	Percentile of Primary Income x Main Income Source
Percentile of Disposable Income x Percentile of Household Assets	Percentile of Disposable Income x Main source of Income
Percentile of Wealth Income x Percentile of Household Assets	Personal Net Income x Work Status
Percentile of Disposable Income x Percentile of Wealth Income	Percentile of Disposable Income x Main Income Source
Percentile of Personal Gross Income x Work Status	Percentile of Wealth Income x Percentile of Household Assets

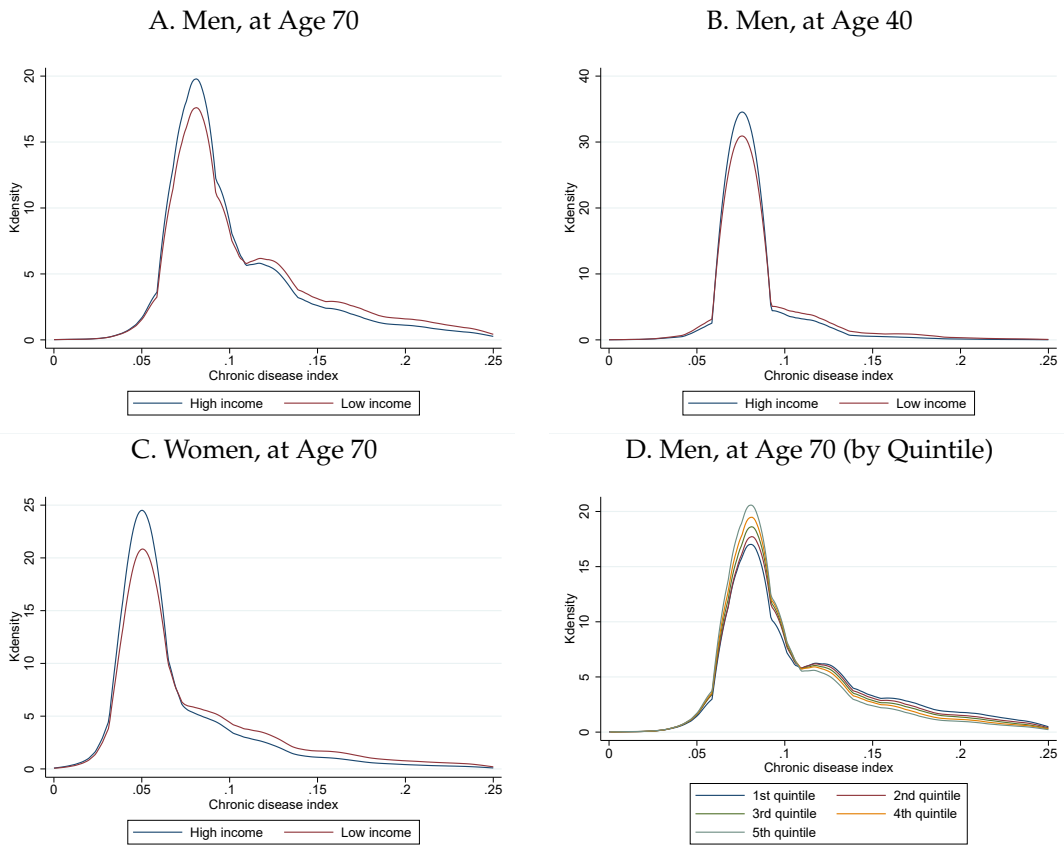
Note: This table presents the list of socioeconomic status variables included through the Lasso selection procedure. Variables for which multiple lags are included are denoted with *. Variables for which multiple lags and higher-order terms are included are indicated with **.

Figure F.1: PREDICTIVE POWER OF CHRONIC CONDITIONS, BY LAG LENGTH



Note: This figure plots the predictive power of varying lags of chronic conditions on five-year mortality. Predictive power is measured as a relative R^2 statistic: $R^2_{s,t}/R^2_{(1,3),Int}$. The numerator $R^2_{s,t}$ is computed from a regression of five-year mortality on the set of chronic-condition lags between s and t, without interactions. $R^2_{(1,3),Int}$ is computed from our preferred specification, using lags 1-3 of chronic conditions with interactions.

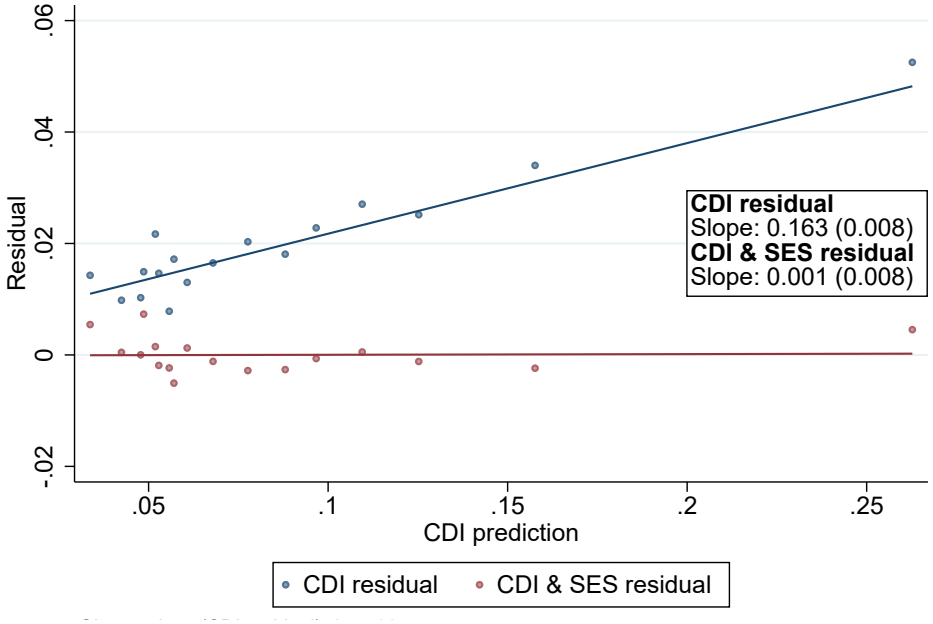
Figure F.2: DISTRIBUTION OF THE CDI BY INCOME GROUPS AND GENDER



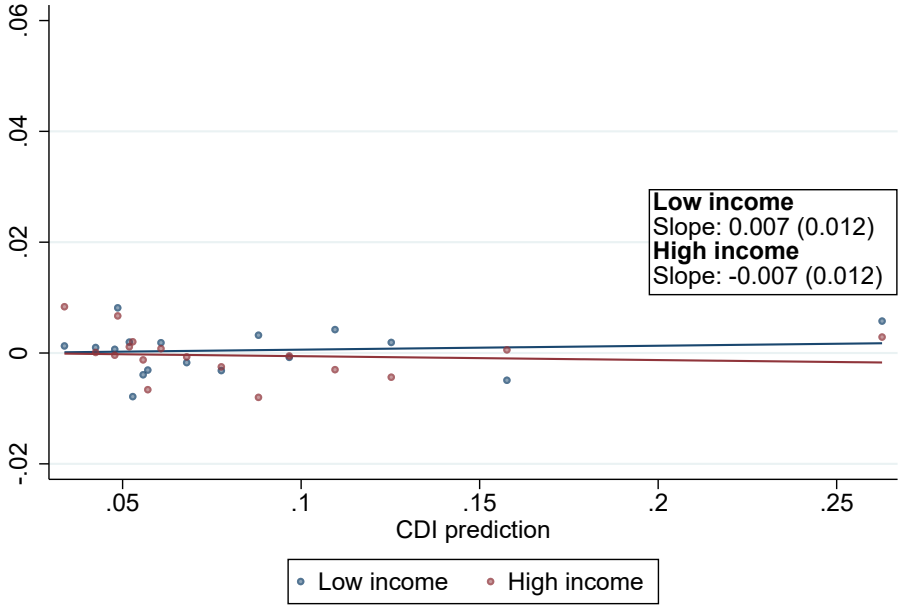
Note: The panels report kernel-density estimates of the chronic disease index at different ages and for different income splits, separately for men and women. The range of the x-axis is limited to the interval [0, 0.25] to avoid showing low-density tails composed of outliers.

Figure F.3: DIAGNOSTIC BINNED SCATTERPLOTS OF THE CDI

A. Residual mortality risk



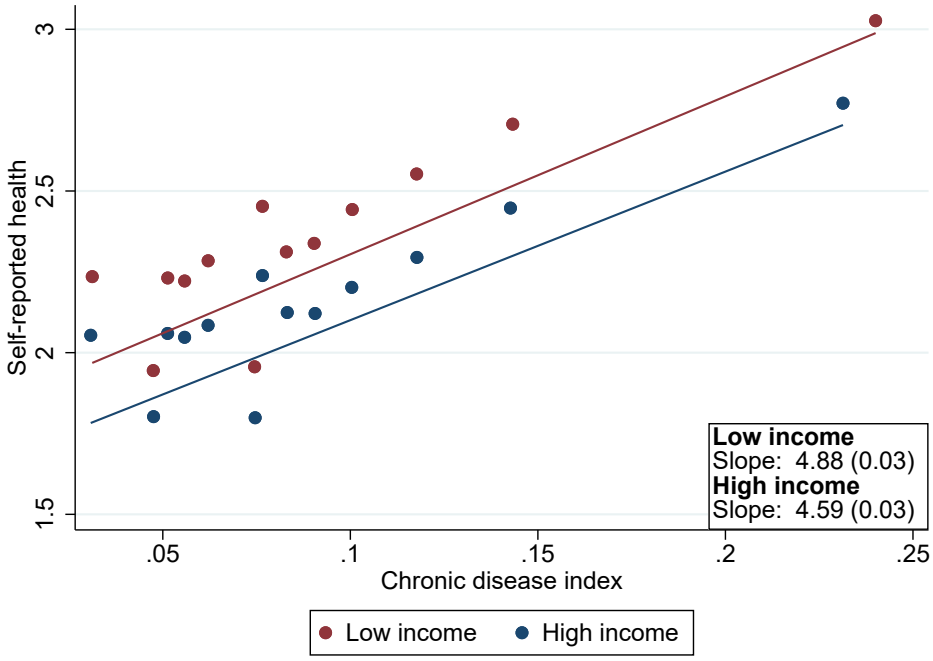
B. CDI & SES residual, by income



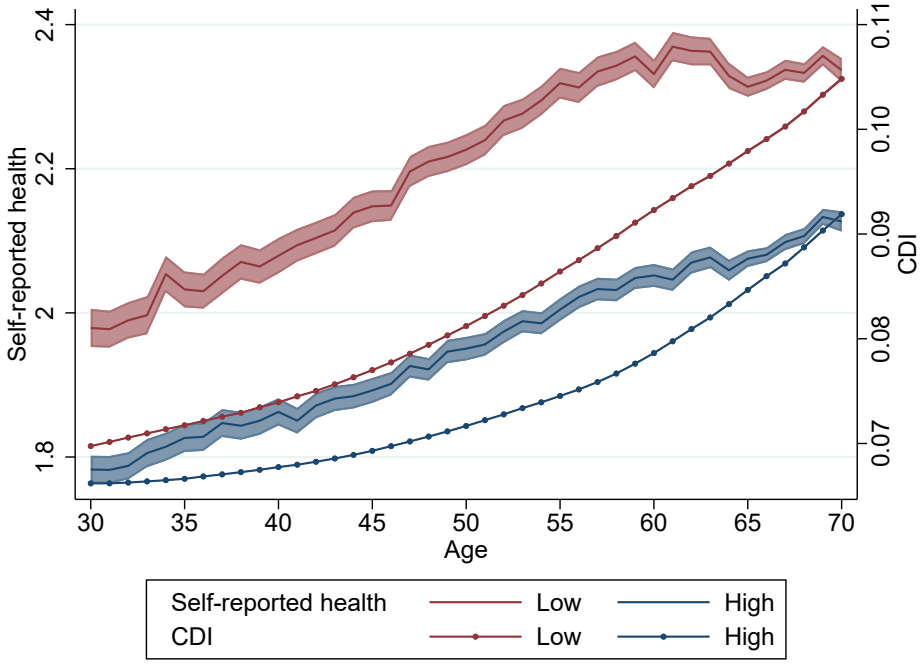
Note: Panel A depicts two series: the CDI residual, $E\left[m_i - CDI_i \mid CDI_i\right]$, and the CDI & SES residual, which is equivalent to the fitted error term in equation (10): $E\left[\hat{\zeta}_i \mid CDI_i\right]$. The wedge of 0.163 represents the bias excluded by the double-selection procedure. The bias arises because socioeconomic status is correlated with chronic conditions. Panel B depicts the CDI & SES residual separately for low- and high-income individuals: $E\left[\hat{\zeta}_i \mid CDI_i, Y_i = Y_L\right]$ and $E\left[\hat{\zeta}_i \mid CDI_i, Y_i = Y_H\right]$.

Figure F.4: DIAGNOSTIC BINNED SCATTERPLOTS OF SELF-REPORTED HEALTH

A. Self-reported health and CDI, by income



B. Self-reported health and CDI over the lifecycle



Note: Panel A shows a binned scatter plot with the chronic disease index on the horizontal axis and average self-reported health from the *Gezondheidsmonitor* survey data on the vertical axis, where higher values denote worse health. Panel B plots the evolution of self-reported health and the CDI over the lifecycle, along with 95% confidence intervals. The CDI series pool observations from 2009 to 2021 and are identical to those reported in Figure 6. The CDI confidence interval is within the thickness of the connector line.

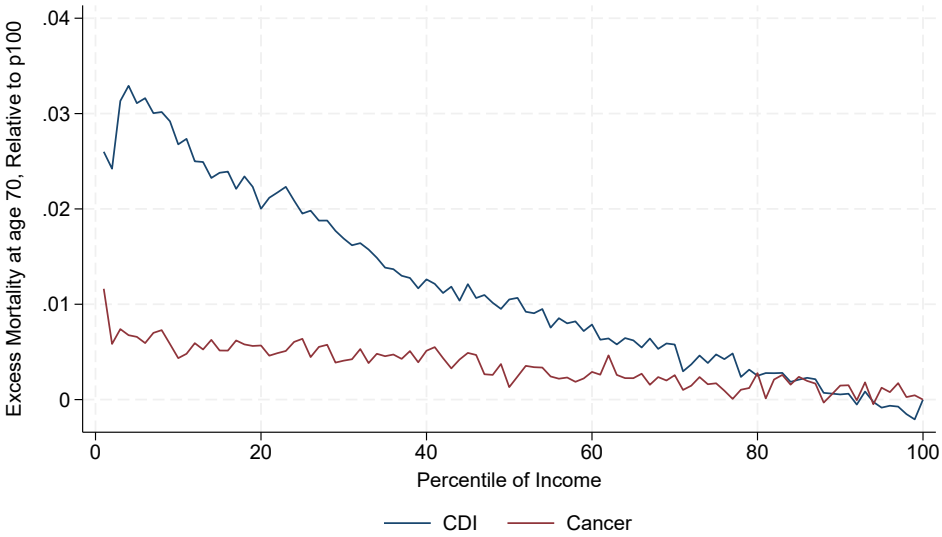
F.1 Comparison with Cancer Registry Data

Cancer is the main morbidity category not comprehensively captured by our medication-based chronic disease measure. In a companion paper (Danesh et al. 2026), we study the socioeconomic gradient in cancer using data from the Netherlands Comprehensive Cancer Organization (Dutch: *Integraal Kankercentrum Nederland*, IKNL). As discussed in Section 4.1, these data are rich but not directly suited to inclusion in the CDI: they capture incident diagnoses rather than repeated prevalence, and are available for a shorter panel than the medication data used to construct the index. We therefore treat cancer as an important external benchmark for the scope of the CDI, rather than as an input into the baseline measure.

We use the overlap between the cancer registry, the medication data, and mortality records to compare the income gradient in the CDI with the mortality-relevant gradient in cancer. For individuals aged 70 between 2016 and 2019, we observe ten years of prior cancer incidence and five-year mortality. We approximate cancer prevalence using the first cancer diagnosis recorded in the previous ten years, including information on cancer type and stage at diagnosis. We then estimate sex-specific regressions of five-year mortality on cancer type-by-stage indicators and income percentile indicators. To isolate the mortality gradient implied by observed cancer histories, we predict mortality for each individual holding income fixed at the median percentile. Averaging these predictions by income percentile gives a measure of the cancer-associated mortality gradient that can be compared to the average CDI gradient at age 70.

This comparison is not a decomposition of the total health gap between these two measures. Cancer and chronic disease burden are not mutually exclusive: many individuals with high CDI values may also have cancer, and the CDI may capture comorbidities that are related to cancer diagnosis, treatment, or survival. The comparison instead asks how large the cancer-related gradient appears to be relative to the broader chronic disease gradient captured by the CDI. Figure F.5 shows that the CDI gradient is almost five times steeper than the gradient in cancer-associated mortality risk. Cancer is therefore a material part of the overall health inequality story, but the CDI captures a substantially broader set of mortality-relevant health differences across the income distribution.

Figure F.5: COMPARING CANCER-ASSOCIATED MORTALITY GRADIENTS TO THE CDI GRADIENT



Note: The figure compares the income gradient in the age-70 CDI with the income gradient in predicted five-year mortality risk associated with cancer history. The cancer measure is constructed for individuals aged 70 between 2016 and 2019 using cancer diagnoses recorded in the previous ten years. Cancer-associated mortality risk is predicted from sex-specific regressions of five-year mortality on cancer type-by-stage indicators and income percentile indicators, holding income fixed at the median percentile.

G Dynamic Decomposition

In Figure 6, the average CDI for both income groups is shown over the lifecycle. The slope of these curves, i.e. the difference in the CDI between two consecutive ages for a given income group, can be denoted as $E_{a+1}[CDI_{i,a+1}|Y_{a+1}] - E_a[CDI_{i,a}|Y_a]$, where Y_a denotes the set of individuals who belong to income group Y at age a. The subscript on the expectation operator indicates that we are taking expectations over those observed at age a.

We can decompose the slope of these curves into several terms:

$$\begin{aligned} E_{a+1}[CDI_{i,a+1}|Y_{a+1}] - E_a[CDI_{i,a}|Y_a] = & [E_{a+1}(CDI_{i,a+1}|Y_{a+1}) - E_{a+1,a}(CDI_{i,a+1}|Y_{a+1})] \\ & + [E_{a+1,a}(CDI_{i,a+1}|Y_{a+1}) - E_{a+1,a}(CDI_{i,a+1}|Y_a)] \\ & + [E_{a+1,a}(CDI_{i,a+1} - CDI_{i,a}|Y_a)] \\ & + [E_{a+1,a}(CDI_{i,a}|Y_a) - E_a(CDI_{i,a}|Y_a, S_{a+1})] \\ & + [E_a(CDI_{i,a}|Y_a, S_{a+1}) - E_a(CDI_{i,a}|Y_a)] \end{aligned} \tag{15}$$

Below, we describe the interpretation for each of these terms.

1. **Aging**: for individuals in Y_a observed during both periods, we can calculate the average change in their outcome measure between a and $a+1$. We call this the Aging effect:

$$Aging = E_{a+1,a}(CDI_{i,a+1} - CDI_{i,a}|Y_a) \tag{16}$$

   Note that $E_{a+1,a}$ denotes the mean outcome for individuals who were observed in the sample both at age $a+1$ and a.

2. **Health-based Sorting**: over the lifecycle, people move between different income groups. Conditional on observing people in both periods, we will see two types of transitions: some people who were in Y_a will not be in Y_{a+1}, and some people who were not in Y_a will now be in Y_{a+1}. The (net) sorting effect is just the difference in mean outcome at age $a+1$ between the members of Y_{a+1} and the members of Y_a:

$$Sorting = E_{a+1,a}(CDI_{i,a+1}|Y_{a+1}) - E_{a+1,a}(CDI_{i,a+1}|Y_a) \tag{17}$$

3. **Attrition due to Death**: some individuals who were in Y_a died at some point during that

year. Call the set of people who survived until age $a+1$, S_{a+1}. The attrition due to mortality is the difference in mean CDI at age a between those individuals in Y_a who survived until age $a+1$ and the mean CDI of all observed in income group Y_a.

$$Attrition = E_a(CDI_{i,a}|Y_a, S_{a+1}) - E_a(CDI_{i,a}|Y_a) \tag{18}$$

4. **Cohort Effect**: This is composed of exit and entry effects out of our sample.

   First, some individuals who were in Y_a and survived into age $a+1$ are no longer in the sample at age $a+1$. This may occur because they emigrated or aged out of the sample period. The exit effect is the difference in mean CDI at age a between individuals in Y_a who remained in the sample and all those who survived; equivalently, it captures the CDI of those who left the sample for reasons other than death.

$$Exit = E_{a+1,a}(CDI_{i,a}|Y_a) - E_a(CDI_{i,a}|Y_a, S_{a+1}) \tag{19}$$

   Second, some individuals in Y_a were not observed in the sample at age $a-1$. This may be because they immigrated, were born, or aged into the sample period. The "entry" effect is the difference in the mean outcome at age a between all individuals in Y_a and the subset who were also observed at age $a-1$ (in other words, it captures the contribution of individuals who were not observed at age $a-1$ to the mean CDI at age a).

$$Entry = E_{a+1}(CDI_{i,a+1}|Y_{a+1}) - E_{a+1,a}(CDI_{i,a+1}|Y_{a+1}) \tag{20}$$

   The exit and entry effects are then combined into the "cohort effects," which capture cohort, time, and migration effects.

$$Cohort = Exit + Entry \tag{21}$$

In our main lifecycle decomposition, we estimate these effects for the low- and high-income groups, pooling observations from 2009–2021. Appendix Figure G.1 presents the results. Aging effects increase steadily over the lifecycle for both groups, while sorting effects are most important around labour-market entry and exit. Mortality attrition becomes relevant at later ages and is

stronger for the low-income group because low-income individuals die at higher rates. Table 2 reports the low-minus-high difference in each component.

The estimated effects can be used to simulate counterfactual CDI evolutions. Panel A of Figure 8 simulates the CDI for both income groups with a) the aging component only, and b) the aging plus health-based sorting components. Figure G.2 repeats this exercise, also accounting for attrition due to death effects. Because those attrition effects are more strongly negative for the low income group, the gap in counterfactual CDIs is smaller when we account for them. Similarly, the biological age gaps in Panel B of Figure G.2 are somewhat smaller than those in Figure C.9.

Using this lifecycle decomposition framework, we can also compare finer income groups and other socioeconomic characteristics. Panel B of Figure G.3 reports aging and sorting effects by income quintile and age. Aging effects are largest for the bottom quintile and decline monotonically with income. Sorting effects have a different sign pattern: they increase the average CDI of the bottom quintile and reduce it in the other four quintiles, consistent with individuals whose health worsens becoming concentrated at the bottom of the income distribution. The positive sorting effect for the bottom quintile is largest between ages 25–30, when young adults are forming households and establishing their careers, and becomes less pronounced towards retirement. It is zero by construction from ages 65–70 because we use pre-retirement income. Appendix Table G.1 confirms that aging effects decline with income quintile; its education-based decomposition similarly shows that more highly educated individuals age more slowly than less-educated individuals at comparable ages.

Furthermore, we also test our decomposition by imposing two robustness checks on the timing of our income variable. First, we restrict the sample to only consider individuals who have been in an age group for at least two years and then run the decomposition again. This means that at age a, only individuals who were in the same income group at a and at $a-1$ are included. Because those individuals are more fixedly in the relevant income group, we might obtain a more 'pure' aging effect. Under this so-called 'Markov Restriction', the aging effect can be written as:

$$Aging_{MR} = E_{a+1,a,a-1}(CDI_{i,a+1} - CDI_{i,a}|Y_a, Y_{a-1}) \tag{22}$$

And sorting is:

$$Sorting_{MR} = E_{a+1,a,a-1}(CDI_{i,a+1}|Y_{a+1}) - E_{a+1,a,a-1}(CDI_{i,a+1}|Y_a, Y_{a-1}) \tag{23}$$

In the second robustness check, we adapt our income definition and use a rolling average of Y_{a-3}, Y_{a-2} and Y_{a-1}. In this definition, we use income at the same ages for which chronic condition indicators are used to predict the CDI. Therefore, we call the second alternative 'Contemporaneous Income'. Panel A of Figure G.3 summarizes aging and sorting effects for both robustness checks, along with the original decomposition. The decomposition results are robust to different income definitions, as the obtained effects are very close to those of the original decomposition.

Finally, we test whether changes in household composition drive the estimated sorting effects by estimating them separately for individuals whose household composition changes and those whose composition remains constant. Appendix Table G.2 shows sorting in both groups. The sorting gap among individuals with unchanged household composition is also close to the gap in our main decomposition, indicating that movements into new households do not drive the result.

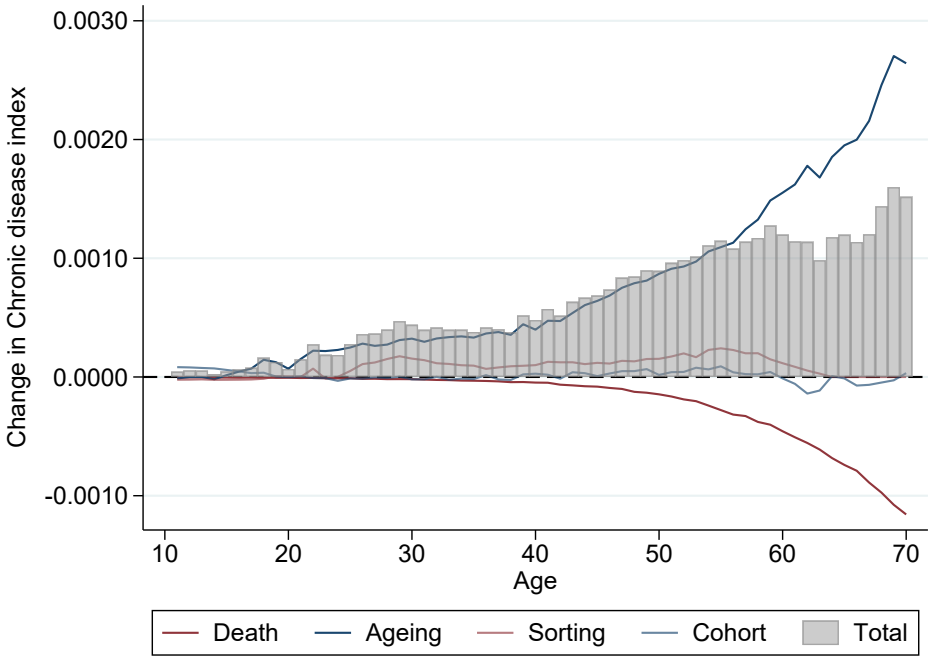
Table G.1: AGING EFFECT BY INCOME QUINTILE AND EDUCATION LEVEL, $\times 100$

	11-20	21-30	31-40	41-50	51-60	61-70	Life-cycle Effect
A.By Income Quintile							
Q1	0.057	0.248	0.425	0.785	1.270	2.196	4.982
Q2	0.038	0.266	0.323	0.623	1.143	2.061	4.454
Q3	0.032	0.215	0.280	0.487	0.988	1.889	3.891
Q4	0.030	0.165	0.227	0.407	0.859	1.685	3.373
Q5	0.052	0.114	0.177	0.272	0.659	1.424	2.698
B. By Education level							
No High School	-	0.374	0.681	1.145	1.472	2.233	5.911
High School	-	0.269	0.421	0.711	1.185	1.926	4.511
Bachelor	-	0.094	0.182	0.372	0.779	1.563	2.990
Master or PhD	-	0.0809	0.1529	0.260	0.6239	1.4294	2.548

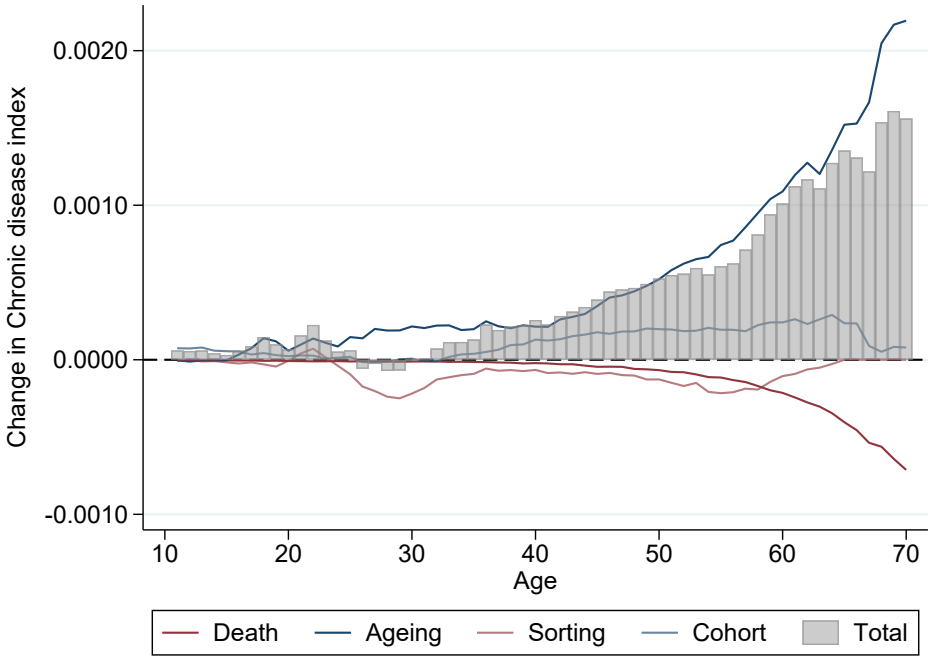
Note: This table reports aging effects for 5 income quintiles and 4 education groups separately. Effects are reported as the total contribution to the change in CDI of the aging effect for 10-year age groups, multiplied by 100. More detail on the methodology used to perform the lifecycle decomposition is provided in Appendix Section G.

Figure G.1: DECOMPOSITION BY INCOME GROUP

A. Low Income



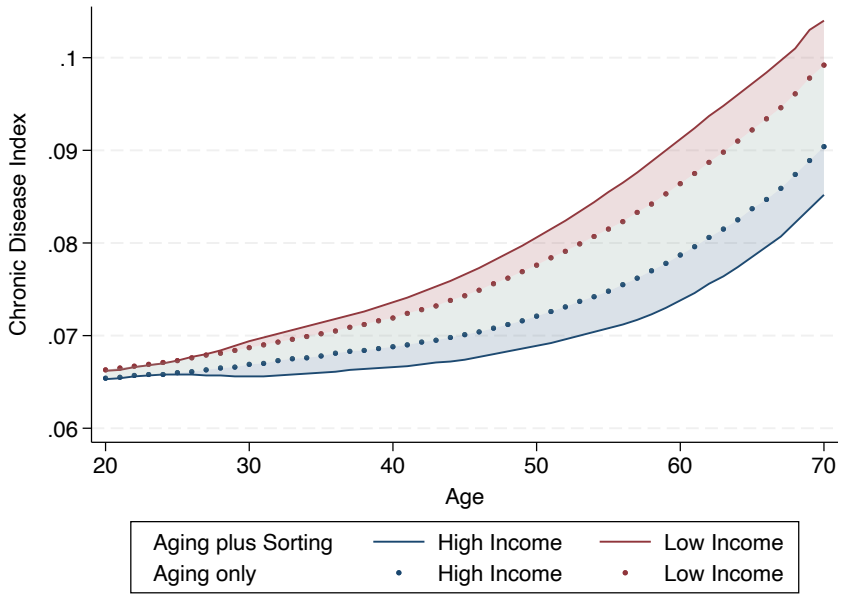
B. High Income



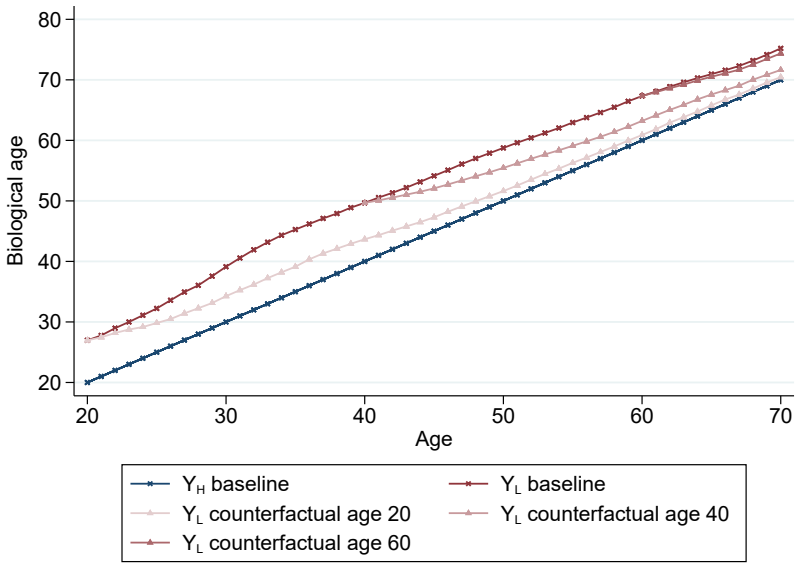
Note: This figure presents the full decomposition of the Chronic Disease Index. Panel A reports the low-income group and Panel B the high-income group. For each group, the total change between ages $a-1$ and a is decomposed into mortality attrition, aging, sorting, and cohort effects. The estimates pool observations from 2009–2021. Table 2 reports the low-minus-high difference in each component.

Figure G.2: BIOLOGICAL AGING, ACCOUNTING FOR ATTRITION DUE TO DEATH

A. Differential Aging and Health-based Sorting



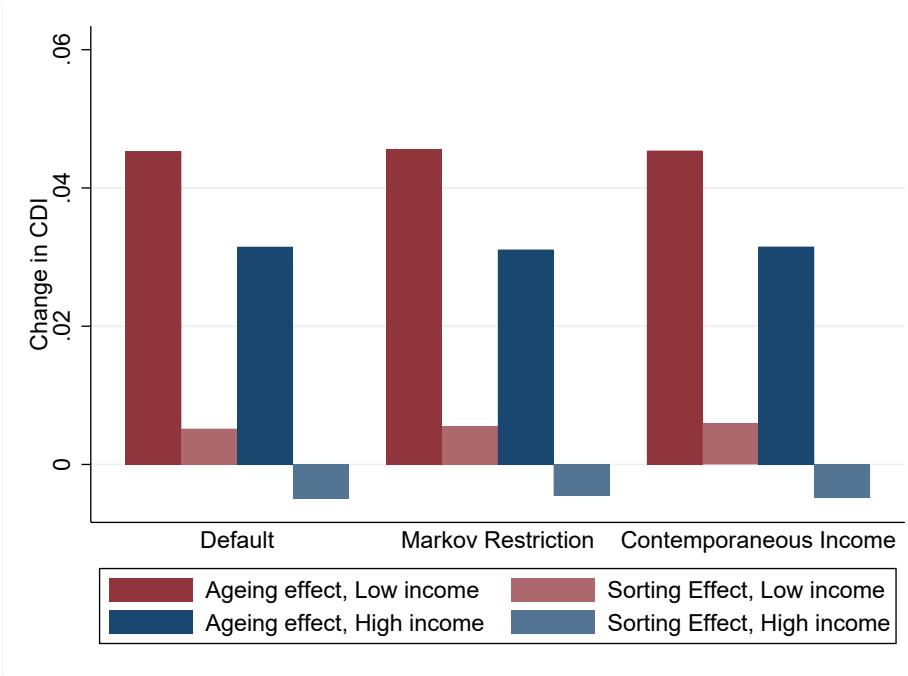
B. Counterfactual biological age



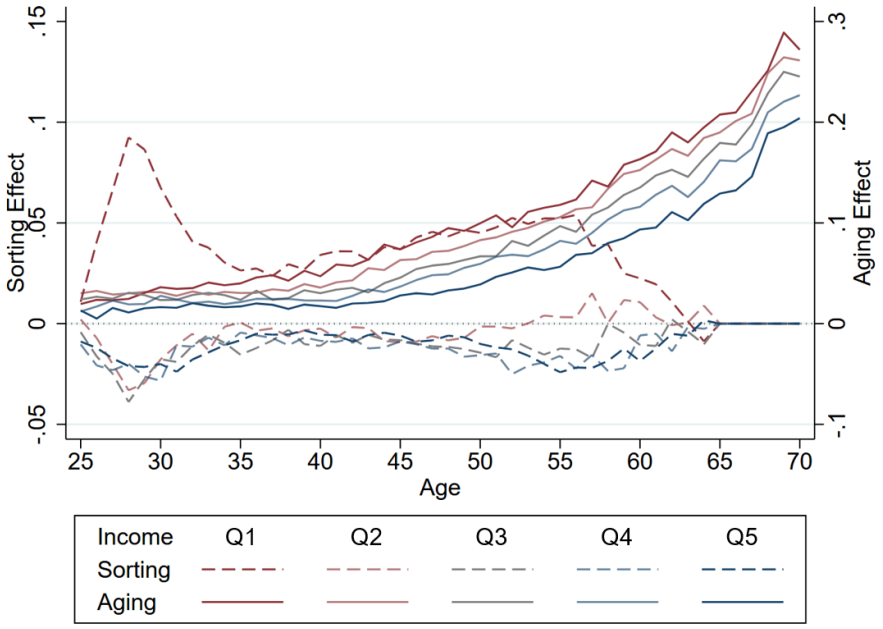
Note: Panel A is equivalent to Panel A of Figure 8, but includes attrition due to death effects. Panel B is equivalent to Figure C.9, with the same attrition adjustment.

Figure G.3: ROBUSTNESS OF SORTING VERSUS AGING EFFECTS

A. Robustness to Income Definition



B. By Income Quintile and Age



Note: Panel A shows cumulative aging and sorting effects for the default decomposition and two alternative income definitions. The 'Markov restriction' robustness check only considers individuals at age a who were in the same income group at a and $a-1$, based on their respective lagged incomes, and repeats the decomposition for this selected sample. The 'contemporaneous income' alternative reduces the income lag by one year, such that the average of Y_{a-3}, Y_{a-2} and Y_{a-1} is used to rank individuals' incomes. Panel B shows the aging and sorting decomposition by income quintile and age. All decompositions pool observations from 2009 to 2021.

Table G.2: SORTING EFFECT BY HOUSEHOLD COMPOSITION, $\times 100$

	11-20	21-30	31-40	41-50	51-60	61-70	Life-cycle Aggregate
1. High income							
Aging	0.04	0.15	0.22	0.37	0.80	1.62	3.18
Sorting, new household composition	-0.07	-0.16	-0.05	-0.10	-0.16	-0.01	-0.55
Sorting, constant household composition	-0.02	-0.11	-0.11	-0.10	-0.18	-0.03	-0.54
2. Low income							
Aging	0.04	0.25	0.36	0.66	1.17	2.08	4.57
Sorting, new household composition	-0.12	0.16	0.26	0.14	0.36	0.18	0.98
Sorting, constant household composition	-0.01	0.07	0.09	0.13	0.17	0.01	0.46
3. Gap							
Aging	0.00	0.10	0.14	0.30	0.37	0.47	1.39
Sorting, new household composition	0.05	0.32	0.30	0.24	0.52	0.19	1.52
Sorting, constant household composition	0.01	0.18	0.20	0.23	0.35	0.03	0.99

Note: The table reports the contribution towards the Chronic Disease Index for the aging and sorting effects. The sorting effects are estimated separately for households which change their composition between $a+1$ and a and those which stay the same. The effects are expressed as the change in the CDI for 10-year age bins, multiplied by 100. That is, the numbers in the table are expressed as percentage points change in the CDI. More detail on the lifecycle decomposition is provided in Appendix Section G.

H Life Expectancy and Lifetime Costs

H.1 Lifetime Income, Transfers, and Healthcare Costs

The counterfactual exercise is motivated by the descriptive gradients documented in the main text. Figure 1 shows the life-expectancy gradient by income, while Panel B of Figure 2 shows the corresponding gradient in chronic disease burden over the lifecycle. These patterns establish that lower-income individuals face both shorter lives and correspondingly higher levels of chronic disease.

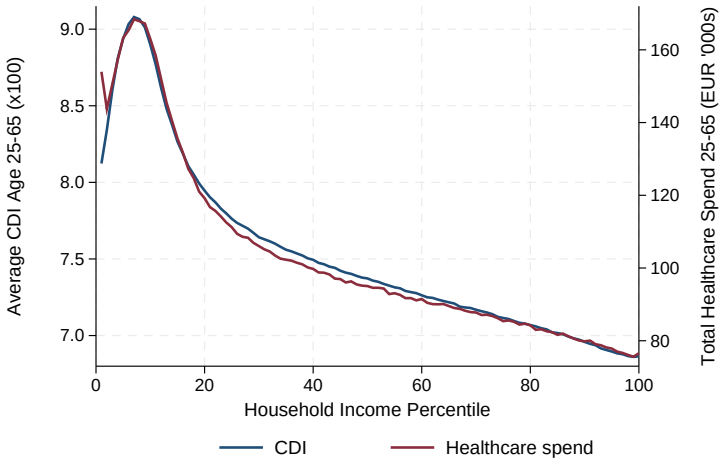
Alongside these health gradients, healthcare spending, household income, and fiscal transfers also vary systematically. Panel A of Figure H.1 orders households by income percentile and reports average CDI and total healthcare spending accrued from ages 25 to 65. Both are highest at the bottom of the income distribution, with a particularly pronounced peak in the lowest decile, and then decline across higher income percentiles. This highlights that the income gradient in chronic disease is accompanied by a gradient in healthcare spending: from the 10th percentile to the 90th percentile, CDI during prime age (25-65) is reduced by 28%, and healthcare spending is reduced by half.

Panel B provides the complementary perspective: how income and social transfers vary by health. Within each gender–age cell, individuals are assigned to CDI percentiles, mean household disposable income and mean total transfers are calculated within each cell, and those percentile-specific means are summed across ages 25 to 65. Disposable income falls and transfers rise with chronic disease burden, especially in the upper tail of the CDI distribution: the top decile contains 60% of the overall income gradient, and 70% of the overall gradient in transfers.

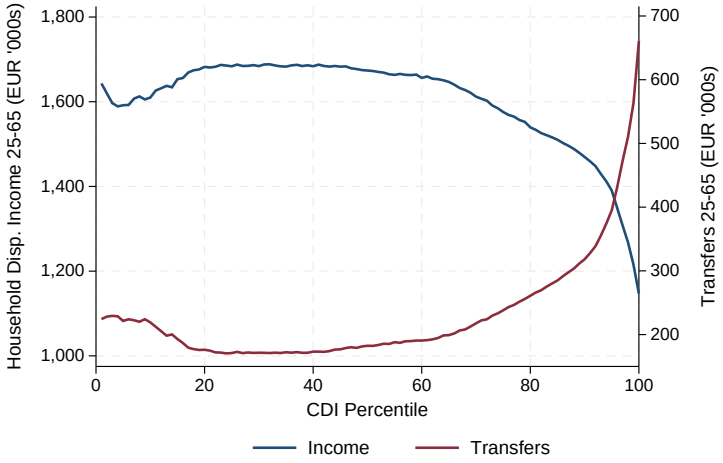
These descriptive gradients motivate a tractable yet informative counterfactual exercise. We isolate the health-accumulation channel by asking how life expectancy and lifetime healthcare costs would change if low-income individuals accumulated chronic disease at the high-income aging rate from age 20, 40, or 60 onwards. Starting from each group's observed CDI at age 20, we construct alternative CDI paths and translate them into age-specific mortality and healthcare-cost profiles using the relationships estimated in the data. This provides a transparent measure of how the timing of slower chronic disease accumulation shapes later-life outcomes. The remainder of this appendix describes this mapping from simulated CDI paths to life expectancy and lifetime healthcare costs.

Figure H.1: LIFETIME INCOME, TRANSFERS, AND HEALTHCARE COSTS

A. Chronic disease burden and healthcare costs by household income percentile



B. Household income and fiscal transfers by CDI percentile



Note: Panel A orders households by income percentile and reports the average CDI and total healthcare spending from ages 25–65. Panel B assigns individuals to CDI percentiles within gender–age cells, calculates mean household disposable income and mean total transfers in each cell, and sums these means across ages 25–65. Values are in thousands of euros where applicable.

H.2 Counterfactual Life Expectancy

In Section 5.4, we estimate how alternative CDI trajectories translate into life expectancy and lifetime healthcare costs. This appendix describes the mapping from simulated CDI paths to age-specific mortality, survival, and costs.

This procedure separates the two roles of mortality weights in the paper. The CDI itself uses fixed age-70 mortality weights to place diagnosed chronic disease burden on a common scale over the lifecycle. The counterfactual life expectancy exercise then allows the mortality consequences of that CDI level to vary by age, using age-specific regressions of mortality on the CDI. In this way we can aggregate the age-specific mortality consequences of chronic disease burden over future ages.

We observe income-specific mortality rates until age 78. Therefore, we run the following age- and gender-specific regressions relating mortality to our Chronic Disease Index:

$$M_{i,a,Y} = \alpha_{a,Y} + \beta_a CDI_{i,a,Y} + \varepsilon_{i,a,Y} \tag{24}$$

where age $a \in [40, 78]$, $M_{i,a,Y}$ is same-year mortality, and $CDI_{i,a,Y}$ is the index for individual i in income group Y, based on lagged chronic conditions. The coefficient β_a is age-specific: it can be small at younger ages, where mortality is rare and less strongly driven by chronic disease, and larger at older ages, where chronic disease burden is more predictive of death. This step converts any specified CDI path into model-predicted age-specific mortality rates. Applying the fitted relationship to observed mean CDI profiles yields fitted mortality rates for those observed profiles; applying it to simulated CDI paths yields baseline or counterfactual mortality predictions.

To construct the counterfactuals shown in Table 3, we simulate alternative CDI paths using the aging and mortality-attrition components estimated in Appendix G. The baseline is itself a simulation rather than the observed CDI age profile: for each income group, we start from its observed mean CDI at age 20 and then cumulate that group's estimated aging and mortality-attrition components, omitting sorting and cohort effects. The counterfactual paths retain this construction but replace the low-income aging component with the high-income component from age 20, 40, or 60 onwards. We apply equation (24) to each simulated path, so the resulting baseline and counterfactual mortality rates are model predictions rather than observed mortality

rates.

The above procedure yields income-specific mortality rates for each counterfactual until age 78 for each alternative. To estimate life expectancy figures, however, we need a full set of same-year mortality rates for income group Y. We estimate the mortality rates at later ages as follows:

- For ages 79 to 90, we use a Gompertz extrapolation to predict mortality. That is, we linearly extrapolate log one-year mortality rates to estimate counterfactual-specific mortality rates between ages 79 and 90. This means that we estimate $\log \tilde{M}_{a,Y} = b_{0,Y} + b_{1,Y}a$, where $\tilde{M}_{a,Y}$ are the one-year mortality rates estimated using equation (24) for ages 40–78. Then, we predict $\log \tilde{M}_{a,Y}$ until age 90.
- For ages between 91 and 110, we rely on the (gender-specific) full-sample one-year mortality rates and set the hazard rate to 1 at age 110.

Once we have a full set of mortality rates, life expectancy at 40 is computed as:

$$\mathbb{E}[A|A \geq 40] \approx \sum_{a=40}^{110} Pr(A = a|A \geq 40) \cdot a \tag{25}$$

where A is age at death. Using this framework, we first compute baseline life expectancy for both income groups from the simulated baseline CDI paths described above.

The resulting life expectancy at age 40 is reported in Table 3. Panel A of Figure H.2 visually shows the survival probabilities for the high and low income baseline, and the counterfactual applying high income aging effects from age 20 onwards.

H.3 Counterfactual Healthcare Costs

Apart from life expectancy estimations, we also calculate counterfactual lifetime healthcare costs. First, we start with the following version of equation (24) :

$$k_{i,a,Y} = \gamma_{a,Y} + \delta_a CDI_{i,a,Y} + u_{i,a,Y} \tag{26}$$

where age $a \in [40, 70]$ and $k_{i,a,Y}$ is logged, detrended healthcare costs for individual i who belongs to income group Y at age a.

Then, we use the same counterfactual CDI evolutions described in Section H.2 above to estimate healthcare costs at each age between 40 and 70. Again, a baseline CDI evolution for each income group using the respective aging and attrition due to death effects is used, along with counterfactuals which apply high-income aging and death effects to the low-income CDI from age 20, 40 and 60 onwards.

We then estimate costs at later ages as follows:

- Between 71 and 90, we calculate yearly costs in two steps. First, we impose the empirical high- and low-income costs to grow at the same rate as the full population costs. Then, we compute a weighted average of both, with linearly increasing weights on the full population costs. This procedure is visually represented in Panel B of Figure H.2 for the high and low income baselines.
- Between 91 and 110, we revert to the overall cost rates $\bar{k}^a$ (not income specific), computed on the full sample, for all sets of costs.

Once we have a full set of mortality rates, lifetime expected cost at age 40 is computed as:

$$\mathbb{E}\left[\sum_{a=40}^{\infty} K_a \mid A \geq 40\right] \approx \sum_{a=40}^{110} S_a \cdot K_a \tag{27}$$

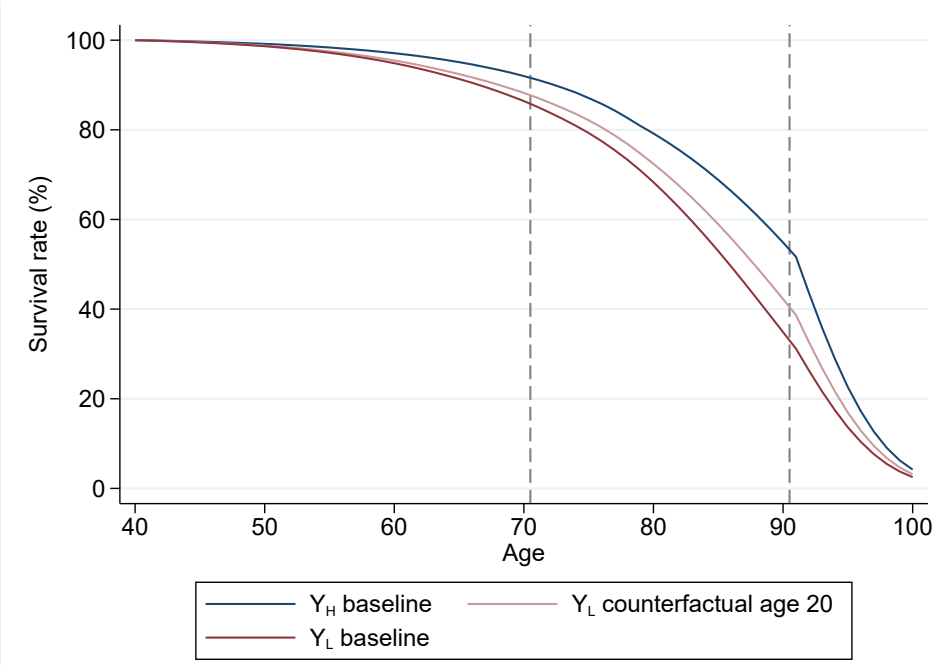
Where A is age at death and S_a is the survival probability, $S_a := Pr(A \geq a|A \geq 40) = S(a|A \geq 40)$.

We use two alternative approaches with respect to the survival probabilities in our counterfactuals. In the first approach, we compute counterfactual healthcare costs when intervening at age 20, 40 or 60, but use the baseline low income survival probabilities computed in the life expectancy calculations. That is, we allow costs to be affected but not survival probabilities by the hypothetical intervention. This approach corresponds to row 3.a in Table 3.

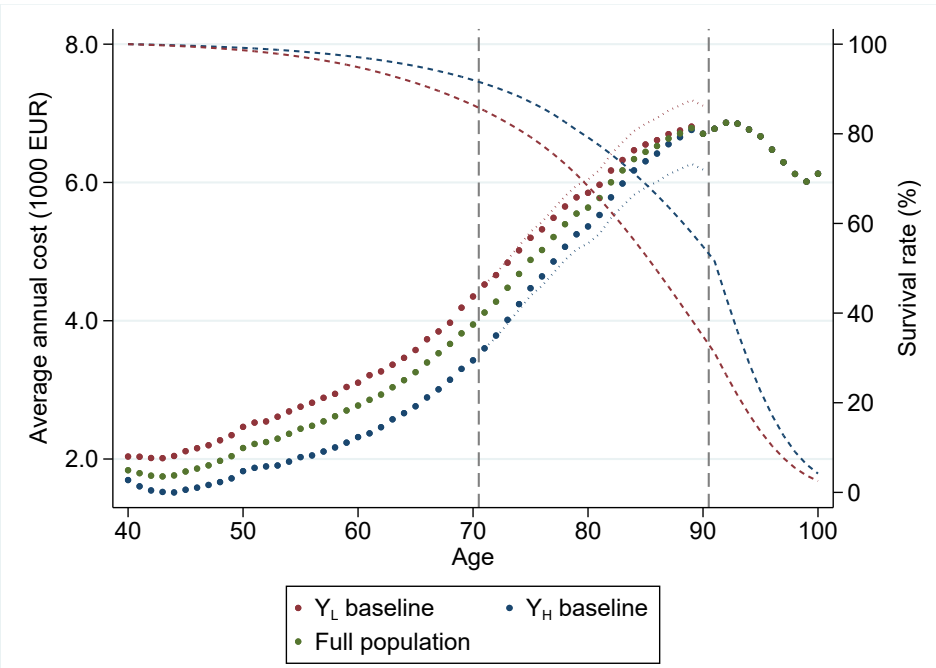
In the second approach, survival probabilities are adjusted for each counterfactual. More specifically, they are taken from the corresponding alternative estimated in the life expectancy calculations, using equation (25). That is, we allow both costs and survival probabilities to be affected by the hypothetical intervention. This approach corresponds to row 3.b in Table 3.

Figure H.2: SURVIVAL RATES AND COSTS OVER THE LIFECYCLE

A. Survival curves, women



B. Weighted costs, women



Note: This figure illustrates the life-expectancy and lifetime-cost calculations for women. Panel A displays model-implied survival curves for the simulated baseline CDI paths of both income groups and for the counterfactual that applies the high-income aging component to the low-income path from age 20. Between ages 40 and 78, one-year mortality rates are predicted from the CDI model. Between ages 79 and 90, they are extrapolated using a Gompertz specification; between ages 91 and 110, observed gender-specific full-sample mortality rates are assigned to each group. Panel B shows model-predicted annual healthcare costs for the same simulated CDI paths between ages 40 and 70. Between ages 71 and 90, income-specific costs are initially grown at the full-sample rate and then combined with full-sample costs using weights that increase with age. Above age 90, full-sample costs are applied to all individuals.

H.4 Alternative Estimates

In our counterfactual analysis, we use simulated CDIs based on aging effects. We can also apply the methodology described in Sections H.2 and H.3 to estimate life expectancy and expected lifetime costs using average CDIs for both income groups. Appendix Table H.1 reports those estimates and shows how they change when we assign high-income CDIs or survival rates to the low-income group.

Table H.1: ESTIMATES USING AVERAGE CDIS

	High Income	Low Income		
	Baseline	Baseline	Y_H Survival	Y_H CDI
1. Life Expectancy	85.2	80.8	85.2	84.3
2. Lifetime Costs	159.0 k	163.9k	171.4k	147.8k

Note: This table shows additional life expectancy and lifetime cost estimations. The first two columns use CDI averages by age to estimate life expectancy and expected lifetime costs. The third column assigns the observed high-income survival rates to the low-income group. Column 4 assigns high-income CDI averages to the low-income group. Each alternative estimate applies the methodology described in Appendix H to estimate costs and survival rates at higher ages.