

# The Economic Value of Eliminating Diseases

Julio Crego, Daniel Karpati, Jens Kvaerner, and Luc Renneboog\*

February 1, 2023

## Abstract

We estimate the causal effects of 334 different types of health shocks on medical expenses, mortality, disability, labor market participation, labor earnings, and the need for nursing home care using detailed data on 6.9 million people diagnosed by medical specialists between 2013 and 2017. We quantify the benefits of eliminating diseases with distinct consequences for people of different social strata by incorporating the estimates into a standard life-cycle model. Our results reveal substantial heterogeneity in welfare gains by types of disease for different people. We discuss the potential implications of our results for the financing of medical research.

**JEL Classifications:** D15, D63, G50, I10, H51

**Keywords:** health shocks, household finance, life cycle, health economics, public health expenditure, labor earnings, labor participation, mortality, medical expenses

---

\*Crego: Tilburg University, Warandelaan 2 5037 AB Tilburg Netherlands; J.A.Crego@tilburguniversity.edu. Kvaerner: Tilburg University, Warandelaan 2 5037 AB Tilburg Netherlands; jkverner@gmail.com. Karpati: Tilburg University, Warandelaan 2 5037 AB Tilburg Netherlands; d.karpati@uvt.nl. Renneboog: Tilburg University, Warandelaan 2 5037 AB Tilburg Netherlands; Luc.Renneboog@tilburguniversity.edu. Results are based on calculations by Daniel Karpati of Tilburg University using non-public microdata from Statistics Netherlands. The research for this publication was partly funded by the Open Data Infrastructure for Social Science and Economic Innovations (ODISSEI) in the Netherlands ([www.odissei-data.nl](http://www.odissei-data.nl)).

# 1 Introduction

Recent breakthroughs in life sciences and medical engineering have greatly contributed to the development of new medical treatments. For example, drugs based on or engineered from biological tissue, are increasingly present in most therapeutic areas, including oncology, rheumatology, diabetes, hematology, neurology, and infectious diseases (Evens and Kaitin, 2015). While new treatments can increase life expectancy and improve quality of life, they often come at very high development costs.<sup>1</sup> This begs the question how to prioritize research funding: if we had the choice, which diseases should we eliminate?

Based on administrative data of 18.5 million distinct medical diagnoses, we show that the answer to this question depends on whom you ask. Our central insight is that people of distinct socioeconomic strata would benefit differently (in terms of labor participation, earnings, mortality, medical costs, or the use of a nursing home) from eliminating a wide range of health shocks. As a result, people's willingness to pay for hypothetical 'vaccines' (medication and treatments) that would eliminate a specific health shock varies substantially across the population. This fact transforms target medical research into a policy-related - and even political - question. Increasing the funding targeted at diseases that cause disability mainly benefits low-income individuals; meanwhile, funding targeted to increase life expectancy tends to benefit more high-income individuals.

To give an example, consider the decision to fund research to eliminate lung cancer or herniated intervertebral disk pain (HNP). These diagnoses have different implications for a specific individual's health and labor market outcomes, and, at the same time, these implications depend on socioeconomic status. Our estimates show that low-income men would pay EUR 198 to eliminate HNP; thereby reducing their risk of becoming disabled. They would pay 2.2 times more (EUR 438) to eliminate lung cancer, which would increase their life expectancy. Quite strikingly, high-income men would pay fifty times more (EUR

---

<sup>1</sup>Between the 1990s and the early 2010s, the total R&D costs of new drugs has been increasing at an annual rate of 8.5% above inflation (DiMasi et al., 2016).

14,866) to eliminate lung cancer than to eliminate HNP (EUR 299). This is because they barely suffer from disability risks. From the government’s perspective, our results are even more striking. We find that the government could spend about EUR 255,000 and EUR 26,000 per born individual to eliminate lung cancer and HNP, respectively, while holding welfare at its current level.

Taken together, our results suggest that public and private medical research funding would favor treatments with high potential benefits for the wealthiest. Given the lower willingness to pay to eliminate diseases such as lung cancer by low-income households, it is unlikely that even a very progressive tax system would justify using public funds to finance the development of new treatments against such diseases.

Our willingness-to-pay estimates—reflecting the present value of the total costs of a particular health shock measured in terms of labor market outcomes and long-term health—come from a two-step procedure. The first step involves estimating the effects of a given health shock on medical expenses, mortality, disability, participation in the labor market, labor earnings, and the need for a nursing home. In the second step, we incorporate the causal effect estimates into a life-cycle model. We use the model to calculate the willingness to pay for a “counterfactual scenario” in which we eliminate one diagnosis at the time.

Our reduced-form estimates contain five independent results. Starting with the average effects (those based on all individuals), we report a range that covers the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the estimated effects. First, medical expenses range from a few hundred euros to more than EUR 18,000.<sup>2</sup> Second, excess mortality reveals a similar spread: it ranges from a practically zero increase in the probability of dying during the three years subsequent to being diagnosed to more than 23%.<sup>3</sup> Third, the average effect of a health shock on labor market participation and dependence on disability insurance ranges from zero to approximately 15%.

---

<sup>2</sup>For example, some of the most expensive diagnoses relate to blood cancers (leukemias, lymphomas), the treatment of which may require expensive anticancer drugs or stemcell transplants, or to subarachnoid hemorrhage (an uncommon type of stroke caused by bleeding on the surface of the brain), where treatment often requires surgery.

<sup>3</sup>Some of the fatal health shocks include pleural mesothelioma (rare cancer caused by asbestos exposure) and malignant cancers of the digestive organs (e.g., pancreatic cancer).

Fourth, conditional on being able to work, the effect on earnings is smaller and less dispersed. Fifth, the probability of moving to a nursing home (above the age of 65) within three years after a diagnosis is between 0 and 7%. Overall, we find substantial cross-sectional variation in the effect of adverse health shocks on health, medical expenses, and labor market outcomes.

We also find substantial heterogeneity in how a specific diagnosis affects people of distinct social strata differently. For example, HNP causes a 10% decrease in the labor participation of people with low permanent income. In comparison, the effect is only 3% for people with high income. Unlike the uneven effect of HNP, lung cancer is similarly severe for everyone.

Until now, the lack of sufficiently rich medical register data has prevented such a study. The necessary data requirements are high. First, one needs precise identification and measurement of a set of health shocks. Survey data, such as the Health and Retirement Study (HRS) in the United States, or the Survey of Health, Aging and Retirement (SHARE) in Europe, only comprise self-reported health outcomes with limited details about specific diagnoses.<sup>4</sup> Second, to examine heterogeneity in the effects of specific health shocks over the life cycle and across the income distribution one needs data on many more individual diagnoses than what is available in surveys. Our dataset satisfies these requirements.

In addition to the rich medical register data available in the Netherlands, studying the Dutch health care system has further advantages. First, the cornerstone of the Dutch institutional setting is the principle of egalitarianism. In the Netherlands, all citizens, irrespective of income and wealth, pay a moderate flat-rate health insurance premium (about EUR 1164 annually in 2015). Individuals who earn less than about 70% of the mean income are eligible for a means-tested subsidy that covers up to 90% of their health insurance premium. Consequently, all citizens including low-income individuals are covered by health insurance and, beyond the mandatory insurance premium and a low deductible, the health care system is essentially free for any citizen. Second, each patient receives the same medical treatment, irrespective of income or wealth. Thus, our research based on data from the Netherlands

---

<sup>4</sup>For example, the survey contains information about whether an individual is diagnosed with “any cancer” rather than a particular type of cancer.

(which is in this aspect prototypical for most countries in the European Union and Switzerland) has the advantage that we can exclude some confounding factors when measuring the economic outcomes of exposure to disease. In our empirical strategy, we can hence ignore that some people remain uninsured or are insufficiently insured when they cannot afford to pay health insurance premiums. We can also ignore the possibility that medical treatments and costs differ between public and private medical institutions, which is the case in some countries such as the U.S.

Our estimates are relevant for health investors and policymakers in countries with both private health care and egalitarian public health systems. Even in countries with an egalitarian public health system (where the direct medical health costs are equal for each patient), health shocks can have substantially different consequences by patient type in terms of future health outcomes and economic effects. Let us illustrate this with an example: appendicitis versus herniated intervertebral disk pain (HNP). The former health shock affects most patients in the same way, and after a recovery period of a few weeks subsequent to surgery patients can usually resume their labor activities. This may not be the case for HNP as losing the ability to move freely and without pain may force a blue-collar worker to exit the labor market, but not, say, an architect. As such, there are good reasons to expect that socioeconomic status moderates or aggravates the effects of specific health shocks. Regardless of the source of heterogeneity, whether it originates from the spectrum of diseases or variation in socioeconomic characteristics, the priorities chosen by policymakers and investors in medical research and health care affects specific social groups of patients differently.

Our paper is related to several strands of literature. First, our reduced form analysis follows the literature that estimates the causal effects of events that are unpredictable within a short period. A naive comparison of treated and non-treated individuals is unlikely to uncover a causal effect because, in our case, many variables that predict exposure to specific diseases also predict the outcomes we study. As a result, the outcome of diagnosed and non-diagnosed individuals is different regardless of health shock. We address this empirical

challenge following (Dobkin et al., 2018; Fadlon and Nielsen, 2019; Doeskeland and Kvaerner, 2022).<sup>5</sup> The idea is to examine the difference in health and economic outcomes between individuals diagnosed with a specific disease today and those who experience the same health shock some years later in order to identify causal effects. For instance, if the occurrence of a specific disease such as lung cancer depends on (latent) life-style factors (e.g. smoking, exercise, eating habits), we control for these life-style factors by considering as a control sample those people who will get lung cancer but only with a delay relative to the treated individuals. The idea is that if life-style affects a disease, the time of occurrence of the disease cannot be accurately predicted such that at the time of the health shock to the people in the treated sample, those in the control sample still live in bliss ignorance of what is awaiting.

Second, we contribute to the literature that studies the impact of health on individuals' saving decisions. The focus of literature (e.g. DeNardi et al., 2010) lies on retired individuals, because they are the most affected by medical expenses. The main objective of this (U.S.) literature has been to assess the consequences of changing policies related to health care and insurance. Koijen et al. (2016) extend the focus on the retired individual to a person's entire life cycle. They leverage the age range in the HRS database to model a person's life-cycle from the age of 51 (when the survey starts) in order to construct optimal portfolios of a range of health insurance products over this life cycle. De Nardi et al. (2017) exploit the PSID and HRS databases to model individuals' health over their life cycles to examine the impact of health on labor income and participation. They show that these economic outcomes are critical to measuring the true cost of poor health. Our main contribution to this literature is to use diagnosis-specific shocks together with causal estimates of the effect of each diagnosis, enabling us to study counterfactual scenarios in which we eliminate a specific health risk (or

---

<sup>5</sup>Dobkin et al. (2018) exploit hospitalization data in the U.S. to analyze the consequences of being insured and the impact of health on bankruptcy. Fadlon and Nielsen (2021) document that fatal events make the surviving spouse work more, with the greatest effects for families who experience significant income losses. Doeskeland and Kvaerner (2022) study the personal investment decisions of 60,000 households after a cancer diagnosis. They find that a cancer diagnosis reduces the willingness of households to take risks with their financial wealth; particularly cancers with a large impact on life expectancy and income. Karpati (2022) exploits the setting of predictive genetic tests to show that households' accumulation of financial wealth reacts negatively to news about increased cancer risks and mortality.

a correlated set of health risks).

Third, we add to the literature by investigating whether socioeconomic status moderates the effect of health shocks on individual health and labor market outcomes. [Heinesen and Kolodziejczyk \(2013\)](#) find that the negative effects of breast and colorectal cancer on labor force participation are stronger for low-educated individuals. [García-Gómez et al. \(2013\)](#) study all types of acute hospitalizations and estimate that individuals from low-income households experience a greater relative loss of labor income following hospitalization. [Lundborg et al. \(2015\)](#) also study the effects of acute hospitalizations and report a greater negative impact on the labour earnings of individuals with a low level of education. Our paper contributes to this literature by presenting evidence on the heterogeneous effects of a broad range of health shocks, including both inpatient and outpatient diagnoses. We also study treatment heterogeneity in a comprehensive set of health outcomes, such as mortality, health costs, and nursing home use. Moreover, our study goes beyond labor market outcomes as we also estimate the heterogeneous effects of health shocks on elderly individuals.

Finally, our study extends the medical literature, which has extensively studied the costs of bad health (cost-of-illness (COI) studies) to guide medical research and policy design. These studies usually focus on one specific disease and use different estimation techniques, outcomes, and discounting, which makes it difficult to compare their conclusions ([Byford et al., 2000](#); [Larg and Moss, 2011](#)). Our study is not subject to this critique because we estimate the effects for all diseases using the same empirical approach, institutional setting, and data sources.

The remainder of the paper is organized as follows. Section 2 presents the data and institutional details, including variable definitions and summary statistics. Subsequently, we explain the reduced-form research design. Section 4 contains the reduced-for results. We present the life-cycle model and quantitative analysis in Section 5.

## 2 Setting and Data

We use administrative data from the Netherlands to estimate the effect of a broad set of health shocks on a variety of economic outcomes. In this section, we explain the institutional setting, define variables, and present summary statistics.

### 2.1 The Institutional Setting

The Dutch institutional setting offers several advantages to our study. First, universal health-care coverage ensures that all people, regardless of socioeconomic status, receive the same treatment for the same diagnosis at the same (insured) cost. Standardized treatment protocols make it possible to study if people from different social strata (e.g. permanent income groups) respond differently (in terms of participation to the labor market, savings behavior, etc.) to the same diagnosis. Consequently, we can use our estimates to analyze the benefits of curing diseases (or in other words, the negative consequences of suffering from diseases in terms of forgone welfare and wealth following the - possibly temporary - inability to work, the reduced wages, declining ability to save etc.) without taking a stand on the design of the health insurance system.

Second, the universal healthcare system generates and stores detailed data on medical diagnoses and expenses. We observe all medical diagnoses made by medical specialists, including diagnoses made not only in inpatient settings but also in outpatient ones. This is important because about  $3/4^{th}$  of all diagnoses are made in an outpatient setting. We also observe all expenses that fall under the comprehensive standard health insurance package. This package covers all necessary medical care, including general practitioner (GP) care, maternity care, hospital care, home nursing care, pharmaceutical care, and mental healthcare. Moreover, we also have information on nursing home care use, which is important because long-term care accounts for more than 20% of all Dutch healthcare expenses.



## 2.2 Data Sources and Variable Definitions

In this sub-section, we summarize our main data sources and define our key variables. The Internet Appendix provides additional details.

### 2.2.1 The Spectrum of Medical Diagnoses

Statistics Netherlands (SN) collects data on diagnoses by medical specialists from the Dutch Healthcare Authority (DHA). Specialist healthcare providers, including hospitals (both in-patient and outpatient care), independent treatment centers, specialized institutions (e.g. epilepsy clinics, cancer clinics), and rehabilitation, dialysis, audiological and radiotherapeutic centers, have a legal obligation to supply information on the care provided to the DHA. Diagnoses are classified using the Diagnosis Treatment Combinations (DTC) classification scheme. The DTC is an adapted type of diagnosis-related group (DRG) system, which Dutch healthcare insurers use to pay for specialist care. The system is based on 27 different medical specialties and numerous medical diagnoses within each specialty.

At any given time, there are approximately 2500 valid DTC diagnoses (speciality-diagnosis combinations). These diagnoses are not automatically comparable to diagnoses in international classification schemes such as the International Classification of Diseases (ICD-10).

We now explain how we group DTC diagnosis codes into medically consistent groups of 334 diagnoses with a good correspondence to the ICD-10 classification.<sup>6</sup> First, we exclude all DTC diagnoses by medical specialties where patients are mostly referred to by other medical specialties (e.g., rehabilitation, anesthesia, radiotherapy) in order to avoid double counting. We also exclude codes that relate to voluntary treatments such as pregnancy-related care (beyond the basic number of consultations and echographies) or cosmetic plastic surgery. Second, we classify the remaining 1761 DTC diagnosis codes into diagnosis groups using

---

<sup>6</sup>Diagnoses can cover multiple ICD-10 codes. In general, we refer to a diagnosis by the most frequently diagnosed ICD-10 code within the diagnosis. For example diagnosis 162 contains three different Dutch DTC codes which can be mapped to the ICD codes I20 (angina), I24 (other acute ischemic heart diseases), and I25 (chronic ischemic heart disease). We refer to this diagnosis as 'angina' because I20 represents about 70% of the cases.

a correspondence table provided by the Dutch Healthcare Authority. We exclude diagnosis groups with ICD codes related to congenital disorders (ICD block Q), signs/symptoms (block R), “factors influencing health status and contact with health services” (block Z), and pregnancy/childbirth (block O). Starting from 2016, our dataset also records ICD-10 codes besides the Dutch DTC codes. We use these codes to split some of the diagnosis groups into subgroups that contain more homogeneous ICD-10 codes. Finally, we identify ten pairs of diagnoses that relatively frequently occur together within the same year.<sup>7</sup> We treat these pairs as separate diagnoses. These steps result in 334 diagnosis groups that had at least 200 people diagnosed annually during the sample period.<sup>8</sup>

### 2.2.2 The Health Shock

We focus on health events that are new diagnoses for an individual and thus exclude follow-up care, in line with [Dobkin et al. \(2018\)](#). We label these events as “health shocks”. If a person receives the same diagnosis twice during our sample period, we only consider the first of these two diagnoses. If a person receives multiple diagnoses of different pathologies during the sample period - we observe an average of 2.7 diagnoses per person, we consider these diagnoses independently.

### 2.2.3 Medical Expenses

The Netherlands has a flat-rate insurance premium for the mandatory standard health insurance package. This package covers all necessary medical care, including general practitioner care, maternity care, hospital care, home nursing care, pharmaceutical care, and mental healthcare. Insurers are not allowed to refuse an applicant. The flat-rate premium and the acceptance obligation prevents insurers from risk selection. Insurers with a relatively risky pool of insured individuals receive additional compensation via a centralized system. A di-

---

<sup>7</sup>An example is heart failure and atrial fibrillation: Among the 88,000 people diagnosed with any of these two diagnoses in 2016, about 3% received both diagnoses.

<sup>8</sup>The Internet Appendix presents the formation of diagnosis groups in detail. It also contains a list of the 334 diagnosis groups and their corresponding ICD-10 codes.

rect implication of the health insurance system is the need to store data on medical expenses and Statistics Netherlands uses these data at the level of the individual to determine the annual aggregate medical expenses.

Medical expenses in the Netherlands, as a percentage of the GDP, are similar to expenses in other OECD countries, with the exception of the United States. In the US, higher costs of labor, goods, and services result in more expensive treatment and higher healthcare spending (Papanicolas et al., 2018). For example, McGuire et al. (2015) estimated that the treatment costs of the most common type of lung cancer in 2007-2008 amounted to about EUR 25,000, 18,000, and 32,000 in France, England, and Germany, respectively. We estimate a cost of about EUR 28,000 in the Netherlands in 2015. In comparison, in the early 2000s, the costs of treating lung cancer in the US was about EUR 43,000 (Kutikova et al., 2005).

#### 2.2.4 The Cost of Nursing Home Care

Although we cannot directly observe the costs of nursing home care at the individual level, we collect an indicator capturing if an individual is living in an “institutional household”, which is de facto a nursing home.<sup>9</sup> We estimate the causal effects of each medical diagnosis on the probability of living in an institutional household/nursing home. We arrive at an estimate of the nursing home care expenditures associated with each diagnosis by multiplying this estimated probability by the average per capita nursing home expenditures in 2015 (EUR 43,545). We further split this expenditure estimate into a personal contribution part paid by the diagnosed individual and a part that is subsidized by the government. We estimate the personal contribution based on the formula applied by the Dutch authority that is responsible for collecting these personal contributions (CAK). Low-income individuals pay 30% of the nursing home costs while high-income individuals pay up to 53%.<sup>10</sup>

---

<sup>9</sup>In rare cases an institutional household can also refer to another type of care institution (e.g., psychiatric hospital, addiction treatment center).

<sup>10</sup>The formula takes into account the individual’s disposable household income, employment income, age, marital status, wealth, among other characteristics.

### **2.2.5 Earnings and Permanent Income**

We collect data on the participation in the labor market and the earnings of these participants. The data begin in 1999 and cover all Dutch individuals. We use two variables to measure labor market outcomes. The first one measures the extensive margin decision, which we label “labor market participation”. A labor market participant is an individual classified as an “employee” in the annual socioeconomic classification of Statistics Netherlands. In this classification, the source of the highest annual income (e.g., labor income, pension benefits, other benefits) determines an individual’s socioeconomic category (employee, pensioner, other). The logarithm of labor earnings is our second measure of labor market participation, which captures the intensive margin decision. We also use labor earnings to classify the Dutch population into permanent income terciles. If data on labor earnings are not available or the labor earnings are significantly lower than the pension income, we use data on pensions (which are available from 2011). The Internet Appendix provides a detailed explanation of the classification of individuals into permanent income terciles.

### **2.2.6 Disability Insurance**

The Dutch disability insurance scheme covers the loss of income from both work-related and non-work-related sickness/injury. During the first two years of an employee’s illness, the employer is required to continue paying the salary. In the third year, the employee can apply for the public disability benefit. A medical doctor and a labor expert, working for the public Employee Insurance Agency, decide on the disability application, and if approved, they also determine the size of the compensation.

## **2.3 Descriptive Statistics**

### **2.3.1 Main Sample**

Individuals in our main sample are at least 25 years old at the time of the diagnosis. We exclude young individuals because they face different economic circumstances, may still be dependent on their parents, and are covered by a distinct disability insurance scheme. We use subsets of the main sample to study different health and economic outcomes subsequent to the diagnosis. For the analysis on mortality risk and medical expenditures, we use all observations. For the analysis of labor market outcomes, we restrict the sample to workers, defined as individuals under 65 years of age with the so-called socioeconomic “employee” status in SN prior to the diagnosis. Individuals classified as employees have the same public disability insurance and similar employee rights after a negative health shock. For the analysis of nursing homes, we include only individuals who are at least 65 years at the time of diagnosis. This age group comprises most users of nursing homes. The main sample contains 6.9 million unique individuals diagnosed at least once between 2013 and 2017. Approximately 4.6 million individuals fall into the sample of workers and 2.5 million individuals into the sample of potential nursing home users.

### **2.3.2 Summary Statistics**

As we explain in the next section, our identification strategy is to compare the outcome of individuals experiencing a health shock today with individuals experiencing the same health shock a few years later. Thus, the key identifying assumption is that these individuals are identical prior to diagnosis; this identification strategy controls for latent behavioral variables such as life style which could be related to future health shocks. As a first test of the validity of this assumption, we present summary statistics in 2012 for two groups diagnosed in subsequent years. The first group, the treatment group, experiences a health shock during 2013. The second group, the control group, experiences a health shock in 2014.

The upper panel of Table 1 shows the number of individual-diagnosis observations, the mean, and the standard deviation for both groups. The last column reports the standardized mean difference (SMD), a statistic that we use to examine the balance of the covariate distribution between the two groups.<sup>11</sup> The lower panel of Table 1 presents the 25th, 50th, and 75th percentiles for age, medical expenses, net wealth, financial assets, and income for the treatment and control groups, respectively.

[Insert Table 1 here]

The average individual is 59 years old.<sup>12</sup> The gender distribution is balanced, and more than one in five individuals obtained a college degree or higher. The average medical expenses per diagnosis amount to about EUR 4,200. Net wealth ranges from EUR 1,000 to approximately EUR 200,000 (25th to 75th percentile), with a median of EUR 42,700. The average income is EUR 37,800 and about 70 percent of the individuals are working. It is reassuring to see that all standardized covariate mean differences (SMD) are well below the rule of thumb, 0.1 (Branson, 2018).

### 3 Empirical Design

We estimate the causal effects of health shocks. Our key empirical challenge is to construct a counterfactual outcome for a diagnosed individual. A naive cross-sectional comparison of diagnosed and non-diagnosed individuals is unlikely to recover causal effects because many variables that predict exposure to certain diseases might also predict the outcomes we study (e.g., participation in the labor market). As a result, the average outcomes of diagnosed and non-diagnosed individuals may be different regardless of the health shock.

---

<sup>11</sup>The SMD is calculated as the difference in means between the treatment and the control groups scaled by the square root of the sum of the group variances. As the SMD is independent of the measurement unit, it is possible to compare variables with different measurement units.

<sup>12</sup>We refer to individuals in this section but it should be kept in mind that the summary statistics are for individual-diagnosis observations. Consequently, individual characteristics are weighted by the number of diagnoses a person has received in the sample period.

We solve the empirical challenge following [Dobkin et al. \(2018\)](#); [Fadlon and Nielsen \(2019\)](#); [Doeskeland and Kvaerner \(2022\)](#). The main idea is to estimate the causal effect of the difference in the outcome between similar individuals, whereby we define similar individuals as those who are diagnosed with the same disease within a 5-year window after controlling for individual fixed effects and year-by-birth cohort fixed effects. The identifying assumption is that conditional on being diagnosed within the next five years, and individual fixed effects and year-by-birth cohort fixed effects, the assignment to the year of the diagnosis is as good as random. There is support for this identifying assumption in the medical literature. Many studies show that it is notoriously difficult to predict whether a specific individual will be diagnosed with a particular disease using statistical models. Some individuals in the low-risk group, or unexposed category, will develop the disease; while the majority of those exposed will remain healthy ([Rockhill et al., 2000](#)). Thus, making predictions about the exact date of a diagnosis is even more difficult.

The research design controls nonparametrically for latent variables that could be correlated both with the treatment status (diagnosis) and the outcome variable. As a consequence, if, say, hernia is preceded by a gradual deterioration in health, which correlates with participation in the labor market, these effects cancel out.

To formalize the research design, let  $i$  index individuals and let  $t$  represent calendar time. We denote the years since the diagnosis year ( $t_i^{diag}$ ) by  $K_{i,t} = t - t_i^{diag}$ . We estimate the effect in  $K_{i,t} \in \{-4, -3, -2\}$  leads of treatment, and  $K_{i,t} \in \{0, 1, 2, 3, 4, 5\}$  lags. The reasons we set the effect at  $K = -5$  and  $K = -1$  to 0 is because we need at least two restrictions for point identification, as discussed formally in [Borusyak and Jaravel \(2017\)](#). For each diagnosis, we restrict the sample to individuals diagnosed in any of the sample years. We estimate the following event-study regression for the 334 different health shocks:

$$y_{i,t} = \alpha_i + \delta_{c(i),t} + \sum_{k=-4, k \neq -1}^5 \gamma_k \{K_{i,t} = k\} + \epsilon_{i,t}. \quad (1)$$

where  $\alpha_i$  stands for individual fixed-effects,  $\delta_{c(i),t}$  represents the cohort-year fixed effects, and  $\gamma_k$  corresponds to the causal effect of the disease  $k$  years after the diagnosis.  $y_{i,t}$  is one of the five different outcome variables: (1) an indicator of whether an individual is classified as working, (2) the logarithm of earnings conditional on labor participation, (3) an indicator of whether an individual is fully disabled, (4) annual medical expenses in euros, and (5) an indicator of whether the individual lives in a nursing home.

For each health shock, we investigate whether individuals who belong to different social strata respond differently to the shock by estimating the above specification in sub-samples. We use sub-sample analysis instead of augmenting the event-study regression with interaction terms given our large sample size. The benefit of sub-sample analysis is that it allows for distinct cohort-year fixed effects in different sub-samples.

We consider slightly different sub-samples in the reduced-form analysis reported in Section 4 and in the analysis that serves as input for our structural model in Section 5. For the model, in principle, we consider all 18 sub-samples formed by the Cartesian product of the two genders, three permanent income groups, and three age-at-diagnosis groups (25 to 44, 45 to 64, and 65+). For the reduced form estimates presented in Table 2, we only consider “univariate” sample splits for illustrative purposes, i.e., we study two sub-samples formed by gender, three by the three permanent income groups, and three by the age at diagnosis. For certain outcomes we do not consider all sub-samples. For the labor market outcomes (labor participation, log labor earnings, full disability), we only consider working-age individuals. This entails that we use only the corresponding 12 sub-samples in the analysis for the model, and that we exclude 65+ individuals from the univariate sub-samples in the reduced form analysis. In contrast, for nursing home use, we only consider 65+ individuals; we thus use only the corresponding 6 sub-samples for the model estimates and exclude the individuals with an age below 65 from the univariate sub-samples in the reduced form analysis.

Estimating the effect of a health shock on mortality requires a different specification. This is because the control group, those diagnosed later, is - by definition in this empirical



setup - still alive in all periods prior to diagnosis. Our choice is a multivariate regression with an indicator of dying in 2016 as the dependent variable:

$$\Delta Dead_{i,2016} = \alpha + D'_{i,2016}\beta_0 + D'_{i,2015}\beta_1 + D'_{i,2014}\beta_2 + D'_{i,2013}\beta_3 + F(age_{i,t}) + \epsilon_{i,t}. \quad (2)$$

Here  $\Delta Dead_{i,2016}$  is a binary indicator if person  $i$ , who was alive on 1 January 2016, died during 2016.  $D_{i,2016} \dots D_{i,2013}$  are vectors of 334 binary indicators of having been diagnosed with the corresponding medical illness (out of the 334 diseases) in 2016, 2015, 2014, and 2013, respectively.  $\beta_0 \dots \beta_3$  are column vectors of coefficients.  $F(age_{i,t})$  is a polynomial of age of the fourth degree.

Because all individuals in the sample of Equation 2 are alive on 1 January 2016, the model provides conditional mortality estimates. For example, the coefficients in the column vector  $\beta_1$  give estimates of death in 2016 for individuals who were diagnosed in 2015 and *who survived* until at least 1 January 2016. Because the average individual is diagnosed in June, the estimated probability of mortality is 0.5 years, 1.5 years (conditional on surviving 0.5 year), 2.5 years (conditional on surviving 1.5 years), and 3.5 years (conditional on surviving 2.5 years). To arrive at multi-year (unconditional) mortality estimates, we cumulate our coefficients. For example, if for a given diagnosis the mortality estimate at 0.5 years is  $\beta_0$  and the 1.5 years (conditional on surviving 0.5 year) estimate is  $\beta_1$ , we estimate  $P(\text{died within 1.5 years}) = 1 - P(\text{survived 1.5 years}) = 1 - (1 - \beta_0)(1 - \beta_1)$ .

The regression in Equation 2 provides mortality estimates up to 3.5 years for each of the 18 sub-samples. For illustrative purposes, in Section 4, we also present mortality estimates for up to 5.5 years. To do so, we use a different empirical strategy. We run cross-sectional regressions for all years from 2013 to 2018, where the dependent variable is an indicator if an individual, who was alive on 1 January 2013, died by the end of the given year  $t$ . The independent variables are a set of indicators for 2013 diagnoses ( $D_{i,2013}$ ), a fourth-degree age polynomial ( $F(age_{i,t}, \gamma)$ ), and gender ( $\gamma_{G(i)}$ ) and permanent income tercile ( $\eta_{PI(i)}$ ) fixed

effects. For example for the year 2013 (0.5-year mortality):

$$Dead_{i,2013} = \alpha + D'_{i,2013}\beta_{2013} + F(age_{i,t}, \gamma) + \gamma_{G(i)} + \eta_{PI(i)} + \epsilon_{i,t}, \quad (3)$$

where  $Dead_{i,2013}$  is an indicator if person  $i$ , who was alive on 1 January 2013, died by the end of 2013.

As with other outcomes in Section 4, we also estimate this regression on the whole sample, and in sub-samples formed by gender, income groups, and age groups. This alternative specification entails a drawback for some diagnoses because it does not control for the mortality effect of previously diagnosed diseases.

## 4 Results

Different health shocks can have distinct effects on personal risks, such as long-term care expenses, income prospects, and mortality. In addition, individuals in distinct socioeconomic strata may be differently affected by the same health shock. Our rich cross-section of health shocks and affected individuals enables us to document such heterogeneous effects.

### 4.1 A Representative Example

We pick three diseases to illustrate the heterogeneous effects of health shocks. As an example of a disease with little impact on income prospects and mortality, but with non-negligible medical expenses, we study appendicitis. We choose lung cancer as an example of a disease with a high mortality risk. Finally, we select herniated intervertebral disk (HNP) as an example of a common disease, in all ages, with a potentially large effect on income.

Figure 1 shows the estimated effect on medical expenses, labor, and mortality.<sup>13</sup> Several observations merit attention. First, for all health shocks, most medical expenses occur in

---

<sup>13</sup>In Section 4, unless otherwise specified, estimates of the effects of medical expenses and mortality are based on the sample of all individuals, while estimates of labor effects are based on the sample of working-age individuals.

the year of diagnosis, and these expenses differ substantially across diseases. Lung cancer is the most expensive diagnosis with medical expenses of EUR 18,084; appendicitis and HNP lead to lower expenses at EUR 4,611 and 1,612, respectively. The expense wedge between these two health shocks reflects that appendicitis requires surgery, while HNP is most often treated with physical therapy and medication.

Second, these three diseases also differ substantially in how much they impact labor income. Here we measure the “income-effect” as the effect of the health shock on the probability of working.<sup>14</sup> Lung cancer has serious consequences, it reduces the probability of working by 11 p.p three years after the diagnosis. HNP has a similar effect with an estimate of 6.3%, which reflects the pain and other work-limiting symptoms associated with HNP, and the difficulty of treatment. On the contrary, appendicitis has little impact on income, which is in line with the usually quick recovery following the surgical removal of the appendix. Third, while the excess mortality of lung cancer reaches 37% after three years, HNP and appendicitis have little impact on life expectancy.

[Insert Figure 1 here]

How do the effects of health shocks differ across various socioeconomic groups? To answer this question, we repeat the previous analysis for individuals who belong to distinct permanent income and age groups.<sup>15</sup> The upper-left sub-figure of Figure 2 shows that medical expenses do not depend on the permanent income group of the individual. This is the expected consequence of the universal healthcare system in the Netherlands. On the contrary, the upper-right sub-figure reveals substantial variation in the effect of the same health shock on labor income by income group. For example, the effect of HNP on labor market participation is 10% for the lowest income group (solid line), but it is only 2.9% for the highest group (dashed line). Similarly, the effect of lung cancer on labor income is larger for

---

<sup>14</sup>The effects on labor earnings are in general very mild, as we will show in Table 2.

<sup>15</sup>We discuss the formation of our three income groups – low, medium, and high – in Section 2.2.5 and in the Internet Appendix.

people with lower income than those with higher income. The bottom sub-figure also reveals some heterogeneity in the effect of lung cancer on life expectancy.

[Insert Figure 2 here]

Figure 3 presents a similar analysis for individuals with different ages at diagnosis. We split the working-age sample into two groups, those diagnosed between 25 and 44 years old and those diagnosed between 45 and 64 years old. The upper-left sub-figure shows that a lung cancer diagnosis later in life (dashed line) is associated with higher medical expenses. There are no age-related differences in the expenses of appendicitis and HNP. The upper-right sub-figure shows the income effects. Lung cancer is more likely to make older people leave the labor market. There is little difference in how younger and older people respond to appendicitis and HNP in terms of labor participation following the diagnosis, but the effect of HNP is large with a coefficient estimate of 6.3%. This means that HNP can have a large impact on human capital early in the life-cycle. The bottom figure reveals that older people are more than twice as likely to die from lung cancer than younger people.

[Insert Figure 3 here]

Taken together, these results illustrate that different health shocks present markedly different effects on medical expenses, labor, and mortality, and that the same health shock can have a distinct impact on people of different social strata.

## 4.2 All Health Shocks

In the previous subsection we illustrated the heterogeneity in the effects of health shocks using three examples, lung cancer, appendicitis, and a herniated disk. In this subsection we summarize the effects across all 334 health shocks in our sample. Table 2 contains the average causal effects (equally weighted). To illustrate the variation in the effects of health

shocks on a particular outcome, we also report the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the estimated effects.

Each panel in Table 2 presents the effects of health shocks for individuals with common characteristics. The upper panel is based on the whole sample. The second panel presents the estimates for men and women separately. In the third panel, we condition on permanent income group. The last panel presents the results by groups formed on the age at diagnosis. The first column shows the results for medical expenses. The second column contains the corresponding results for excess mortality. The third, fourth, and fifth columns show the results for labor market outcomes: labor market participation, being fully disabled, and log labor earnings, respectively. The last column reports the estimates on nursing home use.

[Insert Table 2 here]

The first panel of Table 2 provides several new insights on the range of possible consequences of poor health. First, the range that covers the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the treatment effects is wide for all outcomes. Second, medical expenses range from a few hundred euros to over EUR 18,000, with a mean of EUR 5,300. Some of the most expensive diagnoses include blood cancers (leukemias, lymphomas), the treatment of which might require expensive anticancer drugs or stem cell transplants, and subarachnoid hemorrhage (an uncommon type of stroke caused by bleeding on the surface of the brain), which often requires surgery. Excess mortality reveals a similar spread: it ranges from practically zero to an increase in the probability of dying during the next 3 years of over 23%. Some of the most mortal health shocks include pleural mesothelioma (a rare cancer that is most commonly caused by asbestos exposure) and malignant cancers of the digestive organs (e.g., pancreatic cancer). Third, the effect of a negative health event on participation in the labor market and dependence on disability insurance ranges from zero to approximately 15%, with a mean of 4%. Conditional on being able to work, the effect of earnings is smaller and less dispersed. Fourth, the probability of moving to a nursing home increases by up to 7%

The next panels of Table 2 show the same results for individuals with different socioeconomic characteristics. With a few exceptions, which we now describe, most health shocks have similar consequences for people of different sex, permanent income, and age. In other words, the average causal effect (equally weighted) and the range of causal effects reported in the second (gender), the third (permanent income), and the fourth (age) panel mostly coincide with the above. Three exceptions stand out. First, the average effect of a health shock on labor market participation is decreasing in permanent income. Thus, individuals with higher permanent income are less likely to leave the labor market. Simultaneously, individuals with lower permanent income are more likely to be declared fully disabled. Second, conditional on staying in the labor market, the potential (5<sup>th</sup> percentile) drop in earnings is much more severe for people in the low permanent income group, with an earnings drop of 14% compared with 6% for the other two permanent income groups. Finally, the average causal effect (equally weighted) of a health shock increases with age for all five measures.

### 4.3 Risk Correlations

A unique feature of our data is that we can estimate the correlation in the average effects of health shocks. This is interesting because it makes it possible to quantify health risk over the life-cycle and assess the willingness to pay for insurance products (or preventive options). For example, the willingness to hedge, or prevent, a disease with a low impact on mortality, but with a high impact on disability, is highest early in the life cycle. Table 3 shows the correlation matrix for the estimated average effects.

[Insert Table 3 here]

The key insight is that most correlation coefficients are quite different from each other; and in particular, variation in medical expenses does not explain all the variation in health risks. A notable exception is the correlation between the effect on disability and participation

in the labor market of -0.92.<sup>16</sup>

Figures 4 to 6 plot the effect estimates of various diagnoses on the probability of becoming disabled, labor market participation, and log earnings against medical expenses. The figures show the treatment effects on middle-aged individuals (45 to 64 years old) for both men and women. We label individuals with high (triangle) and low (circle) permanent income and use bold to signal statistically significant estimates.<sup>17</sup> Figures 7 and 8 are slightly different. In Figure 7, we plot the effect estimates of various diagnoses on excess mortality against medical expenses. We use separate points to distinguish groups of individuals based on age and income. In Figure 8, we plot the effect estimates of various diagnoses on the need for a nursing home against medical expenses in the sample of senior people.

[Insert Figure 4 here]

Figure 4 shows a positive correlation between the probability of becoming disabled and medical expenses. Nontraumatic subarachnoid hemorrhage (bleeding in the space that surrounds the brain), labeled by (1), is one example of a diagnosis that often results in disability (e.g., cognitive impairment) and requires expensive surgery. The figure also shows that many diagnoses increase the risk of disability without causing high medical expenses. Points (2) and (3) stand for systemic atrophies affecting the central nervous system (e.g., ALS) in women and men, respectively. These are severe terminal conditions with few treatment options. Points (4) and (5) represent dementia in women and men, respectively. Dementia is another diagnosis with a high impact on the likelihood of becoming disabled but with limited treatment options.

[Insert Figure 5 here]

Figure 5 is almost a mirror image of Figure 4. This is expected due to the strong negative correlation between disability and labor participation documented in Table 3. Points (1)

---

<sup>16</sup>The implication of the strong correlation is that disability insurance subsumes unemployment insurance.

<sup>17</sup>We classify a coefficient as statistically significant if its Bonferroni adjusted (by the number of coefficient estimates in the plot) p-value is below 5%.

to (4) highlight some more of the extreme observations in the figure. Points (1) and (2) correspond to malignant neoplasms of lymphoid, hematopoietic, and related tissue (e.g., leukemias, lymphomas) diagnosed in men and women, respectively. These cancers often require costly treatment (e.g., chemotherapy, bone marrow transplant) and substantially decrease labor participation. Point (3) stands for nontraumatic subarachnoid hemorrhage, which coincides with point (1) in Figure 4. Point (4) stands for systemic atrophies affecting the central nervous system, which coincides with point (2) in Figure 4.

[Insert Figure 6 here]

Figure 6 shows a weak correlation between diseases' effects on labor income for working individuals and the cost of treatment. Most diagnoses have only a modest impact on labor income. The highlighted points (1) to (4) represent a few exceptions. Points (1) and (2) stand for breast cancer diagnosed in high- and low-income women, respectively. Breast cancer treatment may involve costly surgeries, radiotherapy, and chemotherapy, and the disease negatively affects income. Point (3) is stroke (for men), a debilitating condition that might require surgical treatment and can also limit labor earnings. Point (4) stands for the diagnosis of mental and behavioral disorders due to psychoactive substance use (e.g., drug abuse, alcohol abuse).

[Insert Figure 7 here]

Figure 7 displays a modest positive correlation between diseases' effect on excess mortality and medical expenses. Still, some conditions impose very high medical costs and comparatively lower mortality risk (points 1 and 2), while other diagnoses are mortal but are associated with relatively low medical expenses (points 3-7). Point (1) stands for the diagnosis of chronic rheumatic heart disease, which may require costly surgical treatment (e.g., mitral valve replacement). Point (2) corresponds to myelodysplastic syndromes, a group of cancers in which immature blood cells in the bone marrow do not mature. Bone marrow



transplant is the only potential treatment option for myelodysplastic syndromes. Points (3) and (4) stand for mesothelioma diagnosed in high- and low-income women, respectively. Points (5) to (7) correspond to mesothelioma diagnosed in men of different age and income groups. Mesothelioma is a type of cancer that develops from the thin layer of tissue covering many internal organs and is often linked to asbestos exposure.<sup>18</sup>

[Insert Figure 8 here]

Figure 8 shows a modest positive correlation between diseases' effect on the need for nursing and medical expenses for individuals above 65. Again, we see that medical expenses are an imperfect proxy for the severity—or, put differently—the impact of a diagnosis on income and health. Some diagnoses are associated with relatively low medical expenses yet cause a substantial increase in the probability of needing a nursing home (points 1, 2, and 3). Points (1) and (2) stand for dementia, diagnosed among high- and low-income men, respectively. Dementia has limited treatment options, and patients often require long-term care. Point (3) corresponds to nontraumatic intracerebral hemorrhage, a type of stroke in older people mostly treated non-surgically. Contrary examples include diagnoses that generally do not affect the need for a nursing home but is very expensive to cure or treat. Two such examples are acute kidney failure and leukemias/lymphomas, labeled with points (4) and (5).

Overall, the results reveal that medical expenses are far from perfectly correlated with the other risk dimensions of health shocks. Consequently, we cannot hedge health risk with insurance contracts indexed only to medical expenses. Instead, we need alternatives such as disability insurance and nursing home subsidies. The challenge becomes how to aggregate

---

<sup>18</sup>We estimate that a few diagnoses have a large negative effect on excess mortality. An example is chronic subdural hematoma (an old clot of blood on the surface of the brain beneath its outer covering) represented by points (8) (female) and (9)-(10) (male) in Figure 7. Most cases of this diagnosis occur together (in the same year) with the diagnosis 'other and unspecified nontraumatic intracranial hemorrhage', while the reverse is not true. It is possible that the combination of these two diagnoses refers to a more favorable clinical picture of 'other and unspecified nontraumatic intracranial hemorrhage'. This could result in a negative mortality estimate on 'chronic subdural hematoma' because we estimate mortality effects in a multivariate regression. We are currently performing additional data analysis to address this phenomenon.

all the different risks into a number that represents the impact on the welfare of individuals. Because we need to consider the heterogeneity of risk across diseases but also across social strata and age, we calibrate a simple life-cycle model to obtain the certainty equivalent of each disease.

## 5 Quantification of Health Risk

We develop a framework to quantify the benefits of eliminating different health risks for different people. The idea is to compare the current situation with a counterfactual world in which a given health shock is eliminated. Three factors are crucial. First, the different incidences of health shocks in the population. Second, we need to account for health risk heterogeneity across the life-cycle and income distribution. Finally, we need to account for heterogeneity in the possibility of using savings to self-insure against health shocks.

### 5.1 Analytical Framework

#### 5.1.1 The Unit of Analysis

The unit of analysis is the individual. In each period of life, individuals receive utility from consumption and, possibly, from leaving a bequest if the individual is deceased. The model consists of a series of one-year periods, indexed as  $t$ , beginning at age  $M$  and ending at the year of death, which never exceeds  $T$ . The individual works for the first  $(K - M)$  years of life, unless she becomes disabled, and retires at age of  $K$  (deterministic). The state space of the model includes age ( $t \in [M, T)$ ), assets ( $a \in \mathbb{R}^+$ ), and four variables related to health.

We use six groups to account for disease heterogeneity over the life-cycle and across the income distribution. The Cartesian product of three permanent income groups and two genders makes up the groups. We treat group membership as a time-invariant characteristic and use  $g$  to index group membership.

### 5.1.2 The Decision Problem

Given a set of resources and health-related expectations, the individual chooses consumption ( $c \in \mathbb{R}^+$ ) to maximize utility. The consumption of people who live in nursing homes is provided by the government, which we label  $c_{NH}$ . Individuals have CRRA preferences with relative risk aversion  $\rho$ :

$$u(c_{it}) = \bar{u} + \frac{c_{it}^{1-\rho}}{1-\rho}, \quad (4)$$

where  $\bar{u}$  does not affect the policy function, but it is important to capture the value of being alive one more year. We follow [Kraft et al. \(2022\)](#) and let individuals derive utility from leaving a bequest of size  $a$ :

$$v(a_{it}) = \theta \frac{(R \frac{a_{it+1}}{\theta} + \bar{a})^{1-\rho}}{1-\rho}. \quad (5)$$

Here,  $\bar{a}$  measures the extent to which bequests are “luxury goods” and  $\theta$  measures the strength of the bequest motive. To limit the number of parameters, we set the risk aversion of bequest equal to the relative risk aversion coefficient of consumption  $\rho$ .<sup>19</sup>

For individuals not living in nursing homes, the law of motion of assets is:

$$a_{it+1} = a_{it}R + y_{it} - \tau_y(y_{it}) - c_{it} - premium - \min(m_{it}, ded), \quad (6)$$

where  $R \equiv (1 + (1 - \tau_a)r)$  denotes the gross return after taxes,  $m_{it} \in \mathbb{R}^+$  medical expenses,  $y_{it}$  stands for the income of a healthy worker, and  $\tau_y(y) = y - \phi_0 y^{1-\phi_1}$  represents income after tax following [Heathcote et al. \(2017\)](#). The parameter  $\phi_1$  determines how progressive the tax system is.<sup>20</sup> The variables and parameter represent the following: *premium* is the amount paid to the health insurer per year,  $m_t$  are medical expenses, and *ded* determines the maximum annual amount paid by an individual.

<sup>19</sup>The estimates in [Kvaerner \(2022\)](#) shows that one cannot reject that the bequest exponent equals risk aversion for several levels of risk aversion using the same utility function for bequest as we do.

<sup>20</sup>We can see this by noting that under this tax function, pre-tax income ( $y$ ) is mapped to post-tax income ( $\tilde{y}$ ) as  $\tilde{y} = \phi_0 y^{1-\phi_1}$ . Therefore, the elasticity of post-tax income with respect to pre-tax income is  $(y/\tilde{y})(d\tilde{y}/dy) = (1 - \phi_1)$ . The higher  $\phi_1$ , the less elastic post-tax income is with respect to pre-tax income, i.e., the more progressive is the tax system.

Instead, if the individual, of permanent income-gender group  $g$ , enters a nursing home, she moves into an absorbing state in which she consumes annually  $c_{NH}$  and pays a cost  $e_{NH}^g$ .<sup>21</sup> These costs are on top of any health insurance premium and income taxes, and are paid every year until the user runs out of assets. In this case, the user pays the retirement income after taxes and health insurance premium. In addition to the individual's contribution, there is a government contribution up to the total cost of the stay at the nursing home of  $E_{NH}$ . Therefore, for individuals having entered the absorbing state of living in a nursing home, the law of motion of assets is:

$$a_{it+1} = a_{i,t}R + y_t - \tau_y(y_t) - c_{NH} - \text{premium} - \min(m_t, \text{ded}) - e_{NH}^g, \quad (7)$$

We denote the probability of dying as  $p \in [0, 1]$ . We let  $s \in [0, 1]$  be the probability of becoming disabled and use  $d \in [0, 1]$  to denote labor capacity if ill. The probability of moving to a nursing home is  $\pi \in [0, 1]$ .

The set  $\mathcal{H}_{i,t+1} = \{p_{i,t+1}, m_{i,t+1}, \pi_{i,t+1}, d_{i,t+1}, s_{i,t+1}\}$  contains the health-related state-variables, which we assume are first-order Markov with transition dynamics:  $\mathcal{H}_{i,t+1} = g(\mathcal{H}_{i,t})$ .

### 5.1.3 Income and Pension

To quantify the effect of health shocks, we need to define a “baseline” to which we can add the estimated causal effects. In the Netherlands, there is limited cross-sectional variation in pension benefits. This is because the state-provided pension benefit depends mostly on the family composition, whereas the variation in pension benefits accumulated under employer and industry pensions funds is largely accounted for by the permanent income groups.<sup>22</sup> For this reason, all retirees in the model receive pension income  $\bar{y}^g$ .

---

<sup>21</sup>This assumption, which is close to reality, implies that consumption does not differ between income groups, gender, or age for nursing home residents, while costs do differ.

<sup>22</sup>See the description of the Dutch pension system in the Internet Appendix.

Before retirement, people who participate in the labor market receive:

$$y_t = \max\{d_t \tilde{y}_t, \underline{y}\} \text{ where;} \tag{8}$$

$$\underline{y} \equiv \text{minimum income}$$

$$\ln(\tilde{y}_t) \equiv \sum_{i=0}^4 \alpha_i^g \cdot t^i + \gamma^g y_{t-1},$$

where  $d_t \in [0, 1]$  is a factor that represents the labor income reduction due to bad health at age  $t$ . Thus,  $\tilde{y}_t$  is the labor income of a healthy ( $d_t = 1$ ) worker. We estimate the age-dependent  $\tilde{y}_t$  using OLS estimation on annual observations of the labor income of “healthy” Dutch individuals for the years 2013 to 2016. For each year, we only include individuals in the sample who were not diagnosed with any of our 334 diagnoses in the given year, who did not die that year, and who were not receiving disability benefits.

We make two important assumptions. First, we do not distinguish wages from hours worked. This assumption is innocuous as no one in the model chooses how much to work.<sup>23</sup> Second, as our goal is to quantify the utility gains from reducing health risk, and not to explain saving behavior, we omit the stochastic part of labor income.

#### 5.1.4 Consequences of Health Shocks

We now explain how health shocks affect individuals in the model. We let a “healthy individual” be a person who has not received any diagnoses. This person, especially if older, might still need a nursing home, and invest in medical products (e.g. hearing aids, antihistamines, etc.). For this reason, we first model the dynamics of these state variables in the absence of any diagnoses. We refer to this process as the “baseline” process, and denote it by tilde. We estimate each baseline process for each of the six groups separately.

---

<sup>23</sup>Most of the effect of the studied health shocks concentrates in labor participation (see Table 2). In addition to considering the decision to work or not, considering by how much to participate in the labor market complicates the model excessively because this latter decisions depends heavily on households’ current and future composition (e.g. becoming a stay-at-home parent).

Formally, we let medical expenses follow:  $\tilde{m} = \sum_{i=0}^4 \kappa_i^g \cdot t^i$  in the absence of any diagnosis. The parameters  $\kappa_i^g$  are estimated by OLS on annual observations of the total medical expenses of Dutch individuals for the years 2013 to 2016. For each year, we only include individuals in the sample who were not diagnosed with any of our 334 diagnoses in the given year, who did not die that year, and who were not receiving disability benefits either. The aim of these restrictions is to limit the sample to “healthy” individuals. For nursing home use, we assign a zero probability of a healthy individual entering a nursing home before the age of 65. After the age of 65, the nursing home process is:  $\tilde{s} = \sum_{i=0}^4 \nu_i^g \cdot t^i$ . We estimate the parameters  $\nu_i^g$  by OLS on the same sample of “healthy” individuals as we did for the medical expenses (but we restrict the sample to 65+ individuals). For the labor state variables, we require that healthy individuals not be disabled, or experience a sudden income loss.

Once an individual is diagnosed with a disease, all state variables can change. Regarding medical expenses, we add the disease-specific causal effect in the year of diagnosis to the baseline process. In relation to entering a nursing home, we assume that the probability increases in the year of the diagnosis by the size of the causal estimate on the cumulative probability over three years following the diagnosis. We use three years post diagnosis because entering into a nursing home as an absorbing state.

We face a related issue when incorporating the effect of the health shock on labor market participation and earnings among participants. The reason is that under Dutch law, employers must pay at least 70% of the salary of sick employees for the first two years of illness. We address this point by setting the probability of being disabled equal to the causal effect on labor participation three years after the diagnosis. We follow the same approach for earnings. We assume that the effect of the health shock on labor earnings is permanent, meaning that any effects from subsequent diagnoses come on top of the previous losses.

Finally, the baseline process for mortality and the diagnosis-specific effect on mortality is captured by equation (2). These estimates are often negative, although very close to zero, for non mortal diseases. This bias likely arises from individuals who suffer from the

diseases but do not get diagnosed and worsen or even die. To avoid that this small bias leads our results, we truncate the excess mortality at zero; hence, we assume that diseases cannot extend your life. This issue does not arise in the other outcome variables because our identification assumption compares two individuals who were diagnosed at some point.

### 5.1.5 Optimal Behavior

We use the endogenous grid method proposed by [Carroll \(2006\)](#) and backward induction to solve the model. Solving the value function gives a set of optimal decision rules for any given realization of the state variable. To simplify the notation, we now omit the subscript  $i$  and drop the value function in the last period in the exposition.<sup>24</sup> In retirement, the indirect utility and the budget constraint of someone in a nursing home are:<sup>25</sup>

$$V_t^{NH}(a_t, p_t^g) = u(c_{NH}) + \beta [p_t^g v(a_{t+1}) + (1 - p_t^g) \mathbb{E}(V_{t+1}^{NH}(a_{t+1}, p_{t+1}^g))] \quad (9)$$

$$\text{s.t. } a_{t+1} = \max \left\{ Ra_t + \bar{y}^g - \tau_y(\bar{y}^g) - e_{NH}^g - \text{premium} - \min(m_t, \text{ded}), 0 \right\}.$$

At age  $K \leq t < T$ , the value function, and the budget constraint are:

$$V_t(a_t, p_t^g, \pi_t^g) = \max_{c_t \geq 0, a_{t+1} \geq 0} u(c_t) + \beta p_t^g v(a_{t+1})$$

$$+ \beta(1 - p_t^g) \mathbb{E}(\pi_t^g V_{t+1}^{NH}(a_{t+1}, p_{t+1}^g) + (1 - \pi_t^g) V_{t+1}(a_{t+1}, p_{t+1}^g, \pi_{t+1}^g)) \quad (10)$$

$$\text{s.t. } a_{t+1} = Ra_t + \bar{y}^g - \tau_y(\bar{y}^g) - \text{premium} - c_t - \min(m_t, \text{ded}).$$

The expectation  $\mathbb{E}(\cdot)$  reflects uncertainty about future health shocks and their (joint) implication for all risks. Therefore, the expectation  $\mathbb{E}(\cdot)$  takes into account the incidence and the effects of possible future diagnoses.

The year before retirement, the decision problem is almost identical to that of the first

---

<sup>24</sup>In the last period, the value function is deterministic. The individual leaves a bequest to ensure that the marginal utility of giving equals the marginal utility of personal consumption.

<sup>25</sup>There is nothing to maximize in this state because there is nothing to choose.

year of retirement except for the budget constraint; hence we omit it for brevity. Up to the age of retirement (i.e.,  $t < K$ ), the individual might be disabled. If this were the case, the value function is:

$$\begin{aligned}
V_t^{dis}(a_t, p_t^g, t_{dis}, d_{dis}) &= \max_{c_t \geq 0, a_{t+1} \geq 0} u(c_t) + p_t^g \beta v(a_{t+1}) \\
&\quad + (1 - p_t^g) \beta \mathbb{E} (V_{t+1}^{dis}(a_{t+1}, p_{t+1}^g, t_{dis}, d_{dis})) \\
\text{s.t. } a_{t+1} &= Ra_t + \max\{d_{dis}, y_{t_{dis}}^g - \tau_y(d_{dis} y_{t_{dis}}), \underline{y}\} - \text{premium} - c_t - \min(m_t, \text{ded}).
\end{aligned} \tag{11}$$

In this case, income is determined by the time that the person becomes disabled. Other risks, such as mortality and the possibility of needing a nursing home, remain.

If the individual is working, the value function is:

$$\begin{aligned}
V_t(a_t, p_t^g, d_t^g, s_t^g) &= \max_{c_t \geq 0, a_{t+1} \geq 0} u(c_t) + p_t^g \beta v(a_{t+1}) + \\
&\quad (1 - p_t^g) \beta \mathbb{E} (s_t^g V_{t+1}^{dis}(a_{t+1}, p_{t+1}^g, t, d_t^g) + (1 - s_t^g) V_{t+1}(a_{t+1}, p_{t+1}^g, d_{t+1}^g, s_{t+1}^g)) \\
\text{s.t. } a_{t+1} &= Ra_t + \max\{d_t^g y_t^g - \tau_y(d_{dis} y_{t_{dis}}), \underline{y}\} - \text{premium} - c_t - \min(m_t, \text{ded}).
\end{aligned} \tag{12}$$

### 5.1.6 Calibration

We now describe how we parameterize the model. The model captures most health risks, a prerequisite for computing counterfactual scenarios without certain health risks, but that does not explain savings over the life cycle.<sup>26</sup> For that reason, we parameterize the model using the estimates in [Kvaerner \(2022\)](#) and [De Nardi et al. \(2017\)](#).<sup>27</sup>

We set the parameters corresponding to government subsidies and payments to those faced by the average Dutch individual in each group. The minimum income is set at EUR

---

<sup>26</sup>There are several reasons for this. For example, the model does not allow for changes in the household composition and does not include labor income risk (“income shocks”).

<sup>27</sup>[Kvaerner \(2022\)](#) uses health shocks in the Netherlands and Norway to estimate several specifications of “warm-glow” bequest functions. We use the same calibration for all preference parameters except for  $\bar{u}$ , which he does not consider. [De Nardi et al. \(2017\)](#) shows this parameter becomes relevant when we compare health effects on income and mortality. Consequently, we set  $\bar{u}$  to match a value of statistical life of EUR 2 million following [De Nardi et al. \(2017\)](#).



12,700 for all groups.<sup>28</sup> The individual’s contribution to a nursing home’s cost is more complicated as it depends on the wealth of the individual and a few additional characteristics, which is why we allow the individual contribution to differ across social groups. The average personal contribution to the cost of a nursing home stay is set at EUR 12,852, 16,721, 22,881 for poor, medium, and rich women respectively; and EUR 13,000, 13,600, and 18,240 for men.<sup>29</sup>

We estimate the annual cost of nursing home use from aggregate public data published by Statistics Netherlands. Specifically, we divided the total costs by occupants, which amounts to EUR 43,545 ( $E_{NH}$ ). We set the consumption in a nursing home at 70% of the total cost ( $c_{NH} = 30,481$ ). The same percentage corresponds to the income received by workers when disabled ( $\lambda_{dis} = 0.7$ ).

The government plays an important role in the analysis. It finances the health care system and receives income tax from healthy individuals. We equate the health insurance premium to the monthly average rate in 2015 of EUR 97. This flat rate is independent of age, health, and other characteristics. We set the deductible, that is, what the individual has to pay before the health insurance kicks in at the 2015 value of EUR 375.<sup>30</sup> Regarding direct taxes, we set the wealth return tax at 30%. The rate applies to fictive wealth returns under the Dutch Box 3 tax category. For income taxation, we estimate the following parameters

---

<sup>28</sup>This is a close estimation of the pre-tax Dutch “social minimum” in 2015. The Dutch state considers the social minimum the minimum amount one needs to provide for one’s livelihood. The system of social benefits is set up so that it always provides at least the social minimum. For single individuals the pre-tax social minimum equals 70% of the after-tax minimum wage.

<sup>29</sup>We calculate the average personal contribution to the costs of nursing home stay in these six groups using the formula published by the institution responsible for the collection of these contributions (CAK). The formula is based on income, wealth, and marital status, among other factors.

<sup>30</sup>Individuals can opt for a higher deductible, up to a maximum of EUR 875, deductible in return for a lower premium. We do not model this choice because only about 10% of the Dutch population opts for a higher-than-mandatory deductible.

of the income tax function presented above at  $\phi_0 = 6.26, \phi_1 = 0.21$ .<sup>31</sup> Table 4 and Table 5 summarize the model parameters.

[Insert Table 4 and Table 5 here]

In the counterfactual analysis, we combine the model parameters with the estimated causal effects and the processes that determine each disease’s incidence. We focus on three diseases: appendicitis, lung cancer, and HNP. We chose them because they have different effects on our outcome variables for the same individual and, in some cases, different implications for the same outcome variable for people of different social strata (e.g., low- and high-permanent income individuals). With only three diseases, we can represent all health-related state variables with a four-point grid per age and group (one for each disease and one for healthy individuals).

Before we present the results, we need to make a few additional assumptions. First, motivated by the results in Section 4, we let all effects last for one period. The approximation is good for mortality, disability, and medical expenses, but less so for labor income. Second, the likelihood of being diagnosed with a particular condition is independent of the patient’s medical history.<sup>32</sup> Moreover, motivated by the low incidence and low number of diagnoses for aforementioned conditions, we restrict the annual number of health shocks per individual to one. Given these assumptions, we estimate the incidence of each diagnosis independently for each income, age, and gender group conditional on an age polynomial.

---

<sup>31</sup>We observe total income tax due over the sum of income from work/pension/benefits (Box 1), substantial shareholdings (Box 2), and a fictive return on wealth (Box 3). To estimate the tax function parameters for income in Box 1, we deduct tax due in Boxes 2 and 3 from the total income tax. In addition, we deduct income in Boxes 2 and 3 from total taxable income. This adjustment is possible because we observe income in Box 2, can estimate income in Box 3 using wealth data, and because income in both boxes is taxed separately from the income in Box 1, which is taxed at a flat rate of 30%. We estimate the tax function parameters using nonlinear least-squares estimation.

<sup>32</sup>The conditionally independence assumption is reasonable in the case of some diagnoses, such as the ones we now analyze, but it needs to be relaxed when analyzing all health risks

## 5.2 Results

Having solved the model in which an individual can face the risk of any of the three health shocks, we consider a counterfactual scenario in which one is eliminated. For example, people who otherwise would have had lung cancer remain healthy. We then increase the health insurance premium until a 25-year-old individual has the same lifetime welfare as in the baseline scenario. This increase in health insurance resembles an annual certainty equivalent as it represents the amount individuals are willing to pay annually to avoid the risk of contracting a disease. We report the increase in insurance premium in Table 6.

The central insight from Table 6 is that the possibility of removing a particular health risk affects people from different social strata differently. For example, low-income women would pay EUR 157 and EUR 282 to eliminate HNP and lung cancer. In stark contrast, high-income women would pay almost ten times more to eliminate cancer than to eliminate HNP. The considerable heterogeneity in the willingness to pay for eliminating HNP is because its variation in income effects (independently) exceeds the variation in the combined effects of lung cancer on income and mortality. Appendicitis is different. Individuals barely value the disappearance of this disease because it mainly affects medical expenses, which the government insures.

The previous counterfactual is not revenue-neutral from the government's point of view. For instance, eliminating lung cancer increases pension payments and nursing home subsidies due to a longer life expectancy, but also increases savings and, thus, tax revenues. HNP barely increases life expectancy but decreases disability payments. All three diseases reduce the government expenses for medical care to different extents. To account for all these channels, we obtain the extra revenue that the government receives if the government were to set the health insurance premiums so that individuals' welfare remains constant after eliminating the disease. To illustrate this with an example, let us turn to the last panel of Table 6 which shows that, on average, the government would receive EUR 254,888 more from each individual if it eliminated lung cancer. This number suggests that the Netherlands

could fund around EUR 57 billion targeted to research on treatment of lung cancer.<sup>33</sup>

[Insert Table 6 here]

### 5.3 Conclusion

We use data on 6.9 million people diagnosed by medical specialists between 2013 and 2017 to estimate the effects of 334 distinct medical diagnoses on six outcomes (medical expenses, mortality, disability, participation in the labor market, labor earnings, and nursing home use) across several social groups formed by three individual characteristics (gender, age, and permanent income). Across all these groups, diagnoses, and economic and health outcomes, we estimate a total of about 24,000 causal effects of different health shocks.

One key insight from our analysis is that health shocks are heterogeneous and this heterogeneity is not perfectly explained by differences in medical expenses. A second important insight is that people of different economic strata respond differently to the same health shock. Consequently, the progressivity in “eliminating” different diseases (by means of “targeted” medical research) is very dissimilar; a wealthy individual would prefer to eliminate fatal health risks in old age (e.g., lung cancer), while less wealthy people care more about removing health risks with a large impact on labor income early in life (e.g., spinal disk herniation). As a result, governments with redistribution concerns need to design different funding schemes for medical research aimed at distinct health risks. For example, research on disk herniation (which mainly affects the labor income of low-income people early in life) increases overall redistribution towards low-income households unless it is funded by a regressive tax.

The “revealed preference” of current investors and governments is “targeted” research on diseases that benefit the wealthiest the most. For example, the NIH (National Institutes of Health) reports an investment of USD 451 million in research aimed at lung cancer.<sup>34</sup> This

---

<sup>33</sup>The cohort size for people between the age of 20 and 25 is 1,132,885 people (CBS)

<sup>34</sup>We consider US investment data because it is very comprehensive. Source: <https://report.nih.gov/funding/categorical-spending>.

amount is in practice much higher; as we need to include, at minimum, investments in lung research (USD 2.2 billion), cancer research (USD 7.3 billion), and cancer genomics (USD 1.1 billion). In comparison, directed research on spinal disk herniation is absent. If we include research in areas such as “back pain” and “chronic pain”, we end up with only USD 813 million (a clear upper bound).<sup>35</sup>

---

<sup>35</sup>We acknowledge an important caveat: when assessing research investment, we should also consider the expected return of each dollar. It may be that advances in disk herniation research are currently too expensive compared to lung cancer.

## References

- Borusyak, K. and X. Jaravel (2017). Revisiting event study designs, with an application to the estimation of the marginal propensity to consume. *Working Paper*.
- Branson, Z. (2018). Is my matched dataset as-if randomized, more, or less? unifying the design and analysis of observational studies. *arXiv preprint arXiv:1804.08760*.
- Byford, S., D. J. Torgerson, and J. Raftery (2000). Cost of illness studies. *Bmj* 320(7245), 1335.
- Carroll, C. D. (2006). The method of endogenous gridpoints for solving dynamic stochastic optimization problems. *Economics letters* 91(3), 312–320.
- De Nardi, M., S. Pashchenko, and P. Porapakkarm (2017). The lifetime costs of bad health. Technical report, National Bureau of Economic Research.
- DeNardi, M., E. French, and J. B. Jones (2010). Why do the elderly save? the role of medical expenses. *Journal of Political Economy* 118(1), 39–75.
- DiMasi, J. A., H. G. Grabowski, and R. W. Hansen (2016). Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of health economics* 47, 20–33.
- Dobkin, C., A. Finkelstein, R. Kluender, and M. J. Notowidigdo (2018). The economic consequences of hospital admissions. *American Economic Review* 108(2), 308–52.
- Doeskeland, T. and J. S. Kvaerner (2022). Cancer and portfolio choice: Evidence from norwegian register data. *Review of Finance* 26(2), 407442.
- Evens, R. and K. Kaitin (2015). The evolution of biotechnology and its impact on health care. *Health Affairs* 34(2), 210–219.
- Fadlon, I. and T. H. Nielsen (2019). Family health behaviors. *American Economic Review* 109(9), 3162–91.

- Fadlon, I. and T. H. Nielsen (2021). Family labor supply responses to severe health shocks: Evidence from danish administrative records. *American Economic Journal: Applied Economics* 13(3).
- García-Gómez, P., H. Van Kippersluis, O. O'Donnell, and E. Van Doorslaer (2013). Long-term and spillover effects of health shocks on employment and income. *Journal of Human Resources* 48(4), 873–909.
- Heathcote, J., K. Storesletten, and G. L. Violante (2017). Optimal tax progressivity: An analytical framework. *The Quarterly Journal of Economics* 132(4), 1693–1754.
- Heinesen, E. and C. Kolodziejczyk (2013). Effects of breast and colorectal cancer on labour market outcomes average effects and educational gradients. *Journal of health economics* 32(6), 1028–1042.
- Karpati, D. (2022). Household finance and life-cycle economic decisions under the shadow of cancer. *Working Paper*.
- Koijen, R. S., S. Nieuwerburgh, and M. Yogo (2016). Health and mortality delta: Assessing the welfare cost of household insurance choice. *Journal of Finance* 71(2), 957–1010.
- Kraft, H., C. Munk, and F. Weiss (2022). Bequest motives in consumption-portfolio decisions with recursive utility. *Journal of Banking & Finance* 138, 106428.
- Kutikova, L., L. Bowman, S. Chang, S. R. Long, C. Obasaju, and W. H. Crown (2005). The economic burden of lung cancer and the associated costs of treatment failure in the united states. *Lung Cancer* 50(2), 143–154.
- Kvaerner, J. (2022). Intergenerational altruism: estimates based on news about expected mortality. *Available at SSRN 2985465*.
- Larg, A. and J. R. Moss (2011). Cost-of-illness studies. *Pharmacoeconomics* 29(8), 653–671.

- Lundborg, P., M. Nilsson, and J. Vikström (2015). Heterogeneity in the impact of health shocks on labour outcomes: evidence from swedish workers. *Oxford Economic Papers* 67(3), 715–739.
- McGuire, A., M. Martin, C. Lenz, and J. Sollano (2015). Treatment cost of non-small cell lung cancer in three european countries: comparisons across france, germany, and england using administrative databases. *Journal of medical economics* 18(7), 525–532.
- Papanicolas, I., L. R. Woskie, and A. K. Jha (2018). Health care spending in the united states and other high-income countries. *Jama* 319(10), 1024–1039.
- Rockhill, B., I. Kawachi, and G. A. Colditz (2000). Individual risk prediction and population-wide disease prevention. *Epidemiol Rev* 22(1), 176–80.



# A Tables

**Table 1: Descriptive Statistics**

The table reports summary statistics at the end of 2012 for a treatment and a control group. The treatment group is defined as individuals that experience a health shock during 2013. The control group experiences a health shock in 2014. All monetary variables are measured in EUR thousands. The standardized mean difference (SMD) for a given variable is defined as the difference in means between the treatment and the control group divided by the square root of the sum of the two corresponding variances. Section 2.2 contains details about variable definitions.

| Means and Standard Deviations |                 |       |        |               |       |        |       |
|-------------------------------|-----------------|-------|--------|---------------|-------|--------|-------|
| <i>Demographics</i>           | Treatment Group |       |        | Control Group |       |        | SMD   |
|                               | N (000s)        | Mean  | SD     | N (000s)      | Mean  | SD     |       |
| Age (years)                   | 4,432           | 59.6  | 16.5   | 4133.2        | 59.0  | 16.3   | 0.03  |
| Female                        | 4,432           | 0.53  | 0.50   | 4133.2        | 0.53  | 0.50   | 0.00  |
| General secondary education   | 167             | 0.23  | 0.42   | 156.0         | 0.24  | 0.42   | -0.01 |
| College education (HBO, WO)   | 167             | 0.22  | 0.42   | 156.0         | 0.22  | 0.42   |       |
| <i>Medical statistics</i>     |                 |       |        |               |       |        |       |
| Medical expenses              | 4,420           | 4.2   | 9.7    | 4121.5        | 3.7   | 8.5    | 0.04  |
| Voluntary deductible          | 4,420           | 0.041 | 0.2    | 4122.0        | 0.047 | 0.21   | -0.02 |
| <i>Wealth and income</i>      |                 |       |        |               |       |        |       |
| Net wealth                    | 4,375.4         | 200.3 | 1425.7 | 4089.9        | 199.9 | 1588.8 | 0.00  |
| Financial assets              | 4,375           | 69.1  | 450.0  | 4089.9        | 68.0  | 385.0  | 0.00  |
| Labor earnings                | 1,665           | 37.8  | 34.0   | 1665.5        | 37.7  | 33.3   | 0.00  |
| Labor participation           | 2,475           | 0.69  | 0.46   | 2379.0        | 0.70  | 0.46   | -0.01 |
| Permanent Income Group 1      | 4,432           | 0.35  | 0.48   | 4133.2        | 0.35  | 0.48   | 0.00  |
| Permanent Income Group 2      | 4,432           | 0.34  | 0.47   | 4133.2        | 0.34  | 0.47   | 0.00  |
| Permanent Income Group 3      | 4,432           | 0.31  | 0.46   | 4133.2        | 0.31  | 0.46   | 0.00  |
| Distribution (percentiles)    |                 |       |        |               |       |        |       |
|                               | Treatment Group |       |        | Control Group |       |        |       |
|                               | 25              | 50    | 75     | 25%           | 50%   | 75%    |       |
| Age                           | 47.0            | 61.0  | 73.0   | 46.0          | 60.0  | 73.0   |       |
| Medical expenses              | 0.5             | 1.5   | 4.1    | 0.4           | 1.3   | 3.6    |       |
| Net wealth                    | 0.9             | 42.7  | 204.0  | 0.8           | 44.1  | 204.9  |       |
| Financial assets              | 3.1             | 15.7  | 51.9   | 3.1           | 15.8  | 51.8   |       |
| Labor earnings                | 21.3            | 33.1  | 46.6   | 21.4          | 33.1  | 46.7   |       |

**Table 2: Average Effects**

The table contains the summary statistics of the causal estimates. The first panel considers the pooled sample and presents the mean estimate across diseases, the 5th and 95th percentiles (within square brackets) and the proportion of diagnoses for which we reject the hypothesis of no effect against a one-sided alternative. The remaining panels include the mean, 5th, and 95th percentiles for different sub-samples. Medical expenses refer to the dollar amount spent on the year of diagnosis. Excess mortality is the probability of dying within the three years following the diagnosis. Labor participation, disability, and nursing home are the probability of working, being disabled, or living in a nursing home, three years after the diagnoses. Log earnings refers to the logarithm of labor earnings.

| Year                               | Medical expenses<br>(EUR)<br>$t = 0$ | Excess Mortality<br>(%)<br>$t = 3$ | Labor participation<br>(%)<br>$t = 3$ | Disability<br>(%)<br>$t = 3$ | Log earnings<br>(%)<br>$t = 3$ | Nursing home<br>(%)<br>$t = 3$ |
|------------------------------------|--------------------------------------|------------------------------------|---------------------------------------|------------------------------|--------------------------------|--------------------------------|
| <b>All</b>                         |                                      |                                    |                                       |                              |                                |                                |
| Mean                               | 5385                                 | 0.04                               | -0.04                                 | 0.04                         | -0.02                          | 0.02                           |
| 5%;95%                             | [401;18300]                          | [-0.02;0.23]                       | [-0.15;0.00]                          | [0.00;0.15]                  | [-0.06;0.01]                   | [0.00;0.07]                    |
| % Sign.                            | 99%                                  | 54%                                | 66%                                   | 74%                          | 49%                            | 79%                            |
| <b>By gender</b>                   |                                      |                                    |                                       |                              |                                |                                |
| Male                               | 5838<br>[478;19103]                  | 0.04<br>[-0.02;0.22]               | -0.04<br>[-0.16;0.01]                 | 0.04<br>[0;0.17]             | -0.02<br>[-0.06;0.02]          | 0.02<br>[0;0.06]               |
| Female                             | 5036<br>[328;16624]                  | 0.04<br>[-0.03;0.22]               | -0.04<br>[-0.17;0.01]                 | 0.04<br>[0;0.18]             | -0.02<br>[-0.08;0.03]          | 0.03<br>[0;0.1]                |
| <b>By permanent income tercile</b> |                                      |                                    |                                       |                              |                                |                                |
| Low                                | 5436<br>[353;18472]                  | 0.04<br>[-0.02;0.23]               | -0.05<br>[-0.17;0.02]                 | 0.05<br>[-0.01;0.19]         | -0.03<br>[-0.14;0.06]          | 0.02<br>[0;0.07]               |
| Medium                             | 5391<br>[410;17759]                  | 0.04<br>[-0.03;0.22]               | -0.05<br>[-0.16;0.01]                 | 0.04<br>[0.00;0.18]          | -0.02<br>[-0.06;0.02]          | 0.03<br>[0;0.09]               |
| High                               | 5322<br>[343;18099]                  | 0.04<br>[-0.02;0.24]               | -0.03<br>[-0.14;0.02]                 | 0.03<br>[0;0.12]             | -0.02<br>[-0.06;0.02]          | 0.02<br>[-0.01;0.07]           |
| <b>By age at diagnosis</b>         |                                      |                                    |                                       |                              |                                |                                |
| 25-45                              | 4798<br>[323;19325]                  | 0.02<br>[-0.01;0.09]               | -0.03<br>[-0.14;0.02]                 | 0.02<br>[-0.01;0.13]         | -0.01<br>[-0.07;0.05]          |                                |
| 45-65                              | 5753<br>[419;21556]                  | 0.03<br>[-0.01;0.21]               | -0.05<br>[-0.16;0.01]                 | 0.05<br>[0.00;0.17]          | -0.02<br>[-0.06;0.01]          |                                |
| 65+                                | 5761<br>[377;17554]                  | 0.04<br>[-0.04;0.26]               |                                       |                              |                                |                                |

**Table 3: Correlation Matrix: Causal Effects**

The table contains the correlation of the causal estimates across diseases. Medical expenses refer to the amount in EUR spent on the year of diagnosis. Excess mortality is the probability of dying within the three years following the diagnosis. Labor participation, disability, and nursing home are the probability of working, being disabled, or living in a nursing home, three years after the diagnoses. Log earnings refers to the logarithm of labor earnings.

---

|                     | Labor<br>Participation | Log<br>earnings | Medical<br>expenses | Disability | Mortality | Nursing<br>home |
|---------------------|------------------------|-----------------|---------------------|------------|-----------|-----------------|
| Labor Participation | 1                      |                 |                     |            |           |                 |
| Log earnings        | 0.11                   | 1               |                     |            |           |                 |
| Medical expenses    | -0.51                  | -0.11           | 1                   |            |           |                 |
| Disability          | -0.92                  | -0.20           | 0.55                | 1          |           |                 |
| Mortality           | -0.63                  | -0.23           | 0.60                | 0.74       | 1         |                 |
| Nursing home        | -0.39                  | -0.20           | 0.21                | 0.43       | 0.33      | 1               |

---

**Table 4:** State Variables

| State Variable | Values     | Interpretation                                     |
|----------------|------------|--|
| $a$            | $\geq 0$   | Assets. No borrowing                               |
| $p$            | $[0, 1]$   | Probability of dying next period                   |
| $m$            |            | Medical expenses (EUR)                             |
| $s$            | $[0, 1]$   | Probability of becoming disabled next period       |
| $d$            | $[0, 1]$   | Labor capacity                                     |
| $t_{dis}$      | $[25, 64]$ | Age when the individual became disabled            |
| $d_{dis}$      | $[0, 1]$   | Labor capacity when the individual became disabled |
| $\pi$          | $[0, 1]$   | Probability of moving to a nursing home            |

The table lists the state variables in the model.

**Table 5:** Calibrated Parameters

| Parameter         | Value   | Interpretation                  | Source/Note                            |
|-------------------|---|---------------------------------|--|
| $\rho$            | 5   | Risk aversion parameter         | <a href="#">Kvaerner (2022)</a>        |
| $r$               | 2.44%   | Return on savings               | <a href="#">Kvaerner (2022)</a>        |
| $\beta$           | $\frac{1}{1+r}$                                     | Discount rate                   | <a href="#">Kvaerner (2022)</a>        |
| $\bar{a}$         | EUR 20,000  | Bequest threshold               | <a href="#">Kvaerner (2022)</a>        |
| $\theta$          | 83.3  | Bequest intensity               | <a href="#">Kvaerner (2022)</a>        |
| SVL ( $\bar{u}$ ) | 2 million ( $\bar{u} \approx 1.3 \times 10^{-17}$ ) | Utility flow for being alive    | <a href="#">De Nardi et al. (2017)</a> |
| $\underline{y}$   | EUR 12,700  | Minimum income                  | Social minimum (2015)                  |
| $E_{NH}$          | EUR 43,545  | Total cost of nursing home care | Statistics Netherlands                 |
| $c_{NH}$          | EUR 30,481  | Consumption in the nursing home | 70% of $E_{NH}$                        |
| $\lambda_{dis}$   | 0.7   | Dis. insurance replacement rate | WGA wage-related benefit               |
| $\tau_a$          | 0.3   | Tax on assets                   | Tax rate in Box 3                      |
| $premium_t$       | EUR 1164  | Health insurance premium        | Avg. premium (2015)                    |
| $ded$             | EUR 375   | Health insurance deductible     | 2015 minimum deductible                |

The table lists the main calibrated parameters in the model. SVL stands for statistical value of life.

**Table 6: Welfare Gains Associated With Curing Diseases**

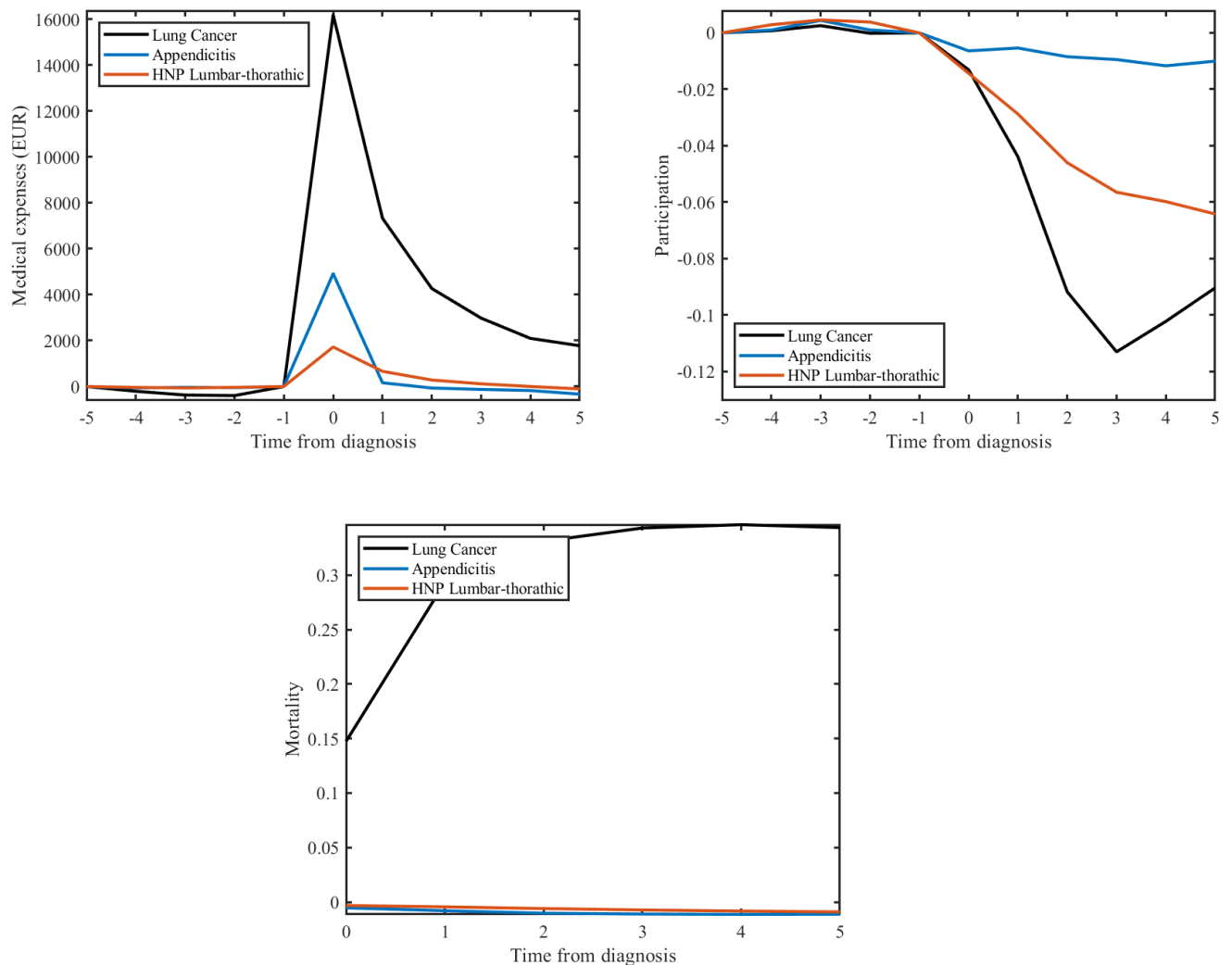
The first two panels of the table show the welfare gains from eliminating one disease of a 25-year-old individual. We measure welfare as the increase in the health insurance premium for which the lifetime welfare of individuals remains constant after curing one disease. Each column represents one of the diseases we consider. Each panel corresponds to a different gender and each row to a different permanent income tercile within the respective gender group. The last panel presents the surplus obtained by the government on average during the whole lifetime of an individual if health insurance premia are set to maintain the welfare of each group constant after curing the disease.

|                           | Lung Cancer | Appendicitis | HNP    |
|---------------------------|-------------|--------------|--------|
| <b>Female</b>             |             |              |        |
| Lowest PI                 | 282         | 6            | 157    |
| Medium PI                 | 454         | 6            | 120    |
| Highest PI                | 7,071       | 31           | 766    |
| <b>Male</b>               |             |              |        |
| Lowest PI                 | 438         | 2            | 198    |
| Medium PI                 | 3,266       | 8            | 375    |
| Highest PI                | 14,866      | 17           | 299    |
| <b>Government surplus</b> |             |              |        |
|                           | 254,888     | 9,256        | 26,313 |

## B Figures

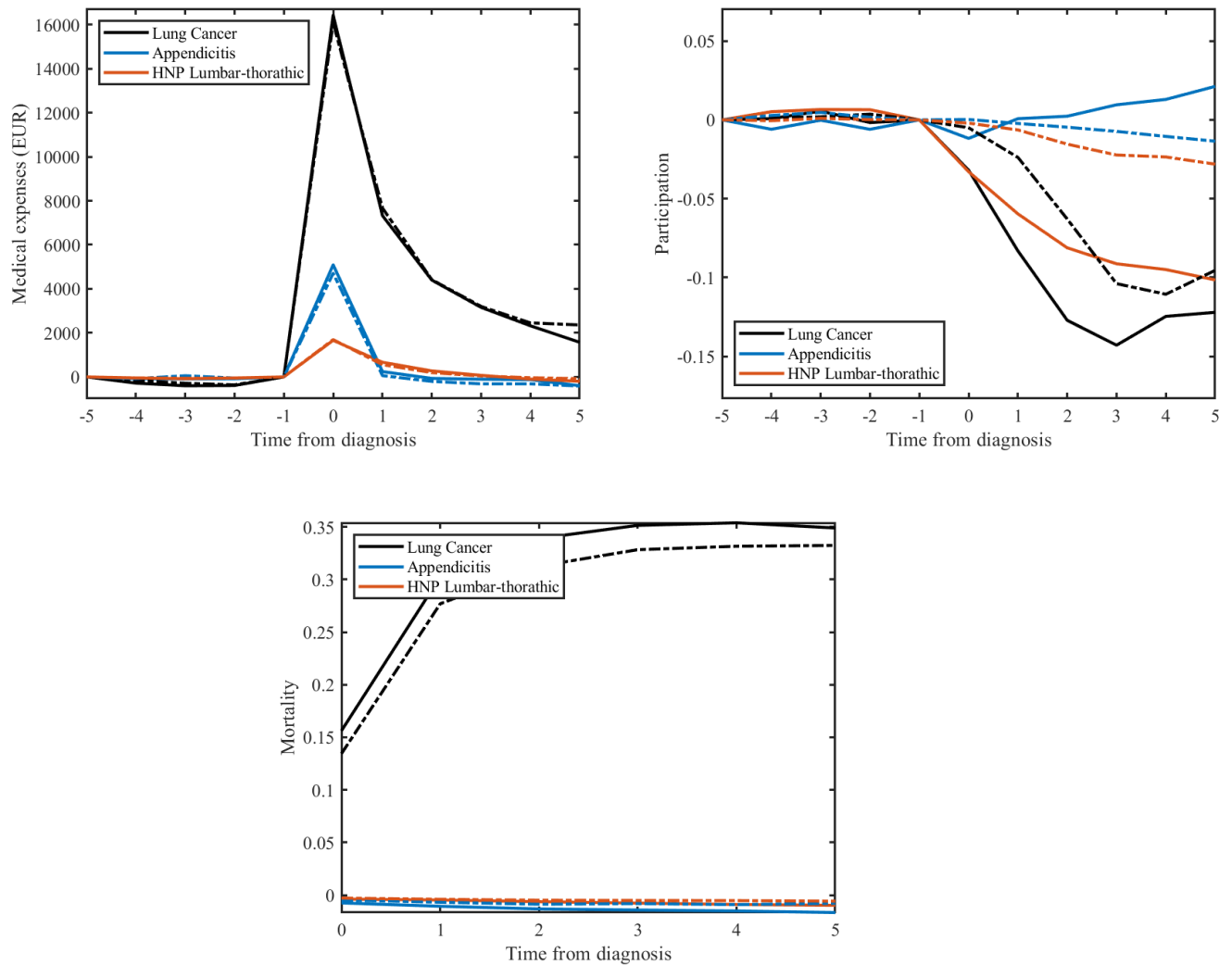
**Figure 1: Average Effects**

The y-axis shows the dependent variable, and the x-axis shows the years around the diagnosis year. The upper plot to the left shows the average medical expenses and the upper plot to the right shows the average labor market participation. The lower plot shows excess mortality. Each colored line corresponds to the diseases displayed in the graph.



**Figure 2: Heterogeneous Treatment Effects: Income**

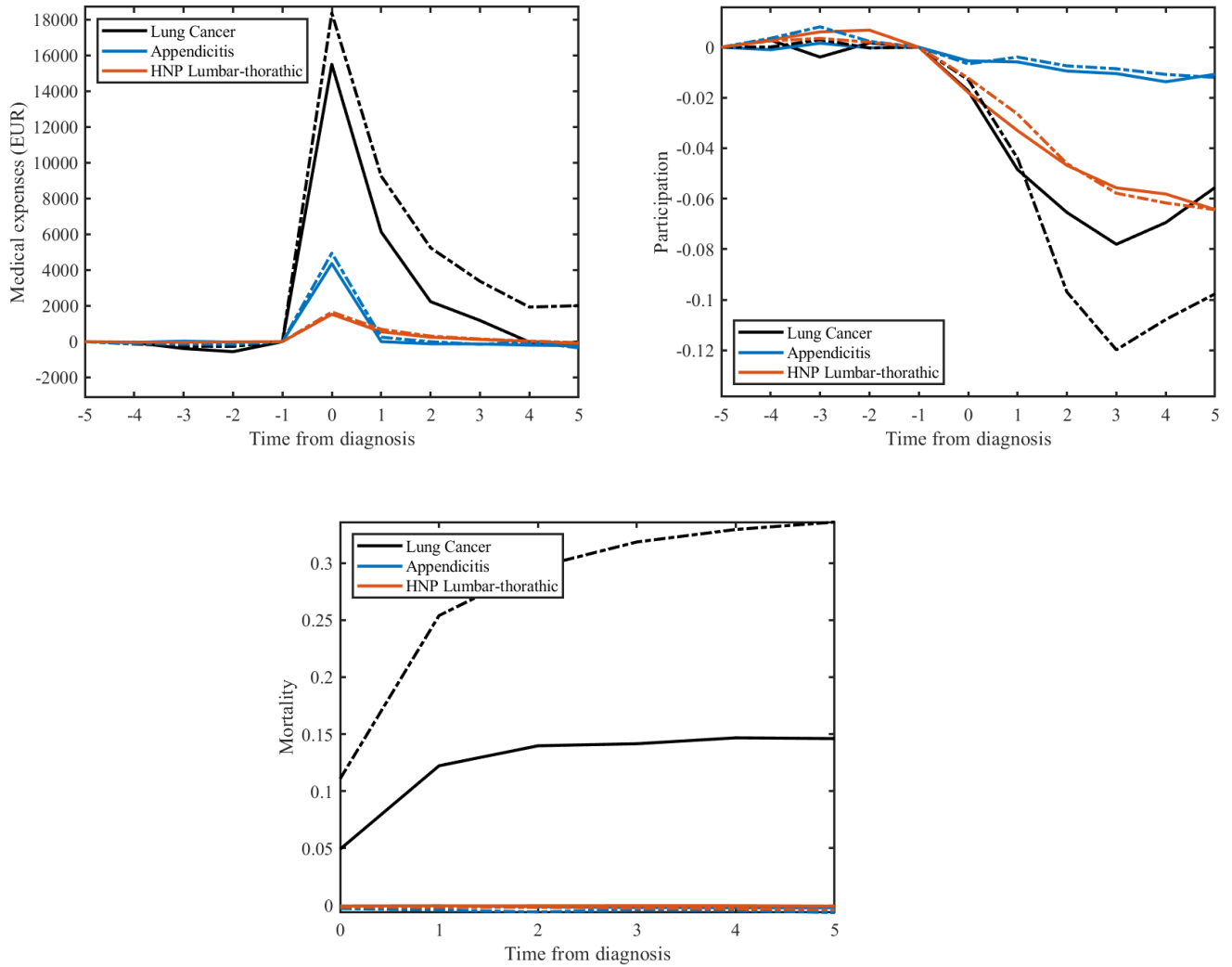
The y-axis shows the dependent variable, and the x-axis shows the years around the diagnosis year. The upper-left plot presents the effects on medical expenses, and the upper-right plot shows the effects on labor market participation. The bottom plot shows the effects on excess mortality. Each colored line corresponds to the health shock displayed in the graph. The solid line shows the effects for individuals in the lowest permanent income tercile, while the dashed line shows the effects for those in the highest permanent income.





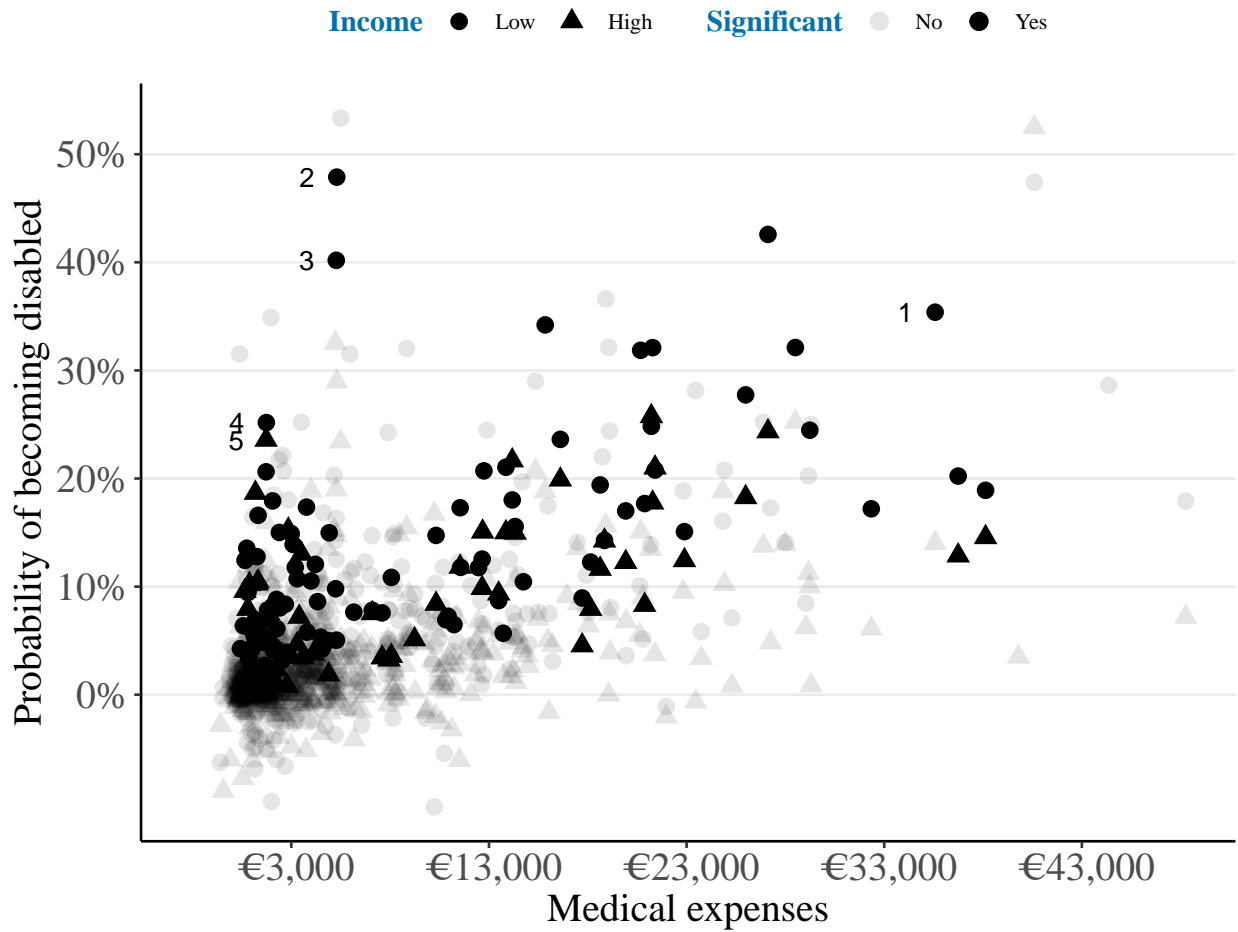
**Figure 3: Heterogeneous Treatment Effects: Age**

The y-axis shows the dependent variable, and the x-axis shows the years around the diagnosis year. The upper-left plot presents the effects on medical expenses, and the upper-right plot shows the effects on labor market participation. The bottom plot shows the effects on excess mortality. Each colored line corresponds to the health shock displayed in the graph. The solid line shows the effects for individuals diagnosed at a younger age (25-44), while the dashed line shows the effects for those diagnosed at an older age (45-64).



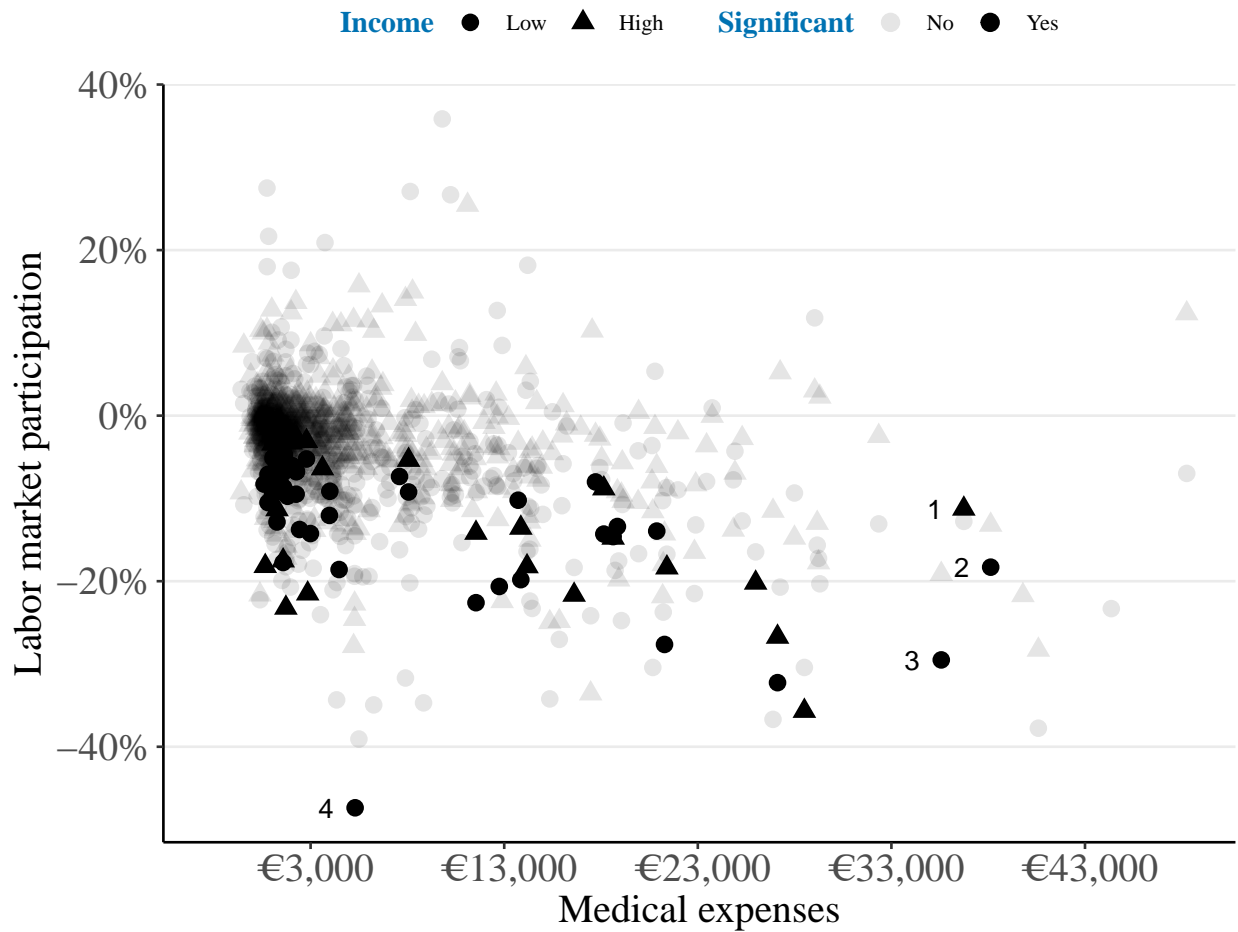
**Figure 4: Probability of Becoming Disabled**

The figure is a scatter plot with x-axis: cost of treating a particular disease and y-axis: the causal effect of that particular disease on the probability of becoming disabled. The numbered points correspond to the diagnoses (1) nontraumatic subarachnoid hemorrhage, female; (2)-(3) systemic atrophies affecting the central nervous system, female and male, respectively; (4)-(5) dementia, female and male, respectively.



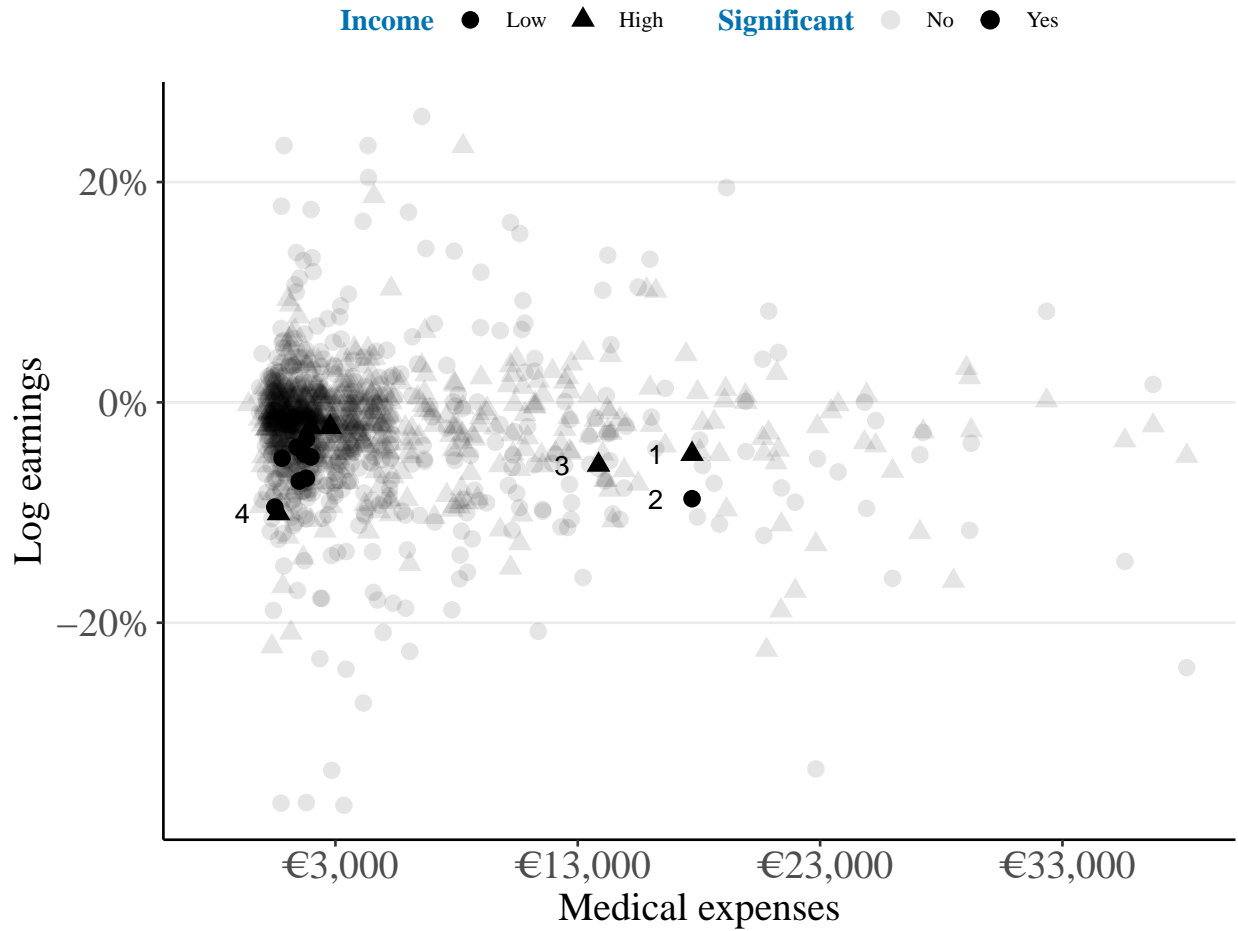
**Figure 5: Labor market participation**

The figure is a scatter plot with x-axis: cost of treating a particular disease and y-axis: the causal effect of that particular disease on labor market participation. The numbered points correspond to the diagnoses (1)-(2) malignant neoplasms of lymphoid, hematopoietic, and related tissue (e.g., leukemias, lymphomas), female and male, respectively; (3) nontraumatic subarachnoid hemorrhage, female; (4) systemic atrophies affecting the central nervous system, female.



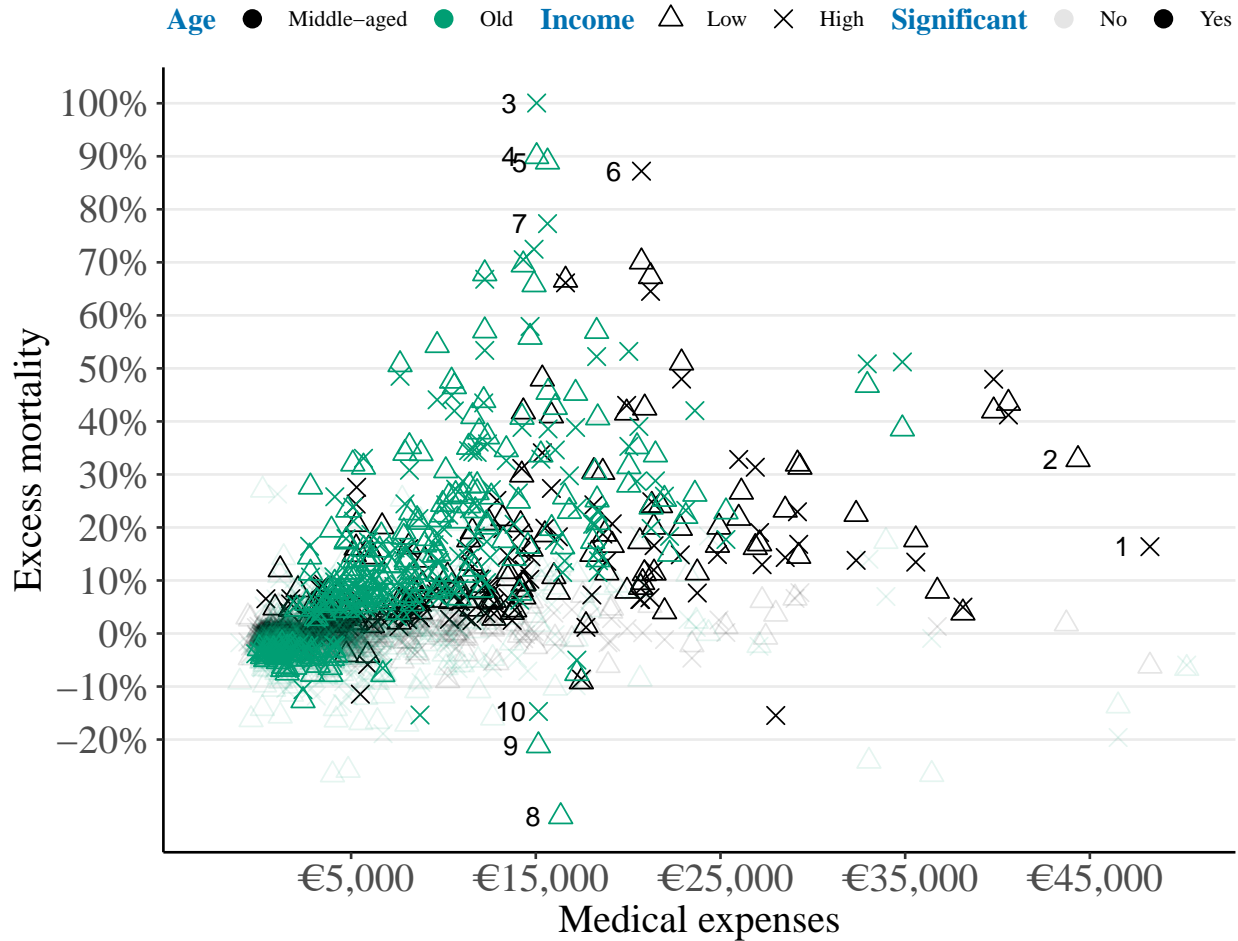
**Figure 6: Log earnings**

The figure is a scatter plot with x-axis: cost of treating a particular disease and y-axis: the causal effect of that particular disease on log earnings. The numbered points correspond to the diagnoses (1)-(2) breast cancer, female; (3) stroke, male; (4) mental and behavioral disorders due to psychoactive substance use, female.



**Figure 7: Excess mortality**

The figure is a scatter plot with x-axis: cost of treating a particular disease and y-axis: the causal effect of that particular disease on excess mortality. The numbered points correspond to the diagnoses (1) chronic rheumatic heart disease, male; (2) myelodysplastic syndromes, female; (3)-(4) mesothelioma, female; (5)-(7) mesothelioma, male; (8) chronic subdural hematoma, female; (9)-(10) chronic subdural hematoma, male.



**Figure 8: Probability of Needing Nursing Home**

The figure is a scatter plot with x-axis: cost of treating a particular disease and y-axis: the causal effect of that particular disease on excess mortality. The numbered points correspond to the diagnoses (1)-(2) dementia, female; (3) nontraumatic intracerebral hemorrhage, female; (4) acute kidney failure, male; (5) malignant neoplasms of lymphoid, hematopoietic, and related tissue (e.g., leukemias, lymphomas), female.

