

# The health effects of demand side cost sharing in European health insurance

Jan Boone\*

September 30, 2023

## Abstract

The rationale for demand side cost sharing in health insurance is to deter patients from using low value care. But if agents are cash constrained, demand side cost sharing can lead them to postpone or forgo valuable treatments. We use data on European (NUTS 2) regions to show that the interaction between poverty rate and out-of-pocket payments leads to unmet medical needs and higher mortality.

**JEL codes:** I11, I13, I18

**Keywords:** out-of-pocket payments, mortality, health insurance, poverty, unmet medical needs

---

\*Tilburg University, Department of Economics, Tilc and CEPR, E-mail: *j.boone@uvt.nl*.

# 1 INTRODUCTION

Most developed economies face rising healthcare expenditures. In many countries the healthcare sector grows faster than the economy as a whole (OECD 2021). One of the instruments that governments have to curb this expenditure growth is demand side cost sharing. The effect of demand side cost sharing on healthcare utilization is well known. As cost sharing increases, healthcare becomes more expensive for the individual and demand for treatments falls. It is less clear whether and to which extent demand side cost sharing induces people to forgo low value care only (Newhouse and the Insurance Experiment Group 1993; Schokkaert and van de Voorde 2011).

The traditional view is that health insurance subsidizes health consumption thereby inducing people to get (expensive) treatments with small health benefits. Economists tend to refer to this as moral hazard. As the social costs (in contrast to an individual's out-of-pocket –oop– expenditure) of such treatments exceed their value (utility gain), an increase in demand side cost-sharing that reduces moral hazard is seen as welfare enhancing. The traditional trade off is between this increase in efficiency (due to reduced moral hazard) and the increased oop risk faced by a risk averse agent.

Here we focus on behavioral hazard which refers to the case where cost-sharing leads patients to forgo valuable treatments (Baicker, Mullainathan, and Schwartzstein 2015). If a patient decides to skip a treatment where value exceeds costs then social welfare is reduced. In this paper, we concentrate on the case where people skip or postpone treatment because it is too expensive.

The goal of this paper is to develop a simple model that can be estimated with aggregate data to identify whether demand side cost sharing has negative health effects. In particular, we are interested in the mechanism where demand side cost sharing reduces health because valuable treatments become too expensive. We start from the following two ideas. First, if demand side cost sharing reduces valuable healthcare by making it (too) expensive, this effect will be stronger for people on low income. Health is a normal good and people with high (enough) income pay for valuable treatments even if they become expensive in terms of oop expenditure. Low income can force a patient to postpone or forgo treatment due to liquidity constraints. Second, if there is a substantial demand reduction for high value care, we should be able to detect this in aggregate mortality statistics.

To identify the health effects of cost-sharing we use mortality statistics of Eu-

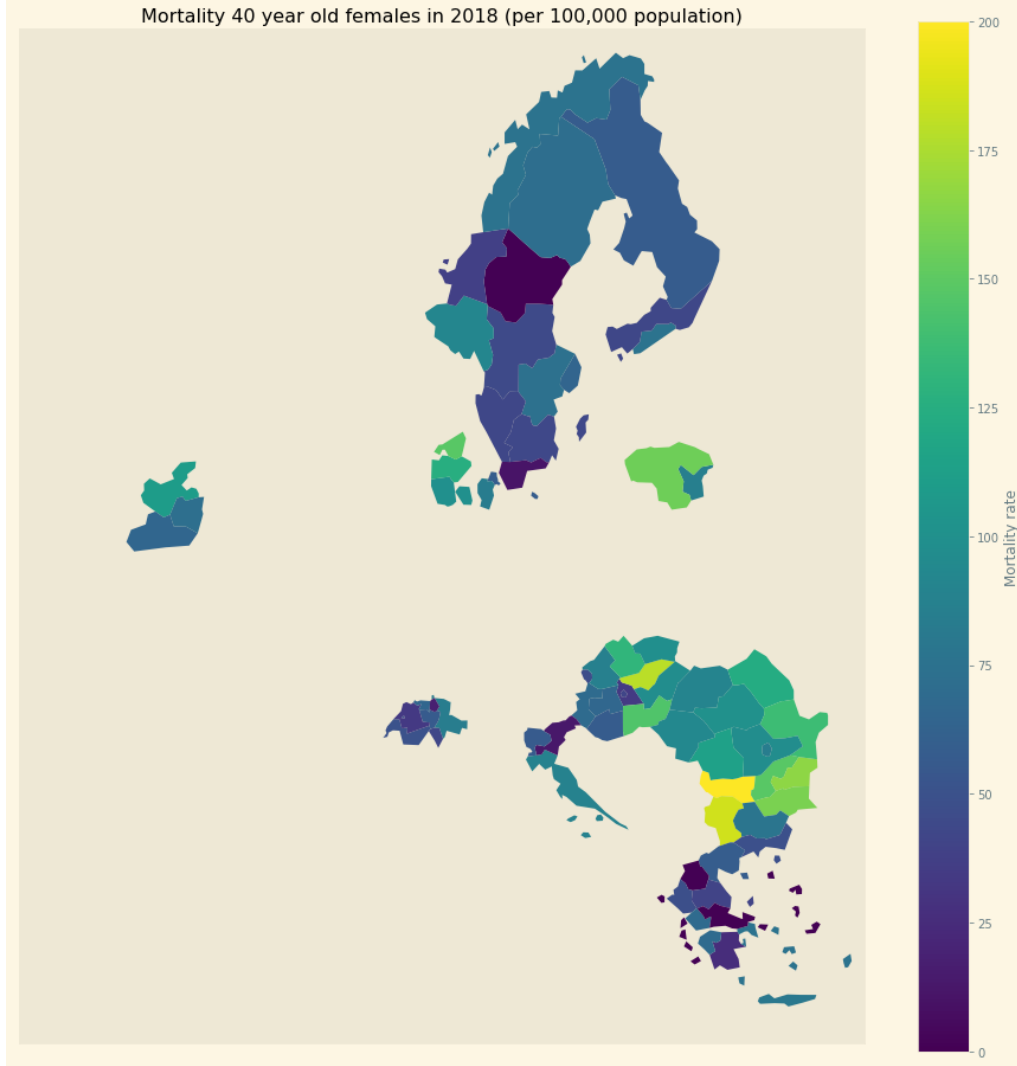


Figure 1: Mortality in NUTS 2 regions in Europe

rostat at the NUTS 2 (Nomenclature of Territorial Units for Statistics) regional level. Figure 1 illustrates NUTS 2 regions used in this paper. Mortality varies by region/year/age/sex. In regions where the percentage of people on low income is high and demand side cost sharing is high, we expect to see high mortality. Since we have panel data, we control for NUTS 2 fixed effects.

Measuring the generosity of a health insurance system is non-trivial. Systems tend to combine coinsurance with health expenditure caps, like a deductible. This leads to non-linearities in the oop price of healthcare. To address this, we introduce a model that links observed variables related to mortality, poverty, oop expenditures and people forgoing treatment because it is too expensive. The combination of the model and these variables allows us to identify the mechanism from reduced health insurance generosity via poverty to people forgoing treatment thereby raising

mortality.

Although details of the analysis follow below, Figure 2 presents the following summary of the results. For each country in our data, we consider the NUTS 2 region where poverty is highest and therefore the effect likely to be the strongest at the regional level. Using the estimated model, we simulate the effect of a 500 euro increase in oop on mortality. We report this effect as the increase in deaths (due to the increase in oop) per 1000 dead. The motivation for this measure is two-fold. First, mortality is –thankfully– low and hence the effect of a change in oop on mortality is going to be (very) small. Reporting the increase in mortality per 1000 dead helps to interpret the numbers. Below we also present this measure for diseases that have similar orders of magnitude, like pneumonia. Second, in our model this measure (per 1000 dead) is age-independent. That is, the number of people dying due to an increase in oop varies with age (as 25 year olds are less likely to die than 80 year olds). But the fraction of people dying due to the oop increase as a fraction of the total number of deceased is the same across age (and gender). This formulation reduces the number of parameters that we need to estimate and fits the data rather well.

The blue bars indicate the average simulated effect of the 500 euro increase for this region within each country; the black lines indicate the 95% probability interval of the effect. The four countries with the biggest effects –Bulgaria, Greece, Hungary and Romania– have the highest poverty levels in our sample. For these countries we can easily see that the 95% probability interval of the effect is bounded away from 0. For the Scandinavian countries, Slovenia and Switzerland the effects are close to zero at the region level because poverty is very low (even in the NUTS 2 region with highest poverty level per country).

The results suggest the following policy implications. An increase in oop has a measurable/significant effect on mortality in regions where poverty is high. Policies to address this include a scheme that subsidizes healthcare expenditure (on top of health insurance) for poor people; e.g. through means-tested cost-sharing. A downside of such a targeted intervention is a higher marginal tax rate at low income levels contributing to a poverty trap. Indeed, if by earning more, the oop subsidy falls, the increase in net income is reduced. This makes such an increase in income less attractive. Alternatively, a government can introduce co-payments that vary with the cost-effectiveness of the treatment. Treatments with high value added would then feature a low co-payment to prevent people from postponing valuable care. This can also help to reduce mortality associated with cost sharing (Chernew

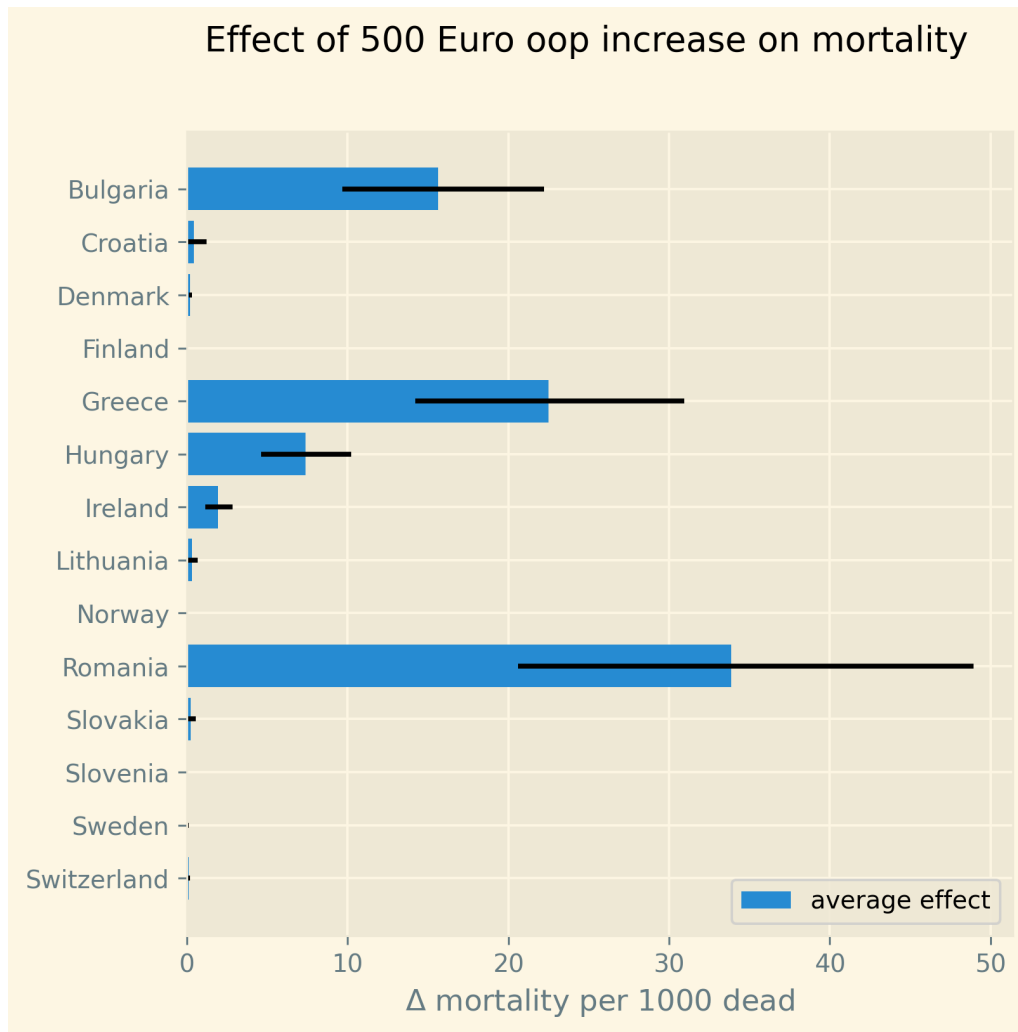


Figure 2: Increase in number of deaths per 1000 dead due to a 500 euro increase in out-of-pocket payment for the region in each country where poverty is highest. Bars present the average predicted effect and black lines the 95% prediction interval.

et al. 2008).

This is not the first paper to consider the effects of demand side-cost sharing on mortality. There is a string of recent papers using innovative methods and (mostly) individual level data to identify the causal effect of health insurance on health and mortality. There are a number of issues when identifying this effect of health insurance on health and mortality using individual level data. First, mortality is a rare event at most ages. Hence, identifying the effect is difficult, especially if the changes in oop are small. Second, there is the selection effect that people with low health status tend to buy (generous) health insurance (as they expect high expenditure). This can bias results in the direction that individuals with (generous) insurance tend

to have adverse health outcomes (e.g. high mortality). Moreover, cost-sharing tends to be non-linear with e.g. a cap on expenditures that have to be paid oop as with a deductible. In this case, people with high care use tend to face low (marginal) treatment prices; for example, because they filled up their deductible. As low health status is likely to increase care use, people facing low marginal prices (suggesting generous insurance) are likely to have experienced adverse health outcomes. This confounds the effect of insurance on health and mortality.

A number of papers use the Medicaid eligibility expansion of the Affordable Care Act which was introduced in different US states at different times. This allows for a diff-in-diffs identification strategy. Using individual level data, these papers have shown that the Medicaid expansion (more generous health insurance coverage) reduced mortality (Borgschulte and Vogler 2020; Miller, Johnson, and Wherry 2021). Others analyze Medicare part D prescription drug coverage where the end-of-year price is non-linear in expenditure. One paper uses enrollment month (related to birth month) to get exogenous variation in end-of-year expenditure for people aged 65 (Chandra, Flack, and Obermeyer 2021). The main finding is that increases in the oop costs of drugs, reduce drug use including use of high value treatments. This, in turn, raises mortality. By using exogenous exit of plans in the Medicare Advantage market to control for endogeneity problems, it is possible to show that more generous prescription drug coverage leads to lower mortality (Abaluck et al. 2020). Finally, Goldin and coauthors use an experiment where a subset of people who should buy health insurance under the Affordable Care Act were reminded that they would face a financial penalty if they did not comply. This reminder tended to induce people to buy insurance instead of remaining uninsured (Goldin, Lurie, and McCubbin 2020). Mortality turns out to be lower among the people who received the reminder compared to the control group who were not reminded in this way.

Compared to these papers on health insurance (generosity) and mortality, our paper differs along the following lines. First, we use European instead of US data. The advantage is that within a European country health insurance is more homogeneous than in the US. Within a US state or county, people may have generous employer sponsored insurance, benefit from Medicaid or Medicare or have no insurance at all. Hence, a change in Medicaid coverage may have no detectable effects at the aggregate level (while an effect can be found with individual level data). In European countries a number of health insurance features are determined nationally. Consider the first two rows of the OECD Health Systems Characteristics Survey (<https://qdd.oecd.org/data/HSC>) showing the share of the population obtaining

basic primary health care coverage through automatic or compulsory insurance coverage. For all European countries this is above 90% and for most 99% or 100%. For the US this is less than one third. Hence, country or region wide statistics in Europe give a better picture of the situation applying to most citizens in that region than in the US.<sup>1</sup>

Second, we show that mortality is high in regions where both oop and poverty are high. This follows the literature showing that healthcare consumption is liquidity sensitive (Gross, Layton, and Prinz 2020; Nyman 2003). People on low income tend to postpone or forgo valuable treatments if these are expensive. This focus on low incomes can imply that we under-estimate the mortality effect of cost-sharing if higher incomes also forgo valuable treatments due to oop (Brot-Goldberg et al. 2017; Chandra, Flack, and Obermeyer 2021). This is then not so much caused by liquidity problems but by other forms of behavioral hazard. In this sense, the results below are a lower bound on the mortality effects of cost-sharing.

Third, we use the regional structure of the Eurostat data. We analyze the effects of our oop variable times poverty interaction on mortality per age-gender class at the NUTS 2 regional level. This helps to solve the following potential endogeneity issue. A country with a population that has low health status (across ages), decides to have, say, generous health insurance to help people improve their health. This causal effect is in the opposite direction from the one we are interested in. We avoid this problem by considering within a country how health per region varies with oop and poverty, while using NUTS 2 fixed effects to correct for other factors affecting health. By analysing health/mortality per age cohort, our results are not affected by a country’s or region’s age distribution. By filtering out these other effects we mitigate power issues associated with the use of mortality data at the regional level (Black et al. 2021). We have a separate section below discussing potential confounding effects when using regional data.

Fourth, Eurostat variables based on the EU-SILC survey allow us to zoom in on the relevant causal mechanism. This survey asks people whether they had unmet medical needs in the past months and if so the reason for the unmet needs. One of the answers is that treatment was postponed or skipped because it was too expensive. This allows us to simultaneously estimate the fraction of people in a NUTS 2 region

---

<sup>1</sup>This does not imply that the aggregate statistics perfectly represent everyone’s insurance situation (e.g. some people may buy complementary insurance where others do not), but it may be representative enough to identify the interaction effect of poverty and oop payments we are interested in.

that forgo treatment because it is too expensive and the effect of unmet medical needs on mortality. In this way, we capture that in regions where the  $\text{oop} \times \text{poverty}$  interaction is high, more people postpone treatment because it is too expensive and these unmet medical needs raise mortality in the region.

Finally, our focus on the  $\text{oop} \times \text{poverty}$  interaction distinguishes our paper from the literature on the effect of income and wealth on health (Chetty et al. 2016; Mackenbach et al. 2008; Semyonov, Lewin-Epstein, and Maskileysen 2013) where papers use cross country data. This literature typically finds that lower income and wealth is associated with lower health status, although the causal mechanism is not clear (Cutler, Lleras-Muney, and Vogl 2011). Two possible mechanisms are that higher income leads to more expenditure on treatments (normal good) and therefore better health. Alternatively, healthier people are more productive and earn higher incomes. The combination of fixed effects and the use of the survey question on unmet medical needs allows us to zoom in on the mechanism where high  $\text{oop} \times \text{poverty}$  interaction leads to unmet medical needs and hence to low health status and high mortality.

Summarizing, compared to papers using individual level data our approach is more broad brush and less precise in estimating the size of the effect of insurance generosity on mortality. To illustrate, we do not aim to determine the mortality effect of a 1% change in a deductible. We estimate the mortality effect of a 500 euro increase in oop. We use the fraction of oop payments in total healthcare expenditure, OOP, as a summary measure of a health insurance system’s generosity. The theory section derives that OOP and the fraction of people postponing treatment because it is too expensive are parametric functions of the underlying “generosity parameters” coinsurance rate and deductible level. This derivation allows us to interpret the relation between OOP and mortality.

Although results based on aggregate data are less precise than those based on individual level data, our approach is more robust in the sense that it applies across a number of countries instead of a particular sub-population (like 65 year old Medicare users in the US). Although we do interpret our results using the size of the effect, our main goal is to establish that an increase in OOP in a poor region increases mortality. In particular, we quantify how sure we are that this effect is positive.

The next section presents a model explaining the relationship between the variables mortality, poverty, OOP and the fraction of people forgoing treatment because it is too expensive. Then we describe the Eurostat data that we use. We explain the empirical model that we estimate. Estimation results are presented for the baseline



model and we show that these are robust with respect to a number of our modeling choices. We conclude with a discussion of the policy implications. The appendix contains the proofs of our results and more details on our data and estimation. The online appendix is the html version of this paper which includes –per section– the python code that is used in each section’s analysis.<sup>2</sup> This is a final advantage of using data at the regional level. The repository contains the python code that gets the data from Eurostat so that each step of this analysis can be replicated. The data used for this paper can be downloaded from DataverseNL (Boone 2022).

## 2 THEORY

The relevant variables in our data are mortality per region/year/age/sex category, OOP measuring the percentage of healthcare expenditure paid out-of-pocket (oop), the poverty rate and the fraction of people per region postponing or forgoing treatment because it is too expensive. We introduce a model to explain how these variables are related.

### 2.1 *simple model*

Consider a population (of a certain age and gender in a particular year) in an EU region where a fraction  $\alpha \in \langle 0, 1 \rangle$  has low income  $y^l$  and fraction  $1 - \alpha$  high income  $y^h$ . Let  $\pi^j$  denote the probability that someone with income  $y^j, j = l, h$  falls ill. As is well known, low income people tend to have a lower health status (Cutler, Lleras-Muney, and Vogl 2011). We capture this by assuming  $\pi^l > \pi^h$ . People on low income may have a less healthy diet, exercise less etc. due to either the cost or knowledge of healthy lifestyle choices. This makes it more likely that they fall ill.

Generally speaking, oop payments tend to take two forms that we want to capture: a coinsurance rate, which we denote  $\xi \in [0, 1]$ , and a maximum expenditure, which we denote  $D$  (for deductible). Some systems have a combination of the two.

Conditional on falling ill, there is a probability  $\zeta_i \in [0, 1]$  that the patient is advised to get treatment  $i$  at cost  $x_i$  for  $i$  in the set of “illnesses”  $I$  with  $\sum_{i \in I} \zeta_i = 1$ .<sup>3</sup> We define  $I_\xi$  as the subset of  $I$  where  $\xi x_i < D$  and  $oop_i = \xi x_i$  and set  $I_D$  where  $\xi x_i \geq D$  and  $oop_i = D$ . To keep things simple, we assume that  $\zeta_i$  is exogenous

---

<sup>2</sup>See the github repository: [https://github.com/janboone/out\\_of\\_pocket\\_payments\\_and\\_health](https://github.com/janboone/out_of_pocket_payments_and_health).

<sup>3</sup>We think of  $I$  as encompassing treatment for every disease and combination of diseases.

to the patient. We model the treatment decision on the extensive margin only: an agent accepts or rejects the treatment proposed by a physician.<sup>4</sup> A pure coinsurance system has  $\xi < 1$  and  $I_\xi = I$ . A pure deductible system  $\xi = 1$  and  $I_D$  non-empty. A combination of the two has  $\xi < 1$  and there is a maximum on the oop payment. Health insurance systems in Europe tend to have such maximum oop expenditure.<sup>5</sup> An increase in either  $\xi$  or  $D$  is interpreted as making health insurance less generous.

Whereas with individual level data one can determine whether an individual faces a positive treatment price at the margin (E.g. using the end-of-year price as in Keeler, Newhouse, and Phelps 1977; Ellis 1986), this is not possible with the aggregate data that we use here. Hence, we rely on an aggregate summary variable, denoted OOP, measured as oop payments over total healthcare expenditure. That is, the fraction of healthcare expenditure paid by patients oop. We interpret this variable as capturing the generosity of the health insurance system. To illustrate, if healthcare is free at point of service, OOP equals zero; if there is no health insurance at all, OOP equals 1. In a pure coinsurance system with rate  $\xi$  applying to all treatments, OOP equals  $\xi$ . It is the cap on oop expenditure (like a deductible) that makes the relation between OOP and healthcare use non-linear. The challenge then is to capture changes in  $\xi$  and  $D$  although we do not directly observe these variables in the data. This is what the model sets out to do.

If a patient receives treatment  $i \in I$ , we denote her (expected) health  $\sigma_i$ , while without treatment (expected) health equals  $\sigma_0$  with  $0 \leq \sigma_0 < \sigma_i \leq 1$ .<sup>6</sup> Health is normalized at value one for a patient who does not fall ill. The trade off between health and oop is captured by  $\sigma_0/\sigma_i < 1$  and we simply assume that utility is multiplicative in health and consumption. That is, consumption yields higher utility if you are healthier. We model the patient's treatment decision as:

$$\nu \sigma_i u(y^j - oop_i) > \sigma_0 u(y^j) \quad (1)$$

where utility  $u(\cdot)$  is determined by how much money can be spent on other goods: income  $y^j$  minus oop in case of treatment and  $y^j$  if no treatment is chosen. The utility function  $u(\cdot)$  is increasing and concave in consumption:  $u(\cdot), u'(\cdot) > 0$  and  $u''(\cdot) < 0$ . Further, parameter  $\nu$  captures other factors than pure financial ones

---

<sup>4</sup>A further simplification is that we do not analyze dynamic incentives like: accepting this treatment fills up my deductible which makes future treatment (weakly) cheaper for me.

<sup>5</sup>See question 12 in <https://qdd.oecd.org/data/HSC> specifying for most European countries a spending cap.

<sup>6</sup>To ease notation we do not let  $\sigma_0$  vary with  $i$ .

affecting a patient's treatment choice.<sup>7</sup> If the inequality holds, the patient accepts treatment  $i$ .

In our data, we have a variable “unmet medical needs” based on a number of motivations: treatment is too far away to travel to, there is a long waiting list, the patient is scared to undergo treatment etc. To make our point, it is enough to assume that such factors affect utility in a simple multiplicative way. To illustrate, if the patient has to travel far for treatment, utility is reduced by multiplying it with a value of  $\nu < 1$ . Agents differ in  $\nu$  and the cumulative distribution function of  $\nu$  is given by  $G(\nu)$  and its density function by  $g(\nu)$ . Other factors can include waiting time till treatment, belief that the condition will resolve itself without intervention, poor decision making e.g. with a focus on the short term thereby undervaluing the benefit of treatment.

The probability that a patient with income  $y^j$  accepts treatment  $i$  offered by a physician equals

$$\delta_i^j = 1 - G\left(\frac{\sigma_0}{\sigma_i} \frac{u(y^j)}{u(y^j - oop_i)}\right)$$

that is,  $\nu$  is big enough that inequality (1) holds. With probability  $G\left(\frac{\sigma_0}{\sigma_i} \frac{u(y^j)}{u(y^j - oop_i)}\right)$  the patient decides to postpone or forgo treatment  $i$ .

The probability that a patient postpones or skips a treatment because it is too expensive is given by

$$G\left(\frac{\sigma_0}{\sigma_i} \frac{u(y^j)}{u(y^j - oop_i)}\right) - G\left(\frac{\sigma_0}{\sigma_i}\right) \quad (2)$$

These are agents  $\nu$  that would have chosen treatment  $i$  if it were free ( $oop_i = 0$  and  $u(y^j)/u(y^j - oop_i) = 1$ ) but who forgo treatment now that it costs  $oop_i > 0$ . The probability  $G(\sigma_0/\sigma_i)$  captures factors like waiting lists or the patient hoping that the health problems resolve themselves without treatment. That is, reasons for postponing treatment not related to oop payments.

In the proof of the lemma at the end of this section, we show that the probability of accepting treatment,  $\delta_i^j$ , is increasing in income  $y^j$  and decreasing in  $oop_i$ , as one would expect.

Note that this model differs from a standard Rothschild and Stiglitz –R&S– health insurance model (Rothschild and Stiglitz 1976) in the following way. In an R&S model income plays no role and people with low health status have generous insurance coverage. Hence, they would not postpone valuable care. In our model,

---

<sup>7</sup>Note that we do not model the decision to buy insurance. In Europe (almost) all citizens are covered by automatic or mandatory insurance.

people with low health tend to have low income and may skip valuable treatment if the oop expense is high. This negatively affects their health.

An agent's health is affected by the probability of falling ill and then getting treatment (or not). Based on the model, we specify that agents' mortality is affected by health in the following way, where we define mortality  $m$  as the probability of dying in a given period.

$$\ln(m_{agt}) = \ln(\eta_{ag}) + \gamma \ln \left( \frac{m_{a-1,g,t-1}}{\bar{m}_{a-1,g}} \right) - (\alpha(1 - \pi^l) + (1 - \alpha)(1 - \pi^h)) \quad (3)$$

$$- \alpha \pi^l \sum_{i \in I} \zeta_i (\delta_i^l \sigma_i + (1 - \delta_i^l) \sigma_0) - (1 - \alpha) \pi^h \sum_{i \in I} \zeta_i (\delta_i^h \sigma_i + (1 - \delta_i^h) \sigma_0)$$

where we use the following subscripts: age  $a$ , gender  $g \in \{f, m\}$ , calendar year  $t$ . In words, log mortality in a region depends on the biology of age and gender, captured by fixed effects  $\eta_{ag}$ . As people get older, they tend to become less healthy and are more likely to die. We define this effect as independent of country or year (in the period that we analyze). Then there are a number of effects that increase or decrease mortality in a region compared to  $\eta_{ag}$ .

The health of the age-gender cohort in the previous period: if in a NUTS 2 region there was a shock in  $t - 1$  –when this cohort was aged  $a - 1$ – that increased mortality above the average (across years and regions) mortality for this cohort, we interpret this as a negative health shock. For the people that survived in this cohort in this region, this health shock can affect their mortality in period  $t$ . This is captured by the coefficient  $\gamma$ .<sup>8</sup>

People who do not fall ill ( $\alpha(1 - \pi^l) + (1 - \alpha)(1 - \pi^h)$ ), have the highest health level (normalized to 1) and hence reduce mortality to the biggest extent. People who do fall ill with  $i$  and get treatment ( $\alpha \pi^l \zeta_i \delta_i^l$  and  $(1 - \alpha) \pi^h \zeta_i \delta_i^h$ ), get health  $\sigma_i \leq 1$  and reduce mortality to a smaller extent. Finally, people falling ill but forgoing treatment lead to the smallest reduction  $\sigma_0$  in mortality. Because poor people tend to have lower health status ( $\pi^l > \pi^h$ ) and more unmet needs ( $\delta_i^l < \delta_i^h$ ), mortality is higher for this group.

As we show in the proof of the lemma below, we can write the expression for log mortality as:

$$\ln(m_{ag2t}) = \ln(\eta_{ag}) + \mu_2 + \gamma \ln \left( \frac{m_{a-1,2,g,t-1}}{\bar{m}_{a-1,g}} \right) + \beta_{poverty} \alpha_{2t} + \beta_{unmet} \text{Unmet}_{2t} \quad (4)$$

---

<sup>8</sup>Although we think of  $\gamma > 0$ , we allow for  $\gamma < 0$ . The interpretation in the latter case would be that some people with low health status in cohort  $a - 1$  passed away early, increasing average health for people remaining in this cohort.

where subscript 2 indicates that the variable varies with NUTS 2 region,  $\mu_2$  denotes NUTS 2 fixed effects, poverty  $\alpha$  varies with NUTS 2 region and calendar year and **Unmet** denotes the fraction of people indicating unmet medical needs in a region in year  $t$ . The model implies  $\beta_{poverty}, \beta_{unmet} > 0$ : we expect mortality (in a region in year  $t$ ) to be higher (for a given age/gender category) if poverty is higher and there are more people with unmet medical needs.

In our data, the variable **Unmet** varies with NUTS 2 region and year and not by age or gender. Hence, in terms of our model, we define this variable as follows:

$$\text{Unmet}_{2t} = \sum_{i \in I} \zeta_i (\alpha_{2t} \pi^l (1 - \delta_{ict}^l) + (1 - \alpha_{2t}) \pi^h (1 - \delta_{ict}^h)) \quad (5)$$

with treatment probability  $\delta_i^j$  for illness  $i \in I$  and income class  $j \in \{l, h\}$  varying with country  $c$  and year  $t$  because oop varies with countries over time.

Further, in our data we have the variable **OOP** defined as oop payments as a percentage of healthcare expenditure. In terms of our model, we write this as

$$\text{OOP} = \frac{\sum_{i \in I} \zeta_i \text{oop}_i (\alpha \pi^l \delta_i^l + (1 - \alpha) \pi^h \delta_i^h)}{\sum_{i \in I} \zeta_i x_i (\alpha \pi^l \delta_i^l + (1 - \alpha) \pi^h \delta_i^h)}$$

where  $\zeta_i (\alpha \pi^l \delta_i^l + (1 - \alpha) \pi^h \delta_i^h)$  denotes the fraction of people accepting treatment  $i$ . If people do not accept treatment, there is no oop and no expenditure. The numerator of **OOP** contains the oop payments  $\text{oop}_i$  and the denominator expenditures  $x_i$ . If  $I_\xi = I$ , it is clear that  $\text{OOP} = \xi$ . Because  $I_D$  is non-empty (European countries have a maximum oop payment), the expression for **OOP** is actually non-trivial. We can also write **OOP** as the ratio of average oop per head and average healthcare expenditure per head:

$$\text{OOP}_{ct} = \frac{\overline{\text{oop}}_{ct}}{\bar{x}_{ct}} \quad (6)$$

In our data these variables vary by country and year.

Finally, using equation (2) our model allows us to formalize the fraction of people that forgo treatment because it is too expensive: fraction of poor people who need treatment,  $\alpha \pi^l$ , forgoing treatment for illness  $i \in I$  because it is too expensive plus the fraction of rich people,  $(1 - \alpha) \pi^h$ , forgoing treatment for this reason:

$$\begin{aligned} \text{TooExp} = & \alpha \pi^l \left( \sum_{i \in I} \zeta_i \left( G \left( \frac{\sigma_0}{\sigma_i} \frac{u(y^l)}{u(y^l - \text{oop}_i)} \right) - G \left( \frac{\sigma_0}{\sigma_i} \right) \right) \right. \\ & \left. + (1 - \alpha) \pi^h \left( \sum_{i \in I} \zeta_i \left( G \left( \frac{\sigma_0}{\sigma_i} \frac{u(y^h)}{u(y^h - \text{oop}_i)} \right) - G \left( \frac{\sigma_0}{\sigma_i} \right) \right) \right) \end{aligned} \quad (7)$$

In our data, **TooExp** varies with NUTS 2 region and year. The following lemma summarizes the main results from the model and presents the equations that we estimate below. The innovation is to view equations (6) and (7) as being parametrized by the underlying parameters  $\xi$  and  $D$  which are not directly observed in our data. We prove that this leads to an equation where **TooExp** is a function of OOP and poverty.

**Lemma 1** *Healthcare demand  $\delta = 1 - G(\cdot)$  is increasing in income  $y^j$  and decreasing in  $\text{oop}_i$  ( $\xi$  or  $D$ ). We write the expression for mortality of age cohort  $\mathbf{a}$  and gender  $\mathbf{g}$  in NUTS 2 region  $\mathbf{2}$  at time  $\mathbf{t}$  as:*

$$m_{ga2t} = \frac{e^{\beta_{ag}}}{1 + e^{\beta_{ag}}} e^{\left( \mu_2 + \gamma \ln \left( \frac{m_{a-1,g,2,t-1}}{\bar{m}_{a-1,g}} \right) + \beta_{poverty} Poverty_{2t} + \beta_{unmet} Unmet_{2t} \right)}$$

where  $\beta_{poverty}, \beta_{unmet} > 0$ . The linear expansion of **TooExp** with respect to OOP can be written as

$$TooExp_{2t} = b_{0,2} + b_{0,t} + OOP_{ct} \bar{x}_{ct} (b_{oop,c} + b_{interaction,c} Poverty_{2t})$$

where  $b_{oop}, b_{interaction} > 0$ .

We model  $\eta_{ag}$  in equation (3) as a sigmoid of age and gender fixed effects,  $\beta_{ag}$ . This makes sure this part of the probability of death is between 0 and 1. We multiply this baseline probability with a multiplier capturing the other effects. In particular, NUTS 2 region fixed effects,  $\mu_2$ , which capture regional variation in the probability of falling ill; whether this age cohort experienced a health shock in the previous period; the region's poverty level and the fraction of people with unmet medical needs in the region in year  $t$  all affect mortality. If the sum of these terms is negative, the multiplier is less than 1 and mortality for this gender/age/region/year combination is reduced compared to the baseline probability given by the sigmoid. If the sum of the terms is positive, mortality for this observation is higher than the baseline probability.

We use a linear expansion of **TooExp** in terms of OOP. The appendix shows how we derive this relation using the policy variables  $\xi$  and  $D$  which affect OOP and **TooExp** simultaneously. It turns out that there is a direct effect of OOP on **TooExp** and an interaction effect with the fraction of people below the poverty line in a region. We show that  $b_{oop}, b_{interaction} > 0$ : a region that lies in a country with high OOP tends to have high unmet needs and especially so if the region features a high poverty rate. As explained in the proof of the lemma, the linear expansion of **TooExp** in OOP and

$\text{OOP} \times \text{Poverty}$  interaction does not determine the intercept  $b_0$ . Therefore, we allow the intercept to vary by region and year:  $b_{0,2} + b_{0,t}$ .

Figure 3 illustrates this approximation of the relation between (log-odds) **TooExp** and **OOP** for simulated values in the model above. We simulate data for a country with varying values for  $\xi$  and  $D$ . Then both **OOP** and expenditure per head vary leading to the graph in the left panel of Figure 3 (see web appendix for details). For this simulated data, the approximation where the (log odds of) fraction of people forgoing treatment because it is too expensive depends linearly on  $\text{OOP} \times \text{Poverty}$  seems reasonable. As shown in the proof of the lemma, we need to multiply **OOP** and  $\text{OOP} \times \text{Poverty}$  by healthcare expenditure per head because the underlying changing variable is not the endogenous **OOP** but the parameters  $\xi$  and  $D$ . As illustrated in equation (6), the relation between changes in  $D$  and **OOP** is multiplied by expenditure per head:  $d\text{OOP}/dD \propto 1/\bar{x}_{ct}$ .

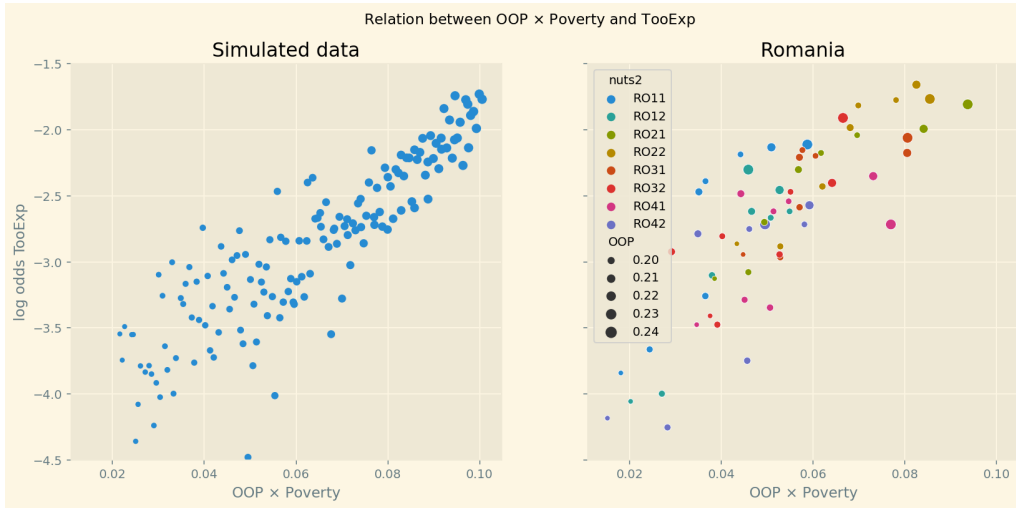


Figure 3: The simulated relation between fraction of people who forgo treatment because it is too expensive and poverty times the **OOP** measure for different values of  $\xi$ ,  $D$  (left panel) and this relation for NUTS 2 regions and years in Romania (right panel).

The right panel of Figure 3 illustrates this relation for regional data from Romania. Again a linear approximation looks reasonable. The size of the dots indicates the level of **OOP** for that observation. To identify the colors for the different Romanian regions, a color version of the pdf (or the website) is useful.

## 2.2 confounding effects

The model sets the stage for the empirical analysis. First, it helps us specify functional forms. To illustrate, the way the poverty rate in a region affects the percentage of people indicating they postponed treatment because it was too expensive is via the term  $b_{interaction} \cdot Poverty \cdot OOP \cdot \bar{x}$ .

Second, it helps us to avoid “causal salad” (McElreath 2020) where variables are added to control for “all other effects”. To illustrate this, consider adding variables like gdp or unemployment to the mortality equation. Isn’t unemployment likely to affect mortality? Indeed, it does and the model is clear about the causal mechanism: due to unemployment, income falls and poverty increases which has two effects on mortality. First, poverty is associated with less healthy lifestyle choices, e.g. because fresh fruit and vegetables tend to be expensive. Second, poverty makes it more likely that people forgo valuable treatments, especially when OOP is high.

Because the model is clear on the mechanisms that are covered, we can also identify potential mechanisms that are missing which can confound our estimates. Note that differences between regions that are more or less constant over time –think of lifestyle (smoking habits), pollution etc.– are captured by region fixed effects. Effects are potentially confounding if they vary over time and simultaneously affect both unmet needs and mortality (in the first equation) or both **TooExp** and the interaction  $oop \times poverty$  (in the second equation). In the robustness analysis, we focus on two such mechanisms: economic shocks and health shocks.

A negative economic shock can reduce government resources available to finance healthcare in the region or country. If this increases waiting lists due to reduced healthcare capacity, **Unmet** is likely to rise. At the same time, this can also reduce the quality of care, say because equipment is replaced less often or high quality physicians leave and are replaced by lower quality staff. This reduction in care quality can affect health and mortality introducing a different mechanism from the one we focus on: unmet medical needs negatively affect health and raise mortality.

Recall that an economic shock which raises poverty, the fraction of people that forgo treatment because it is too expensive and increases mortality is captured by the model and hence is not a confounding effect.

A health shock can increase both unmet medical needs (as more people fall ill, more people can have unmet needs) and increase mortality directly. In the same vein, a health shock can increase the fraction of people that forgo medical treatment because it is too expensive and can raise poverty through reduced productivity. Al-



though our data does not include the Covid years, we cannot exclude the possibility of other shocks that tend to reduce health and raise mortality.

In the robustness section we introduce variables to control for these potentially confounding economic and health effects. Then we compare our baseline estimate of the OOP effect with the effect that follows from these equations with extended variable set.

### 3 DATA

The data that we use is from Eurostat’s regional database and provides for NUTS 2 regions population size and number of deaths per age-gender category. In principle, we have data on 14 countries and 78 NUTS 2 regions for the years 2009-2019, ages 35-85 for women and men. The years 2009-2019 were chosen because, at the time of the analysis, data on poverty was available from 2009 onward and data on the number of deaths ran till 2019. Further, we want to exclude the corona years which were exceptional in terms of mortality. We start at age 35 because at ages below 35, mortality is so low that there is hardly a difference between mortality in regions with different poverty levels (see Figure 4 below). For ages above 85 population numbers per region get rather low.

We drop NUTS 2 region-year combinations where for an age-gender category –due to reporting issues or people moving– the number of deaths in a year exceeds the population size at the start of the year. For the baseline analysis, we focus on observations where we have complete records on mortality, the fraction of people indicating they postponed treatment because it was too expensive and oop expenditure.

Table 1 shows the summary statistics for our variables. We have more than 50k observations.<sup>9</sup> The average population size per region-age-gender category is about 7500 and the average number of deaths 100. Median population size per category equals 6500 and median number of deaths 56. In our data, the percentage of people dying in a NUTS 2/year/age/gender category (**mortality**) equals 2% on average with a maximum of 20% for some region and age combination.

We use two measures for poverty; each of these measures comes from the EU statistics on income and living conditions (EU-SILC) survey. The first is “at-risk-

---

<sup>9</sup>A rough estimate of the max. number of observations that we could have is: 78 (regions) \* 10 (years) \* 50 (ages) \* 2 (genders) = 78k. Missing observations on some of the key variables reduces this to 50k.

Table 1: Summary statistics main variables

	count	mean	std	min	median	max
population	52612	7491.3	4805.3	440	6477	36117
deaths	52612	103.2	126.5	0	56	1033
mortality (%)	52612	2.1	2.9	0	0.8	20.7
poverty (%)	50878	16.5	6.6	2.6	15.3	36.1
deprivation (%)	52612	11.2	12.8	0	3.4	52.3
too exp. (%)	52612	2	3.1	0	0.6	16
unmet (%)	52612	5.8	4.1	0	4.8	20.9
out-of-pocket (%)	52612	22	8.9	8.8	19.5	47.7
voluntary (%)	52612	3.1	3.1	0.3	1.6	15.2
expend. per head	52612	3386.6	2691.3	307.7	3559.5	8484.9
cdd	46186	72	97.1	0	18.8	615.3
hdd	46186	3467.1	1179.9	493.9	3089.8	7011.1
infant mortality (‰)	52612	4.3	2.3	0.8	3.6	11.6
bad health (%)	52612	12.8	12.2	0.8	8.3	78.9

of-poverty rate” that we refer to as **poverty**. This is a relative poverty measure: the share of people with disposable income after social transfers below a threshold based on the national median disposable income. The material deprivation measure (denoted **deprivation**) refers to the enforced inability to pay unexpected expenses, afford adequate heating of the home, durable goods like a washing machine etc. See the appendix for details.

In our data, the (unweighted) average (across regions and years) percentage of people at risk of poverty equals 16% with a maximum of 36%. For material deprivation the numbers are 11% and 52%. These measures vary by NUTS 2 region and year but not by age or gender. We use **deprivation** in our baseline analysis because it captures more closely the idea of postponing treatment due to financial constraints. The **poverty** variable is used in a robustness check.

Also from the EU-SILC survey, we use the variable capturing unmet medical needs because the forgone treatment was too expensive (**too exp**). The variable **unmet** measures percentage of people in need of healthcare that postpone or forgo treatment because it is either too expensive, the hospital is too far away, there is a waiting list for the treatment, the patient hopes that symptoms will disappear without treatment, the patient is afraid of treatment or has no time to visit a

physician. As explained in the model above, our analysis uses both `too_exp` and `unmet` (which includes `too_exp` as reason for unmet medical needs) as variables.

The measure `OOP` that we use in the baseline model, is based on household oop payments (`out-of-pocket`). In particular, this measures the percentage of healthcare expenditures paid oop. This varies by country and year. The higher `OOP`, the less generous the healthcare system is (in terms of higher coinsurance  $\xi$  or deductible  $D$  in the model above). We expect that high `OOP` is especially problematic in regions with a high percentage of people with low income.

In a robustness analysis we consider the sum of oop and payments to voluntary health insurance (`voluntary`) as a percentage of health expenditures as our `OOP` measure. The reason why we also consider voluntary insurance is that basic or mandatory insurance packages can differ between countries. If people are willing to spend money on voluntary insurance, it can be the case that this voluntary insurance covers treatments that people deem to be important. Put differently, a country that finances all expenditure (“free at point of service”) for a very narrow set of treatments would appear generous if we only used oop payments. The narrowness of this insurance would then be signalled by people buying voluntary insurance to cover other treatments.

As can be seen in Table 1, `out-of-pocket` is the most important component of the two `OOP` inputs. Percentage of healthcare expenditure paid oop is a multiple of the percentage financed via voluntary insurance (both in terms of the mean and of the minimum, median and maximum reported in the table). Therefore, the baseline model works with oop payments (only).

As shown in Lemma 1, healthcare expenditure per head (`expend_per_head`) affects how `OOP` influences the fraction of people forgoing treatment because it is too expensive. Expenditure per head is on average 3300 euro for the countries in our data. But the variation is big with a standard deviation of almost 2700 euro.

The last four variables are used in our analysis of confounding effects. The variables `cdd`, `hdd` measure cold and hot days per year in the NUTS 2 region. Days with extreme temperatures are known to affect health and mortality and vary year by year. From the Eurostat website: the cooling degree days (`cdd`) index measures “the severity of the heat in a specific time period taking into consideration outdoor temperature and average room temperature (in other words the need for cooling)”<sup>10</sup> and similarly the heating degree days (`hdd`) index measures the severity of the

---

<sup>10</sup>See [https://ec.europa.eu/eurostat/cache/metadata/en/nrg\\_chdd\\_esms.htm](https://ec.europa.eu/eurostat/cache/metadata/en/nrg_chdd_esms.htm)

cold. Days with extreme heat are correlated with mortality and we will use these variables as health shocks in the robustness analysis. The index is calculated as the sum over days with extreme temperatures and then sums the difference between the temperature and “room temperature”. The mean and standard deviation for `hdd` (cold days) is order of magnitude larger than `cdd`. Whereas there exist regions in Europe without (extremely) hot days, all regions experience cold days.

Infant mortality is a well known measure of healthcare quality. In contrast to quality measures like treatable and preventable mortality, infant mortality is not directly correlated with our mortality measure which considers people above age 35. Infant mortality is defined as the number of deaths of infants (younger than one year of age at death) per 1000 live births in a given year.

Finally, `bad health` gives the percentage of people who answer bad or very bad when asked about their health status in the EU-SILC survey. Around 12% report self-perceived bad or very bad health and this ranges from less than 1% in some regions to almost 80% in others. This variable is also used to control for changing health over time as a potential confounding variable.

Figure 4 (left panel) shows average mortality as a function of age for women and men. This is the pattern that one would expect: clearly increasing with age from age 40 onward and higher for men than for women (as women tend to live longer than men). Figure 4 (middle panel) shows the effect we are interested in: mortality is higher in regions where the interaction  $00P \times \text{Poverty}$  is high than where it is low and this difference increases with age.

Both for women and for men, we plot per age category the difference between average mortality in regions that are at least 0.5 standard deviation above the mean for  $00P \times \text{Poverty}$  and regions that are 0.5 standard deviation below the mean. Around age 82, this mortality difference equals approximately 4 percentage points. In the raw data, for 100 (wo)men aged 82, there are 4 additional deaths in regions with high interaction  $00P \times \text{Poverty}$  compared to regions with low interaction. Note that this plot of the raw data does not correct for other factors, like the poverty level itself, and thus over-estimates the size of the effect of  $00P \times \text{Poverty}$  on mortality. The right panel in this figure does a similar exercise with the fraction of people reporting unmet medical needs. Mortality is higher in regions where unmet needs are at least 0.5 standard deviation above the mean compared to regions where it is 0.5 standard deviation below the mean.

The observation from the figure that the difference between the two sets of regions is approximately zero for people below 35, is our motivation to include ages above 35

only in our data. Further, the difference in mortality between the regions increases with the mortality level in the left panel. This is in line with our specification in Lemma 1 where unmet needs has a multiplicative effect on the underlying mortality rate modeled by  $\eta_{ag} = e^{\beta_{ag}} / (1 + e^{\beta_{ag}})$ .

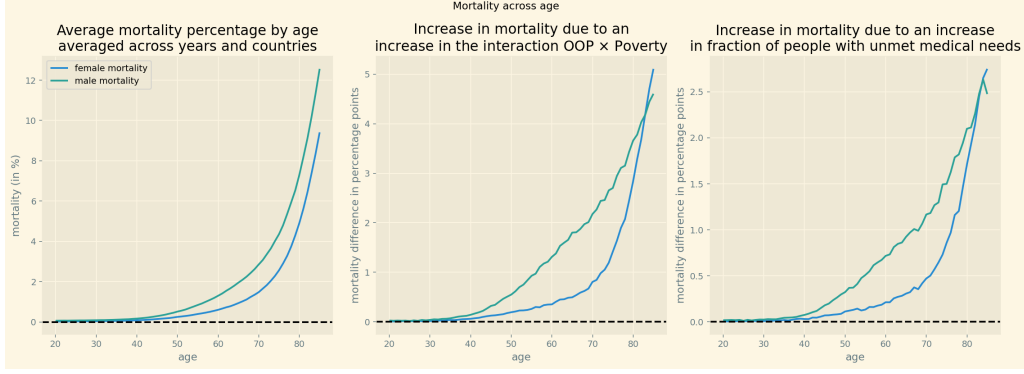


Figure 4: Mortality and difference in mortality between regions with high and low interaction  $OOP \times Poverty$  and high and low unmet medical needs.

## 4 ESTIMATION

In this section, we explain how we estimate the equations in Lemma 1.

### 4.1 Empirical model

The first equation estimates a binomial model with population size as the number of draws and deaths as the number of events. We do this for every combination of NUTS 2 region, calendar year, age and gender in our data. The probability of  $k \leq n$  deaths out of a population  $n$  is then given by

$$\binom{n}{k} m^k (1 - m)^{n-k}$$

where  $m$  denotes mortality: the probability of death. The equation that we estimate for mortality  $m_{ga2t}$  is given in the lemma above. The coefficient we are especially interested in is  $\beta_{unmet}$ . This is the coefficient through which an increase in unmet medical needs because of financial problems affects mortality.

Figure 4 illustrates that without the multiplicative specification for  $m_{ga2t}$  in the lemma, the coefficients for  $\beta_{unmet}, \beta_{poverty}$  would have to vary with age. Indeed, for the young mortality is low even in regions with high poverty or high unmet needs.

Specifying coefficients that vary with age would considerably increase the number of parameters that we need to estimate.

The second equation captures how an increase in `OOP` affects the fraction of people in a region that postpone or skip treatment because it is too expensive. This fraction `TooExp` is based on (EU-SILC) survey data where we do not know the number of people interviewed. Hence, we cannot model this as a binomial distribution. In our estimation we want to ensure that `TooExp` is between 0 and 1. For this we assume that `TooExp` in the lemma above has a logit-normal distribution. That is, the log-odds of `TooExp` is normally distributed.

Details of the estimation can be found in section C.1, the python code in the online appendix.

## 4.2 *Bayesian estimation*

We use Markov Chain Monte Carlo (MCMC), in particular the NUTS sampler to explore the posterior distributions of our parameters. For this sampler, we have the guarantee that the whole posterior distribution is captured as long as we have enough samples. Although this is an asymptotic result, we are confident that drawing four chains of 2000 samples (1000 samples of which are used for tuning) is enough to cover the posterior distribution. In the appendix we discuss a number of checks on this convergence.

It is not straightforward to put priors on the coefficients of the two equations in Lemma 1. To illustrate, how strong is the reaction of mortality to a 0.1 increase in the fraction of people reporting unmet medical needs? We are not aware of previous studies looking into this and have no a priori information on the strength of this effect. We use two principles when setting priors. First, we use regularizing priors (“seat belt priors”): priors close to zero with small standard deviations. Hence, a coefficient differs from zero only if there is clear evidence for this in the data. Second, we use a hierarchical model to determine the parameters of the prior distributions. Details on the priors can be found in the appendix.

As it is hard to judge how sensible a prior for one particular coefficient is, the online appendix shows the prior predictive distributions. That is, the predictions for mortality and `TooExp` that the model generates without having seen the data. Comparing the prior predictive distributions with the observed distributions we show that our priors do not exclude relevant possible outcomes nor do they put (much) weight on likely outcomes (say, mortality close to 0.9).

## 5 RESULTS

In this section we present the results of the estimation of the baseline model. Before presenting the outcome of our estimation, we present two graphical checks of our model.

### 5.1 *model fit*

Figure 5 gives an idea of the fit of the model in terms of predicting deaths per gender/age/region/year category and the fraction of people postponing treatment because it is too expensive.

The left panel shows observed number of deaths per category on the horizontal axis and the posterior predictive for this on the vertical axis. For each row in our data, we have observed number of deaths and a prediction of this number. In the figure, we show the average prediction of deaths across the posterior samples. The predictions are not perfect but do follow the 45-degree line closely.

The right panel shows the (log odds of the) fraction of people per region/year indicating they went without treatment (for a while) because it was too expensive. The difference between this panel compared to the one on the left is that this fraction does not vary by gender and age. Hence, we do not have a prediction for each row in our data. The right panel shows the observed and predicted fraction for **TooExp** per region/year. The dots indicate the average posterior prediction of this log-odds ratio. For small observed values of the **TooExp** (log-odds below  $-5$  in the figure) there is a range of predicted values. Although this range seems wide in log-odds space, both the observed and predicted values are basically equal to zero. To illustrate, for practical purposes it does not matter if a probability equals 0.0001 (log-odds of  $-9$ ) or 0.002 (log-odds of  $-6$ ). Both values are basically zero. Given our log-odds specification, the model cannot predict an exact zero probability.

A related observation is that in the data **TooExp** equals 0 for a number of region/year combinations. To handle this numerically, we use a lower bound for the log-odds. This corresponds to a probability of 0.0001 which is close enough to zero for our purposes. The right panel shows this bunching for a number of observations slightly below  $-9$ . The bunching for other values of observed log-odds between  $-5$  and  $-7$  corresponds to regions reporting rounded fractions of 0.001, 0.002 etc.

Compared to the observed number of deaths, the predictions for **TooExp** seem less accurate. This is to be expected as there are (a lot) fewer observations for this variable compared to mortality. But all in all the fit does not seem unreasonable as

the points cluster around the 45-degree line.

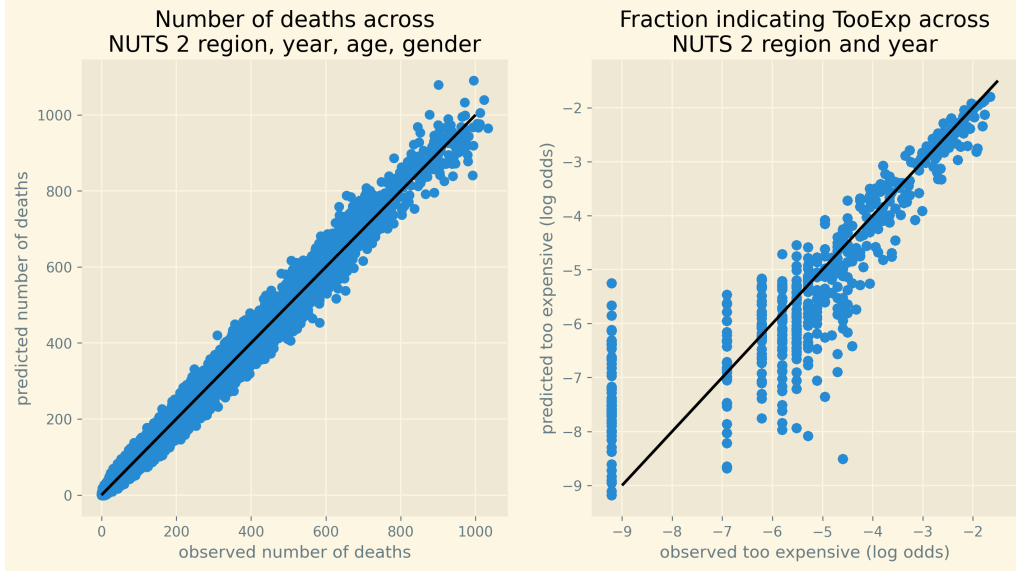


Figure 5: Fit of estimated and observed mortality across all observations and observed and predicted fraction of people indicating TooExp across NUTS 2 regions.

Another way to check how well the model fits, is to see how well it captures the age profile of mortality. This we present in Figure 6. The left panel shows the age profile  $\eta_{ag} = e^{\beta_{ag}} / (1 + e^{\beta_{ag}})$ . If the other terms in equation (4) equal 0,  $\eta_{ag}$  gives the probability of death for age/gender category  $ag$ . The right panel includes for every region and calendar year the correction on  $\eta_{ag}$  to yield mortality for that combination of gender/age/region/year. On average, the model captures the age profile perfectly.

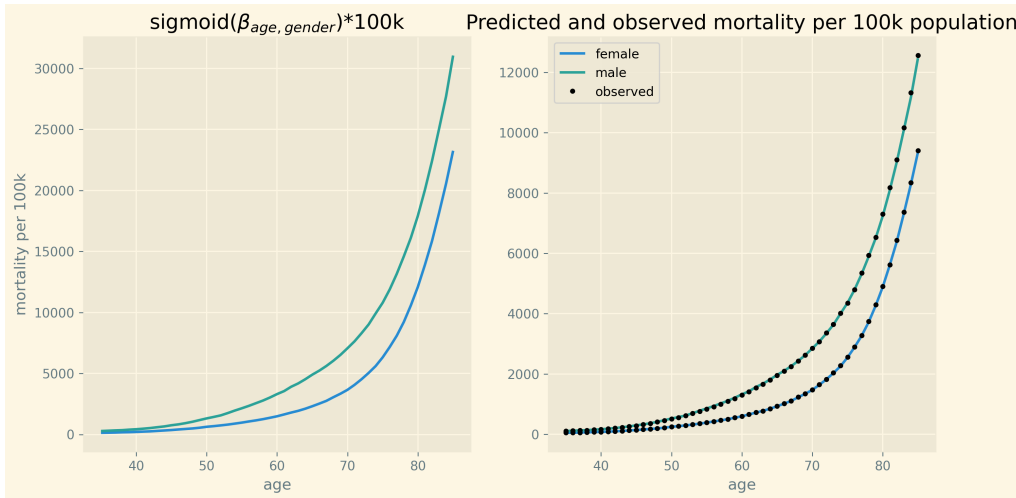


Figure 6: Fit of average mortality by age



The appendix presents two further checks of the model. Figure 10 shows the trace plots for the parameters of interest. The figures in the left panel show the posterior distribution of the parameters in the figure. The coefficients `b_oop`, `b_interaction` vary by country and hence we have different colors for the distributions in these graphs. The `beta` parameters do not vary with country (or another index) and hence there is one color only. In the `beta` figures it is easy to see that there are four distributions per parameter. These correspond to the four chains that are sampled by the NUTS algorithm.

The right panels show the same samples but now ordered across the horizontal axis as they were drawn. We check these plots for the following three features. First, the plot should be stationary; that is, not trending upward or downward. This implies that the posterior mean of the coefficient is (more or less) constant as we sample. Second, there should be good mixing which translates in condensed zig-zagging. In other words, the algorithm manages to draw values across the whole domain of the posterior quickly after each other. Finally, the four chains cover the same regions. All three features are satisfied for all coefficients in the right panel of the figure.

Another check on the convergence of the algorithm are the  $\hat{r}$  values in Table 5 in the appendix. This table summarizes the posterior distribution for the slopes that we are interested in. It provides the mean and standard deviation for each of these parameters, the 95% probability/credibility intervals and the number of effective samples for each parameter. As the number of these samples is roughly above 500 for all and above 1000 for most parameters, this looks fine. The final column presents the values for  $\hat{r}$  for each parameter. Since these are all equal (close) to one, we can be confident that the NUTS algorithm converged for these parameters.

## 5.2 size of effects

Table 5 in the appendix presents the posterior values for each of the parameters. Here we focus on the effect we are interested in: what is the increase in mortality due to an increase in oop? As we show in the appendix, a 500 euro increase in oop leads to the following increase in mortality:

$$\frac{dm_{ga2t}}{m_{ga2t}} = \beta_{unmet} \text{TooExp}_{2t} (1 - \text{TooExp}_{2t}) 500 (b_{oop,c} + b_{interaction,c} \text{Poverty}_{2t}) \quad (8)$$

Note that this increase in the number of deaths  $dm_{ga2t}$  per the number of deaths  $m_{ga2t}$  is independent of age. This is due to our formulation of mortality in equation

(4) where we have a baseline mortality  $\eta_{ag}$  depending on age/gender and a deviation from this baseline based on poverty and unmet medical needs etc. Figure 2 reports the expression in the equation above multiplied by 1000. That is, we report the increase in deaths due to the oop increase per 1000 deaths.

Note that the 500 euro change in OOP enters multiplicatively. In other words, dividing the effect in Figure 2 by ten gives the effect of a 50 euro increase in OOP for each country. In this sense, the choice of 500 euro is a matter of presentation.<sup>11</sup>

As the expression for  $dm/m$  varies with country, year and NUTS 2 region, Figure 2 summarizes our main findings in the following way. For each country we focus on the region where deprivation is highest. This is the region where we expect the mortality effect of an oop increase to be highest as many people could have problems paying medical bills. Table 2 presents this region for each country in our data together with the value of deprivation, the fraction of people with unmet medical needs due to financial constraints and the country's value for OOP. As the table illustrates, the fraction of people indicating that treatment was too expensive tends to be high when both deprivation and OOP are high.

Substituting these values from the table into the expression for  $dm/m$  we get the numbers in Figure 2. As mentioned, the blue bars give the average effect of the 500 euro increase in oop on mortality. As we have the posterior distributions for each of the parameters, we also have the posterior distribution for the mortality effects per country (taking the uncertainty for all parameters into account). The black horizontal lines present the 95% intervals around the mean effect.

The first observation is that for Bulgaria, Greece, Hungary and Romania the 95% probability interval is bounded away from zero. For these countries we can clearly see that an increase in oop negatively affects health and increases mortality.

Why are the effects smaller for the other countries? The effects are basically zero for the Scandinavian countries, Slovenia and Switzerland. As shown in Table 2, for these countries both deprivation and the fraction of people indicating unmet medical needs because treatment is too expensive are small. For the Scandinavian countries in the region with highest deprivation, TooExp is basically zero. It then follows from equation (8) that the effect on mortality is (close to) zero.

The equation also features the following positive second derivative effect. As OOP increases, TooExp increases (especially in regions with high deprivation). This

---

<sup>11</sup>Therefore, the observation in Table 1 that in some countries expenditure per head is below 500 euro is not a problem here.

Table 2: Fraction of people indicating material deprivation, foregoing treatment because it is too expensive and the country wide fraction of health expenditure paid out-of-pocket per NUTS 2 region with highest fraction of material deprivation per country.

region	country	deprivation	too expensive	OOP
BG33	Bulgaria	0.40	0.08	0.43
HR04	Croatia	0.13	0.01	0.11
DK02	Denmark	0.04	0.00	0.14
FI1C	Finland	0.03	0.00	0.18
EL63	Greece	0.28	0.07	0.37
HU31	Hungary	0.32	0.02	0.28
IE06	Ireland	0.07	0.02	0.12
LT02	Lithuania	0.12	0.01	0.32
NO01	Norway	0.02	0.00	0.14
RO22	Romania	0.32	0.11	0.21
SK04	Slovakia	0.11	0.01	0.20
SI03	Slovenia	0.05	0.00	0.12
SE22	Sweden	0.02	0.00	0.15
CH01	Switzerland	0.02	0.02	0.26

increases the mortality effect of a further increase in OOP.<sup>12</sup> Hence, the effect of an increase in OOP is bigger, the higher the starting point of OOP. Countries where OOP is already high, should be careful increasing it further because the detrimental health effects are stronger. In contrast, low initial values for OOP and TooExp imply that an increase in oop expenditure hardly affects health and mortality.

Another reason why the effects are small for some countries is that the underlying parameters `b_oop`, `b_interaction` are small for these countries. This can be seen in Table 5 in the appendix. If countries have policies to subsidize healthcare for poor families, the effect of country wide OOP on these families' unmet medical needs is small as they actually pay a lower fraction (than the national average) of their treatments' costs oop.

Summarizing, we can identify in our data the effect that an oop increase, raises the number of people with unmet medical needs due to financial constraints and

---

<sup>12</sup>Note that we use here that TooExp is smaller than 0.5:  $d(x(1-x))/dx = 1-2x > 0$  for  $x < 0.5$ . As tables 1 and 2 show, TooExp is indeed below 0.5 in our data.

hence increases mortality. This is especially the case in regions with high poverty and high initial OOP. Documenting this effect was the main objective of the paper.

A follow up question is: how big is this effect? In order to interpret the size of the oop effect, Table 3 presents the number of people dying from a particular cause per 1000 dead.<sup>13</sup> If we would consider all causes and add them up, the sum of the second column in Table 3 would equal 1000. The table focuses on causes of death with an order of magnitude comparable to the effects in Figure 2. The table is based on EU wide data in 2017 for ages 35-85.

Note that the comparison of the numbers in the figure with the numbers in the table is just to get an idea of the order of magnitude. But –strictly speaking– the causes are not comparable. Nobody dies of an increase in oop in the way people die from pneumonia. Due to an increase in oop, people may have gone without treatment which can then lead to death from, say, lung cancer. Hence, one should be careful in comparing the simulation results with the numbers in Table 3. But the table does provide some context in interpreting the size of the simulated effects.

Table 3: Number of people dying by cause (per 1000 dead) for ages 35-85 (EU average) using WHO’s icd10 disease classification.

icd10	per 1000
Malignant neoplasm of breast	23.66
Malignant neoplasm of prostate	16.16
Malignant neoplasm of bladder	9.72
Diabetes mellitus	22.59
Mental and behavioural disorders	26.85
Parkinson disease	9.17
Alzheimer disease	13.08
Pneumonia	19.74
Transport accidents	5.90

The average mortality effect due to a 500 euro increase in oop in Romania is approximately 33 (per 1000 dead). This exceeds deaths due to each of the causes in the table. The average effects in Bulgaria and Greece are around 15 and 22 resp. which places them between deaths due to Alzheimer disease and diabetes. In Hungary the order of magnitude is comparable to deaths due to transport accidents.

However, these are effects aggregated at the regional level (of the regions with

---

<sup>13</sup>We use the icd-10 classification here.

highest poverty levels). Suppose we are willing to assume that the incidence of the increase in mortality due to the 500 euro increase in oop falls mainly in the group of people who live in material deprivation. Table 2 shows this is around 30% for the relevant regions in Greece, Hungary and Romania. To get these effects at the region level, the effects among this specific group is an order of magnitude bigger.

Finally, there is also the following dynamic effect. As oop increases, 35 year olds postpone treatments thereby lowering their health status. Part of this reduced health leads to higher mortality among 35 year olds but some of these people survive this year. Next year, they start with lower than average health which can then raise mortality among 36 year olds. These dynamic feedback effects are captured by the parameter  $\gamma$  in Lemma 1. As shown in Table 5 in the appendix, the estimated value for  $\gamma$  is approximately 0.5 (coefficient `beta_lagged_log_mortality`). As effects accumulate across age and time, the effect for 85 year olds almost doubles ( $1 + \gamma + \dots + \gamma^{50} \approx 2$ ). To illustrate, the long run effect of a 500 euro increase in OOP leads to 66 deaths per 1000 dead for 85 year old Romanians in its poorest region.

One of the advantages of doing a Bayesian analysis is that we can easily show the uncertainty surrounding our estimated effects. This is illustrated in Figure 7 where we show for eight Romanian regions the probability that the mortality effect exceeds a certain value. Region RO22 tends to have the biggest effect (reported in Figure 2), while the effect per 1000 dead is smallest in NUTS 2 region RO42. We are pretty sure (probability close to 1) that in RO22 the effect is at least 15 per 1000 dead. While in RO42 this probability is less than 40%. In RO42 we are 80% sure that the effect exceeds 10 per 1000 dead. This is due to the fact that both `deprivation` and `too expensive` are substantially lower in RO42 compared to RO22.

Summarizing the discussion on the size of the effect, we find the following. In countries where poverty and OOP are high, a 500 euro (further) increase in oop leads to an increase in mortality (per 1000 dead) that is comparable to causes varying from Alzheimer disease to diabetes or breast cancer.

## 6 ROBUSTNESS

In this section we analyze confounding effects and discuss three different robustness checks on choice of variables. We explain each robustness check and discuss the results. The details of these checks can be found in the online appendix.

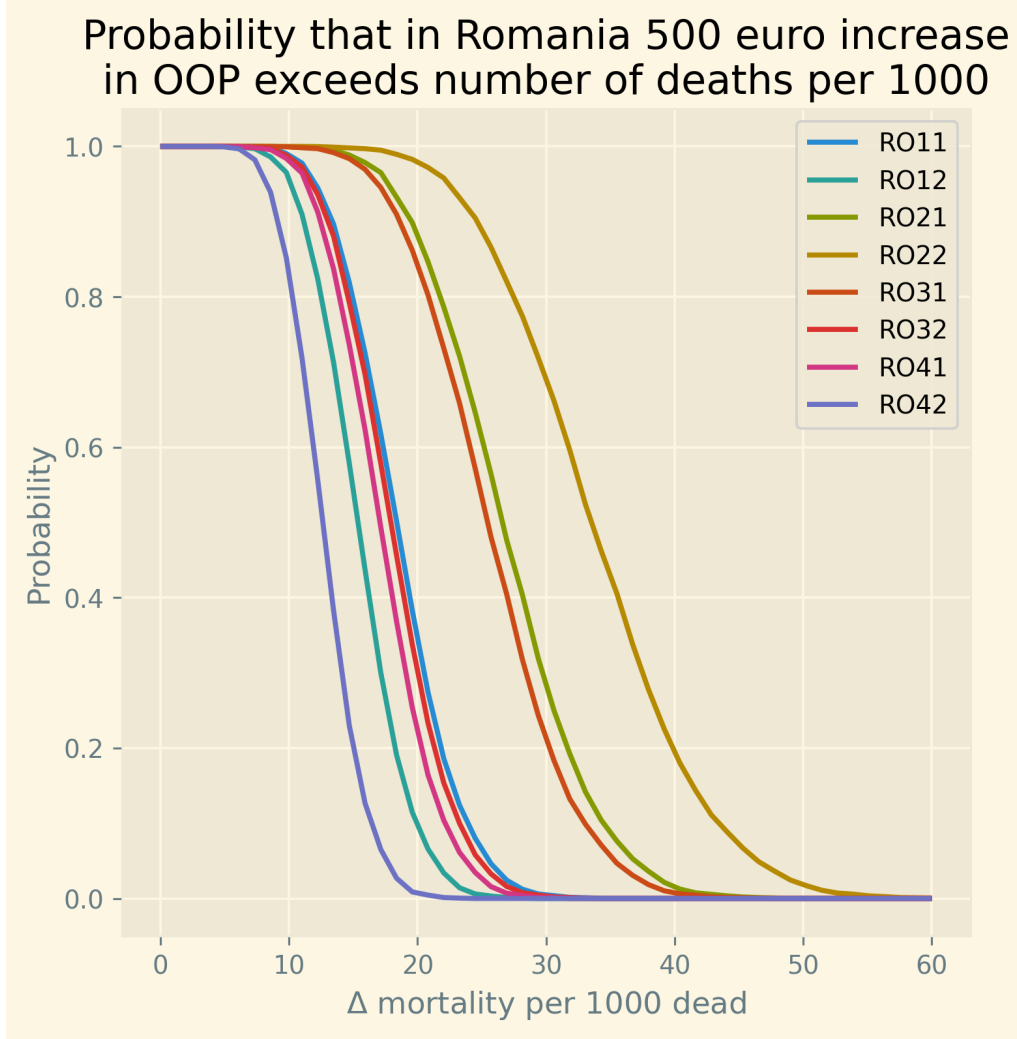


Figure 7: Uncertainty of the mortality effects in eight NUTS 2 regions in Romania.

### 6.1 confounding effects

As discussed in section 2.2 we focus on two mechanisms that may confound our estimated effect of oop on health and mortality: an economic shock and a health shock.

As described above, a plausible path for the economic shock goes via the quality of care (negative shock reduces resources available for healthcare thereby reducing quality). Controlling for healthcare quality closes this backdoor path (Cinelli, Forney, and Pearl 2022).

We use infant mortality to control for the quality of healthcare. This is a well known measure of care quality and is not directly related to our mortality measure which considers people older than 35.<sup>14</sup>

<sup>14</sup>There is a potential effect going in the other direction: infant mortality may be higher if

Note that we cannot use quality measures like number of physicians or MRI scanners per 100k population. If due to high oop and poverty many people forgo treatments, these measures will be low and mortality is high. But this is exactly the mechanism that we are analyzing here. In DAG-terminology, this would be controlling for a mediator which is a bad idea (Cinelli, Forney, and Pearl 2022). Similarly, we cannot use quality measure like preventable and treatable mortality because these are (part of) the mortality effects we are estimating.

In order to close the health shock path we use indices for hot and cold days. Periods with extreme temperatures (especially hot days) tend to raise mortality. If the estimated effect of unmet medical needs on mortality is caused by health shocks, including these indices should reduce the effect thereby signaling that our baseline estimates overestimate the true effect. We also use the fraction of people who state their (perceived) health is “bad or very bad” to control for health shocks.

Hence we re-estimate our two equations (for mortality and TooExp) by adding to each equation the variables `cdd`, `hdd`, infant mortality and self-perceived health. With the newly estimated parameters we calculate the percentage increase in mortality due to a 500 euro increase in oop using equation (8). Then we divide this percentage increase in mortality by the baseline estimated mortality increase. This ratio should be around 1 if the confounding effects are not important.

Estimated increase in mortality due to 500 euro increase in oop  
in the extended relative to the baseline model

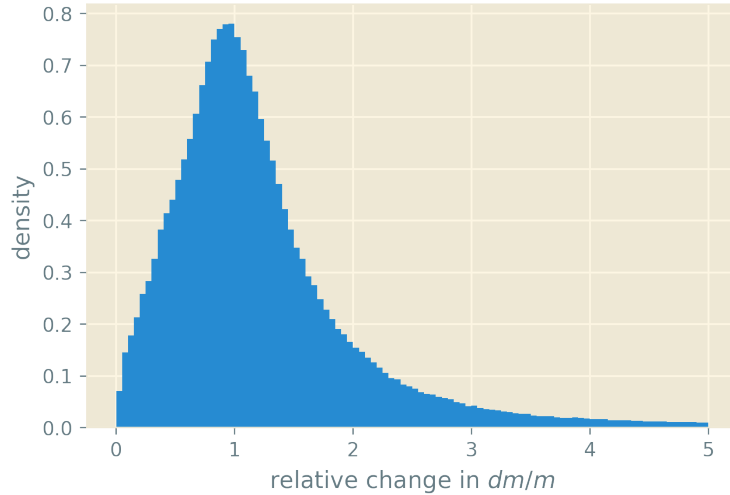


Figure 8: Mortality effect of OOP in extended equations relative to baseline estimate  
mothers’ health is compromised. This could suggest a bias while actually our estimates are correct. As explained below, we do not find evidence of a bias in the first place.

As shown in Figure 8, the distribution of the relative effects (across traces and countries) has a mode close to one and quite some variation around it. We know from Figure 2 that there is quite some uncertainty about this effect. This is reflected in the dispersion of the histogram (which is clipped at 5 to improve readability).

As the mass of the effect in the histogram is clustered around one, there is little reason to believe that the estimated effect in the baseline model is biased due to economic or health shocks. Once we control for such shocks we find effects that are on average in line with the baseline effects.

## 6.2 *robustness checks*

First, we use the at-risk-of-poverty variable, instead of deprivation as our variable to capture the fraction of people on low income. Second, we extend our definition of oop costs with expenditures on voluntary health insurance. Third, instead of focusing on the coefficient  $\beta_{unmet}$  in the mortality equation, we work with `TooExp` in this equation. The idea here is that other reasons for unmet needs (than too expensive) may have bigger effects on mortality. If this would be the case, the baseline model overestimates the effect of too expensive (and hence of oop) on mortality.

In the robustness check with at-risk-of-poverty we estimate our model taking missing observations into account. As shown in Table 1, poverty has fewer observations than deprivation and we prefer not to shrink our sample size here.

In Bayesian analysis there is a natural way to deal with missing variables which we view as an improvement on interpolating the missing values. Interpolating data, say by replacing a missing value with the mean value of the variable makes the estimation method “too confident” about this value, thereby negatively affecting the quality of the inference.

We use the following method to deal with missing values (McElreath 2020). The uncertainty surrounding the value of a missing observation is taken into account in the posterior distributions of our parameters. When sampling the posterior, if we encounter a missing value in a variable, this value is drawn from its distribution. We work with 4000 samples for the posterior and hence we draw 4000 different values for each missing observation. In this way, the uncertainty about the (missing) value translates into posterior uncertainty of the parameters and predictions.

We summarize the robustness checks in Figure 9. The figure has the baseline average effects of Figure 2 on the horizontal axis. For a number of countries these effects are smaller than 5 (per 1000 dead). In all of the robustness checks, these



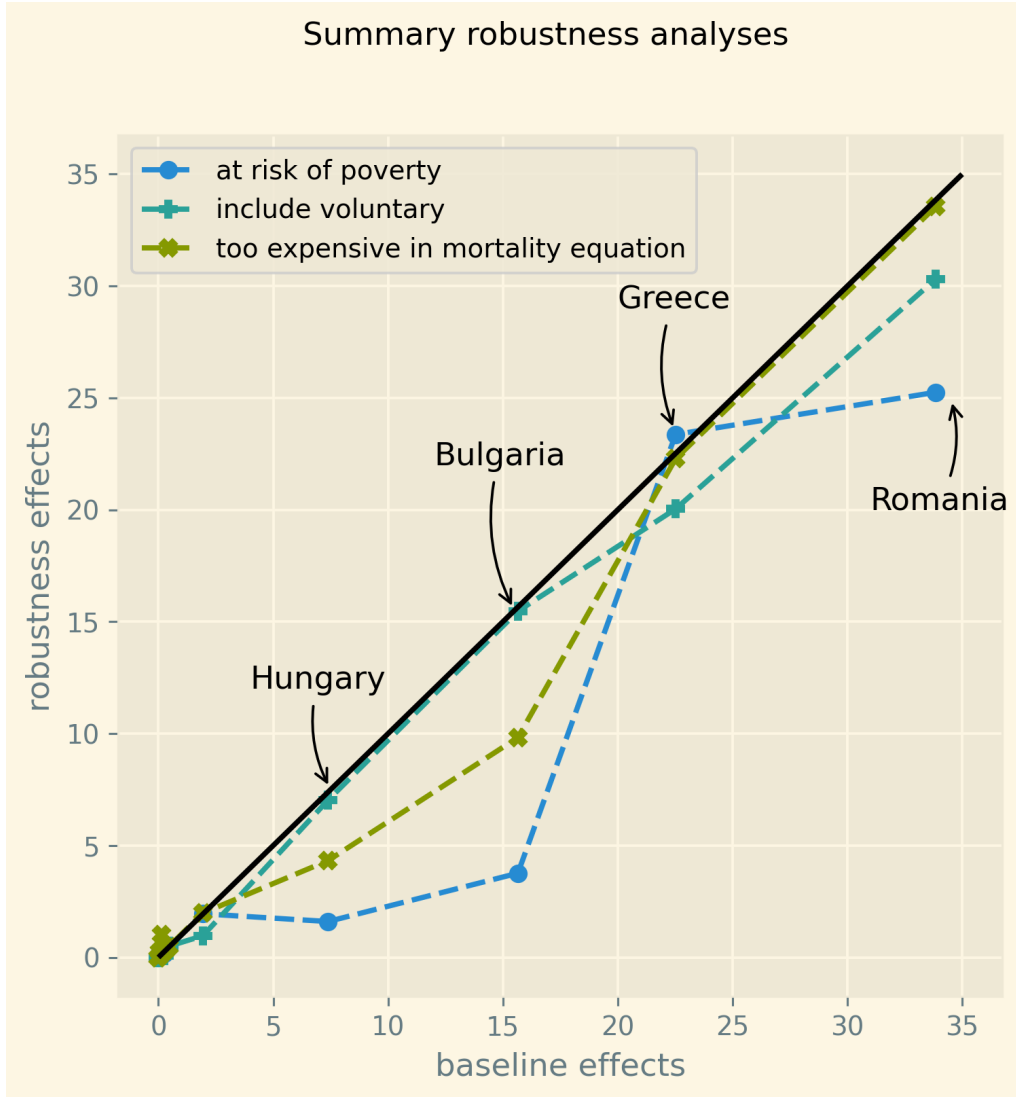


Figure 9: Summary of four robustness checks

effects remain small.

The four countries with baseline effects exceeding five are explicitly named in the figure. For a given baseline effect, there are three (vertically aligned per country) robustness effects. Ideally, all three points would lie on the 45-degree line. But, obviously, there is some variation in the effect size for different specifications.

Including voluntary health insurance payments in our oop measure has the smallest effect on outcomes. The points are very close to the 45-degree line: the effects are basically the same as with the baseline specification.

Using `too expensive` instead of `unmet` in the mortality equation yields smaller effects for Hungary and Bulgaria, but the differences with the baseline are relatively small. The ranking of the effect size across the four countries (with significant

effects) remains the same.

The differences are bigger if we use the at-risk-of-poverty rate instead of deprivation. Especially in Bulgaria and Romania, the effects are substantially smaller. As at-risk-of-poverty is a relative poverty measure, one would expect the effect of this variable on `too_expensive` and hence on unmet needs to be smaller than with deprivation. This also reduces the overall effect of oop on mortality.

Overall the picture is that the baseline results are fairly robust to the use of different variables and different specifications.

## 7 DISCUSSION AND POLICY IMPLICATIONS

The Introduction discusses a recent literature using individual level data analyzing whether demand side cost sharing reduces expenditure on low value treatments (usually referred to as moral hazard) or whether it leads patients to postpone or forgo valuable treatments thereby negatively affecting their health. This literature is focused on US individual level data and argues that it is hard to identify negative health effects of cost sharing in aggregate data. We use European data at the (NUTS 2) regional level to show that a high share of out-of-pocket expenditures (in total healthcare expenditures) has a clear effect on mortality in regions where the fraction of low income households is high. The size of the mortality effect increases with the initial level of out-of-pocket expenditure and hence the initial fraction of people indicating they skip or postpone treatment because it is too expensive.

Healthcare costs keep increasing in most, if not all, developed countries. Demand side cost sharing is a well known instrument to curb the growth in expenditure. This paper shows that there is a threshold of out-of-pocket expenditure beyond which regions with high poverty levels start to show increased mortality rates. To avoid this mortality effect, policy makers in countries with high out-of-pocket expenditure need to search for alternative instruments. Possible alternatives are means tested cost sharing or co-payments that are lower for cost-effective treatments. In the latter case, high value treatments are cheaper than low value care and hence less likely to be postponed by people on low income.

## 8 BIBLIOGRAPHY

Abaluck, Jason, Mauricio M. Caceres Bravo, Peter Hull, and Amanda Starc. 2020. “Mortality Effects and Choice across Private Health Insurance Plans,” July. Na-

tional Bureau of Economic Research. doi:10.3386/w27578.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. 2015. "Behavioral Hazard in Health Insurance." *The Quarterly Journal of Economics* 130 (4). Oxford University Press (OUP): 1623–67. doi:10.1093/qje/qjv029.

Black, Bernard, Alex Hollingsworth, Leticia Nunes, and Kosali Simon. 2021. "Simulated Power Analyses for Observational Studies: An Application to the Affordable Care Act Medicaid Expansion," February. National Bureau of Economic Research. doi:10.3386/w25568.

Boone, Jan. 2022. "Replication Data for: European Data on Mortality, Unmet Medical Needs and Healthcare Expenditure." DataverseNL. doi:10.34894/AABEBD.

Borgschulte, Mark, and Jacob Vogler. 2020. "Did the Aca Medicaid Expansion Save Lives?" *Journal of Health Economics* 72 (July). Elsevier BV: 102333. doi:10.1016/j.jhealeco.2020.102333.

Brot-Goldberg, Zarek C., Amitabh Chandra, Benjamin R. Handel, and Jonathan T. Kolstad. 2017. "What Does a Deductible Do? the Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics." *The Quarterly Journal of Economics* 132 (3). Oxford University Press (OUP): 1261–1318. doi:10.1093/qje/qjx013.

Chandra, Amitabh, Evan Flack, and Ziad Obermeyer. 2021. "The Health Costs of Cost-Sharing," February. National Bureau of Economic Research. doi:10.3386/w28439.

Chernew, Michael E., Mayur R. Shah, Arnold Wegh, Stephen N. Rosenberg, Iver A. Juster, Allison B. Rosen, Michael C. Sokol, Kristina Yu-Isenberg, and A. Mark Fendrick. 2008. "Impact of Decreasing Copayments on Medication Adherence within a Disease Management Environment." *Health Affairs* 27 (1). Health Affairs (Project Hope): 103–12. doi:10.1377/hlthaff.27.1.103.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016. "The Association between Income and Life Expectancy in the United States, 2001-2014." *Jama* 315 (16). American Medical Association (AMA): 1750. doi:10.1001/jama.2016.4226.

Cinelli, Carlos, Andrew Forney, and Judea Pearl. 2022. "A Crash Course in Good and Bad Controls." *Sociological Methods & Research* nil (nil): 004912412210995. doi:10.1177/00491241221099552.

Cutler, David M., Adriana Lleras-Muney, and Tom Vogl. 2011. *Chapter 7 - Socioeconomic Status and Health: Dimensions and Mechanisms, in S. Glied and P. Smith, Editors, Oxford Handbook of Health Economics*. Oxford University Press.

Ellis, R.P. 1986. "Rational Behavior in the Presence of Coverage Ceilings and Deductibles." *Rand Journal of Economics* 17 (2): 158–75.

- Goldin, Jacob, Ithai Z Lurie, and Janet McCubbin. 2020. "Health Insurance and Mortality: Experimental Evidence from Taxpayer Outreach." *The Quarterly Journal of Economics* 136 (1). Oxford University Press (OUP): 1–49. doi:10.1093/qje/qjaa029.
- Gross, Tal, Timothy Layton, and Daniel Prinz. 2020. "The Liquidity Sensitivity of Healthcare Consumption: Evidence from Social Security Payments," October. National Bureau of Economic Research. doi:10.3386/w27977.
- Keeler, E. B., J. P. Newhouse, and C. E. Phelps. 1977. "Deductibles and the Demand for Medical Care Services: The Theory of a Consumer Facing a Variable Price Schedule under Uncertainty." *Econometrica* 45 (3): 641–55.
- Mackenbach, Johan P., Irina Stirbu, Albert-Jan R. Roskam, Maartje M. Schaap, Gwenn Menvielle, Mall Leinsalu, and Anton E. Kunst. 2008. "Socioeconomic Inequalities in Health in 22 European Countries." *New England Journal of Medicine* 358 (23). Massachusetts Medical Society: 2468–81. doi:10.1056/nejmsa0707519.
- McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Second. Chapman and Hall/CRC.
- Miller, Sarah, Norman Johnson, and Laura R Wherry. 2021. "Medicaid and Mortality: New Evidence from Linked Survey and Administrative Data." *The Quarterly Journal of Economics*, January. Oxford University Press (OUP). doi:10.1093/qje/qjab004.
- Newhouse, J.P., and the Insurance Experiment Group. 1993. *Free for All? Lessons from the RAND Health Insurance Experiment*. Cambridge, Massachusetts: Harvard University Press.
- Nyman, J.A. 2003. *The Theory of Demand for Health Insurance*. Stanford University Press.
- OECD. 2021. *Health at a Glance 2021*. doi:<https://doi.org/https://doi.org/10.1787/ae3016b9-en>.
- Rothschild, M., and J. Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *The Quarterly Journal of Economics* 90 (4): 629–49.
- Schokkaert, Erik, and Carine van de Voorde. 2011. "Chapter 15 - User Charges." In *Oxford Handbook of Health Economics*, edited by S. Glied and P. Smith, 329–53. Oxford University Press.
- Semyonov, Moshe, Noah Lewin-Epstein, and Dina Maskileyson. 2013. "Where Wealth Matters More for Health: The Wealth-Health Gradient in 16 Countries." *Social Science & Medicine* 81 (March). Elsevier BV: 10–17. doi:10.1016/j.socscimed.2013.01.010.

## A PROOF OF RESULTS

**Proof of Lemma 1** First, we show that the probability of treatment,  $\delta_i^j = 1 - G(\sigma_0/\sigma_i u(y^j)/u(y^j - oop_i))$ , is increasing in  $y^j$  and decreasing in  $oop_i$ . Taking the derivative

$$\frac{d\left(\frac{u(y^j)}{u(y^j - oop_i)}\right)}{dy^j} = \frac{u'(y^j)u(y^j - oop_i) - u(y^j)u'(y^j - oop_i)}{u(y^j - oop_i)^2} < 0$$

because  $u$  is positive and increasing in  $y$ ,  $u' > 0$  is decreasing in  $y$  and  $oop_i > 0$ . Hence, the probability of treatment is increasing in income  $y$ . Similarly, the treatment probability falls with  $oop$ .

The expression for mortality follows from equation (3) which we can write as:

$$\begin{aligned} \ln(m_{agt}) = & \ln(\eta_{ag}) + \gamma \ln\left(\frac{h_{a-1,g,t-1}}{\bar{g}_{a-1,g}}\right) - (1 - \pi^h) - \pi^h \sum_{i \in I} \zeta_i \sigma_i + \alpha(\pi^l - \pi^h)(1 - \sum_{i \in I} \zeta_i \sigma_i) \\ & + \sum_{i \in I} \zeta_i (\sigma_i - \sigma_0)(\alpha \pi^l (1 - \delta_i^l) + (1 - \alpha) \pi^h (1 - \delta_i^h)) \end{aligned}$$

We capture  $\eta_{ag}$  with a sigmoid of age and gender fixed effects,  $\beta_{ag}$ . The NUTS 2 fixed effects capture  $(1 - \pi^h) + \pi^h \sum_{i \in I} \zeta_i \sigma_i$ . As  $\alpha$  denotes poverty, we have

$$\beta_{poverty} = (\pi^l - \pi^h)(1 - \sum_{i \in I} \zeta_i \sigma_i) \sigma_0 > 0$$

With the expression for **Unmet** in equation (5), we find that

$$\beta_{unmet} = \frac{\sum_{i \in I} \zeta_i (\sigma_i - \sigma_0)(\alpha \pi^l (1 - \delta_i^l) + (1 - \alpha) \pi^h (1 - \delta_i^h))}{\sum_{i \in I} \zeta_i (\alpha \pi^l (1 - \delta_i^l) + (1 - \alpha) \pi^h (1 - \delta_i^h))}$$

which is a weighted average of the  $\sigma_i - \sigma_0 > 0$  terms: if the medical needs were met, health would equal  $\sigma_i$  but for this group it is  $\sigma_0 < \sigma_i$ . Hence,  $\beta_{unmet} = E(\sigma_i - \sigma_0) > 0$ .

Finally, we derive how the fraction of people that forgo treatment because it is too expensive depends on OOP. We first derive this for an increase in  $D$ . We start from

$$\frac{d\text{TooExp}}{d\text{OOP}} = \frac{d\text{TooExp}}{dD} \left( \frac{d\text{OOP}}{dD} \right)^{-1} = \frac{d\text{TooExp}}{dD} \frac{\sum_{i \in I} \zeta_i x_i \delta_i}{\sum_{i \in I_D} \zeta_i \delta_i} \quad (9)$$

where we use that in Europe oop payments tend to be small relative to yearly income and hence we use the approximation that  $\delta_i^j$  is constant across  $j$ ,  $\delta_i^j \approx \delta_i$ :

$$\text{OOP} = \frac{\sum_{i \in I_\xi} \zeta_i \xi x_i \delta_i + \sum_{i \in I_D} \zeta_i D \delta_i}{\sum_{i \in I} \zeta_i x_i \delta_i}$$

Further, equation (7) implies we can approximate the slope of TooExp with respect to  $D$  around 0 as:

$$\frac{d\text{TooExp}}{dD} = \sum_{i \in I_D} \zeta_i \left( \alpha \pi^l g_i^l \frac{\sigma_0}{\sigma_i} \frac{u'(y^l)}{u(y^l)} + (1 - \alpha) \pi^h g_i^h \frac{\sigma_0}{\sigma_i} \frac{u'(y^h)}{u(y^h)} \right)$$

where we use notation  $g_i^j = g(\sigma_0/\sigma_i * u(y^j)/u(y^j - D))$  for  $i \in I_D$ . Again using the approximation that  $y^j$  is big compared to  $D$ , we have that  $g_i^j = g_i$ :

$$\frac{d\text{TooExp}}{dD} = \sum_{i \in I_D} \zeta_i g_i \frac{\sigma_0}{\sigma_i} \left[ \pi^h \frac{u'(y^h)}{u(y^h)} + \alpha \left( \pi^l \frac{u'(y^l)}{u(y^l)} - \pi^h \frac{u'(y^h)}{u(y^h)} \right) \right]$$

where  $\left( \pi^l \frac{u'(y^l)}{u(y^l)} - \pi^h \frac{u'(y^h)}{u(y^h)} \right) > 0$  because  $\pi^l > \pi^h$ ,  $u'(y^l) > u'(y^h)$  and  $u(y^l) < u(y^h)$ . A similar derivation shows

$$\frac{d\text{TooExp}}{d\xi} = \sum_{i \in I_\xi} \zeta_i x_i g_i \frac{\sigma_0}{\sigma_i} \left[ \pi^h \frac{u'(y^h)}{u(y^h)} + \alpha \left( \pi^l \frac{u'(y^l)}{u(y^l)} - \pi^h \frac{u'(y^h)}{u(y^h)} \right) \right]$$

Using equation (9), we find that

$$\frac{d\text{TooExp}}{d\text{OOP}} = \left[ \pi^h \frac{u'(y^h)}{u(y^h)} + \alpha \left( \pi^l \frac{u'(y^l)}{u(y^l)} - \pi^h \frac{u'(y^h)}{u(y^h)} \right) \right] \sum_{i \in I_D} \zeta_i g_i \frac{\sigma_0}{\sigma_i} \frac{\sum_{i \in I} \zeta_i x_i \delta_i}{\sum_{i \in I_D} \zeta_i \delta_i} \quad (10)$$

And, similarly, for  $\xi$ :

$$\frac{d\text{TooExp}}{d\text{OOP}} = \left[ \pi^h \frac{u'(y^h)}{u(y^h)} + \alpha \left( \pi^l \frac{u'(y^l)}{u(y^l)} - \pi^h \frac{u'(y^h)}{u(y^h)} \right) \right] \sum_{i \in I_\xi} \zeta_i x_i g_i \frac{\sigma_0}{\sigma_i} \frac{\sum_{i \in I} \zeta_i x_i \delta_i}{\sum_{i \in I_\xi} \zeta_i \delta_i x_i}$$

Using that  $\delta_i = 1 - G_i$ , we can write the last term in (10) as:

$$\sum_{i \in I_D} \zeta_i g_i \frac{\sigma_0}{\sigma_i} \frac{\sum_{i \in I} \zeta_i x_i \delta_i}{\sum_{i \in I_D} \zeta_i \delta_i} = \left( \sum_{i \in I} \zeta_i x_i \delta_i \right) \sum_{i \in I_D} \frac{\zeta_i \delta_i}{\sum_{i \in I_D} \zeta_i \delta_i} \frac{g_i \frac{\sigma_0}{\sigma_i}}{1 - G_i} = \left( \sum_{i \in I} \zeta_i x_i \delta_i \right) \kappa$$

where  $\kappa > 0$  denotes the parameter of the Pareto distribution:

$$1 - G(\nu) = (\underline{\nu}/\nu)^\kappa$$

if  $\nu \geq \underline{\nu}$  and 1 otherwise. With this distribution we have  $g(\nu)\nu/(1 - G(\nu)) = \kappa$  for each  $\nu > \underline{\nu}$ . It is routine to verify that we get the same expression for the expansion with respect to  $\xi$ . If we assume that  $G(\nu)$  denotes the cumulative distribution function of a Pareto distribution, the linear expansion that we use below is exact. Otherwise, it is an approximation as  $\kappa$  is a function of  $\delta_i$  and hence of  $\xi$  and  $D$ .

Hence, irrespective of whether we expand with respect to  $D$  or  $\xi$ , we find that

$$\frac{d\text{TooExp}}{d\text{OOP}} = \kappa \left[ \pi^h \frac{u'(y^h)}{u(y^h)} + \alpha \left( \pi^l \frac{u'(y^l)}{u(y^l)} - \pi^h \frac{u'(y^h)}{u(y^h)} \right) \right] \sum_{i \in I} \zeta_i x_i \delta_i$$

where we capture  $\sum_{i \in I} \zeta_i x_i \delta_i$  with expenditure per head in our data.

Using this, we estimate the following linear expansion  $\text{TooExp} = b_0 + \frac{d\text{TooExp}}{d\text{OOP}} \text{OOP}$ :

$$\text{TooExp}_{2t} = b_{0,2} + b_{0,t} + \text{OOP}_{ct} \bar{x}_{ct} (b_{oop,c} + b_{interaction,c} \text{Poverty}_{2t}) \quad (11)$$

where

$$b_{oop,c} = \kappa \pi^h u'(y^h) / u(y^h) > 0$$

and

$$b_{interaction,c} = \kappa \left( \pi^l \frac{u'(y^l)}{u(y^l)} - \pi^h \frac{u'(y^h)}{u(y^h)} \right) > 0$$

As it is hard to know what determines the intercept for this linear expansion, we allow it to vary with NUTS 2 region and calendar year:  $b_0 = b_{0,2} + b_{0,t}$ . Finally, to facilitate the estimation of this equation we assume that **TooExp** has a logit-normal distribution. That is, the log-odds of **TooExp** are normally distributed with the mean given by equation (11). This ensures that **TooExp** in the estimation always lies between 0 and 1.

*Q.E.D.*

## B DATA

All our variables come from Eurostat. Table 4 shows the dimensions over which our variables vary: country, NUTS 2, calendar time, age and sex. We also present a clickable link to the variable on the Eurostat website for ease of reference. The file `./getting_data.org` presents the code to download the Eurostat data.<sup>15</sup>

The variables on poverty, deprivation and access to care (unmet and too expensive) come from the EU statistics on income and living conditions (EU-SILC) survey.

From the Eurostat Glossary: “The at-risk-of-poverty rate is the share of people with an equivalised disposable income (after social transfers) below the at-risk-of-poverty threshold, which is set at 60 % of the national median equivalised disposable income after social transfers. This indicator does not measure wealth or poverty, but low income in comparison to other residents in that country, which does not necessarily imply a low standard of living. The equivalised disposable income is the total income of a household, after tax and other deductions, that is available for spending or saving, divided by the number of household members converted into

---

<sup>15</sup>This file can be found in the github repository: [https://github.com/janboone/out\\_of\\_pocket\\_payments\\_and\\_health](https://github.com/janboone/out_of_pocket_payments_and_health).

Table 4: Variables and the dimensions over which they vary.

variable	country	NUTS 2	time	age	sex	reference
population		x	x	x	x	link
deaths		x	x	x	x	link
at-risk-of-poverty		x	x			link
material deprivation		x	x			link
fraction too expensive		x	x			link
unmet		x	x			link
out-of-pocket	x		x			link
voluntary	x		x			link
expenditure per head	x		x			link
cdd, hdd		x	x			link
infant mortality		x	x			link
bad health	x		x	x	x	link

equalised adults; household members are equalised or made equivalent by weighting each according to their age, using the so-called modified OECD equivalence scale.”

“Material deprivation refers to a state of economic strain and durables, defined as the enforced inability (rather than the choice not to do so) to pay unexpected expenses, afford a one-week annual holiday away from home, a meal involving meat, chicken or fish every second day, the adequate heating of a dwelling, durable goods like a washing machine, colour television, telephone or car, being confronted with payment arrears (mortgage or rent, utility bills, hire purchase instalments or other loan payments).” Our variable “material deprivation” equals the share of people in a NUTS 2 region in material deprivation.

Fraction of people with self-reported unmet needs for medical examination is based on the same survey. In particular, the definition of this item is “Self-reported unmet needs for health care: Proportion of people in need of health care reporting to have experienced delay in getting health care in the previous 12 months for reasons of financial barriers, long waiting lists, distance or transportation problems.”. We use both the general definition of unmet needs and the specific reason that treatment was too expensive.

We characterize how generous a health insurance system is using the variable `00P` in our analysis. This variable is derived from data on health care expenditure by financing scheme. For our `00P` measure we focus on voluntary healthcare payment



schemes (**voluntary**) and household out-of-pocket payment (**out-of-pocket**). Both measured as share of total current health expenditure. The baseline specification uses **out-of-pocket** only.

Expenditure per head refers to healthcare expenditure per head at the country level.

The exact definitions for **cdd**, **hdd** can be found here: <https://ec.europa.eu/eurostat/cache/metadata>. To illustrate, for the Heating Degree Days the calculation starts from a base temperature of 18 deg  $C$ . Then for days  $i$  with mean air temperature  $T_m^i \leq 15$ ,  $HDD = \sum_i (18 - T_m^i)$ . And similarly for Cooling Degree Days, which starts from a base temperature of 21 deg  $C$ .

Bad health equals the fraction of people indicating that their (self-perceived) health is either bad or very bad.

## C ESTIMATION

This section presents the full specification (including priors) and trace plots for the baseline model, the table with the relevant coefficients and the derivation of the change in mortality as a function of the change in oop.

### C.1 Bayesian model

The mortality equation is specified as follows:

$$k_{ga2t} \sim \text{Binomial}(n_{ga2t}, m_{ga2t})$$

where  $k_{ga2t}, n_{ga2t}$  denote the number of deaths, population size resp. in a gender/age/region/year category. Mortality is given by

$$m_{ga2t} = \frac{e^{\beta_{ag}}}{1 + e^{\beta_{ag}}} e^{\left( \mu_2 + \gamma \ln \left( \frac{m_{a-1,g,2,t-1}}{\bar{m}_{a-1,g}} \right) + \beta_{poverty} \text{Poverty}_{2t} + \beta_{unmet} \text{Unmet}_{2t} \right)}$$

where

$$\beta_{ag} \sim N(-3.0, 0.3)$$

with  $e^{-3}/(1 + e^{-3}) = 0.05$  implying that people above 35 have –on average– still 20 years to live. This is an under-estimation of life expectancy that will be easily adjusted by the data that we have. Both here and for the **TooExp** equation, we chose relatively narrow priors for the fixed effects. The prior for the regional fixed effects is given by:

$$\mu \sim \text{Normal}(0, 0.3)$$

We allow  $\gamma$  to be positive or negative with a relatively small standard deviation (determined in a hierarchical way):

$$\gamma \sim \text{Normal}(0, \text{sd\_prior\_beta})$$

Note that  $\gamma$  in the main text is denoted `beta_lagged_log_mortality` in the code. The standard deviation of this prior is determined by the following prior distribution (this is the hierarchy):

$$\text{sd\_prior\_beta} \sim \text{HalfNormal}(0.1)$$

Finally, we know from the theory that the following two parameters are non-negative:

$$\beta_{unmet}, \beta_{poverty} \sim \text{HalfNormal}(\text{sd\_prior\_beta})$$

In words, `sd_prior_beta` captures the extent to which (social) external factors (like poverty, previous period health etc.) can affect mortality at all, instead of it being mainly determined by biology (age).

Next, we turn to the equation for `TooExp`. We estimate `TooExp` with a logit-normal distribution.

$$\text{TooExp}_{2t} \sim \text{LogitNormal}(b_{0,2} + b_{0,t} + \text{OOP}_{ct} \bar{x}_{ct} (b_{oop,c} + b_{interaction,c} \text{Poverty}_{2t}), \sigma)$$

where the parameter  $\mu$  of the Normal distribution of the log-odds of `TooExp` is given by the equation in Lemma 1 and the parameter  $\sigma$  has a prior

$$\sigma \sim \text{HalfNormal}(1.0)$$

which allows for a range of values for  $\sigma$  that are positive. For the regional fixed effects, we have

$$b_{0,2} \sim \text{Normal}(-5.0, 0.3)$$

where the  $-5.0$  is approximately equal to the mean log-odds of `TooExp` (which equals  $-5.17$ ) and we chose a relatively narrow standard deviation for the fixed effects.

$$b_{0,t} \sim \text{Normal}(0, 0.3)$$

Since we know that the following two parameters are non-negative, we specify a half-normal distribution

$$b_{oop,c}, b_{interaction,c} \sim \text{HalfNormal}(\text{sd\_prior\_b})$$

where the prior of `sd_prior_b` is given by (hierarchical model):

$$\text{sd\_prior\_b} \sim \text{HalfNormal}(0.1)$$

### C.2 trace plots

Figure 10 gives the trace plots for the parameters that we are interested in. That is, we leave out the traces for the (age, calendar year and region) fixed effects.

As explained in the main text, we are interested in three features in the plots on the right. First, the plots are stationary; second, condensed zig-zagging and third, the four chains cover the same regions of the parameter space. These features are satisfied for our coefficients.

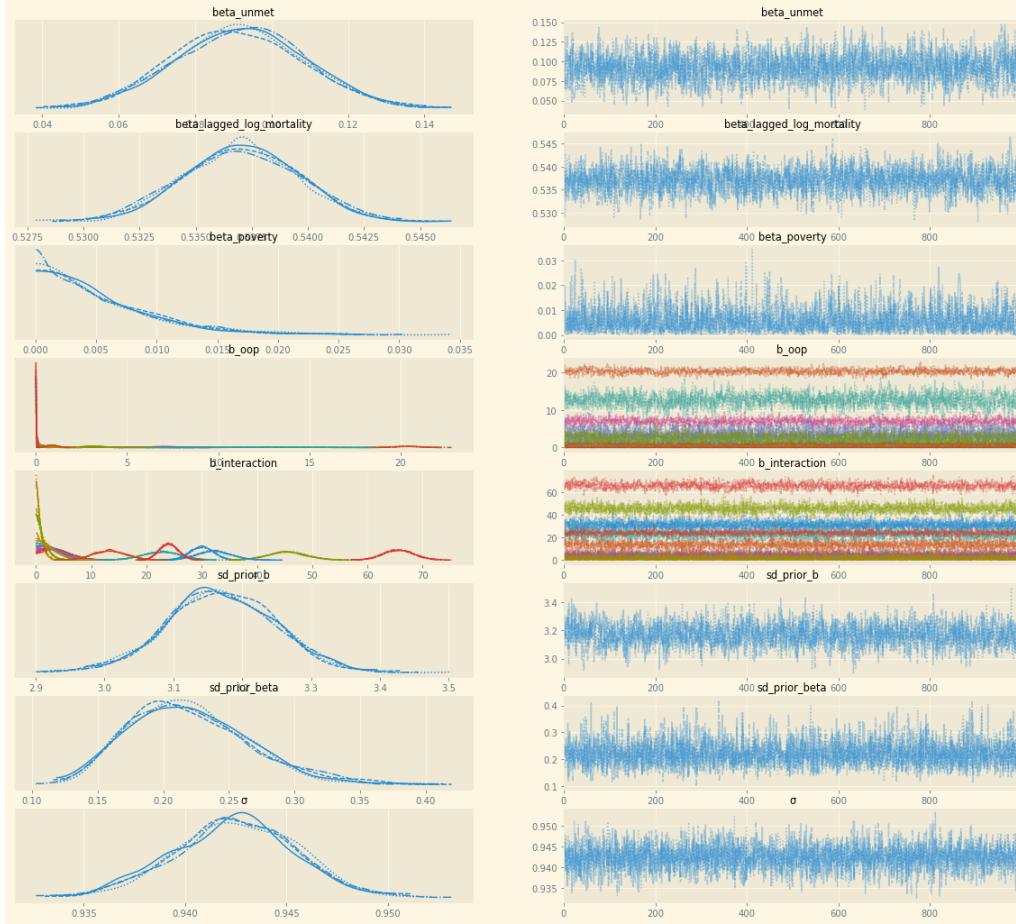


Figure 10: Trace plots of the coefficients of interest

### C.3 table of coefficients baseline model

Table 5 provides summary statistics for the posteriors of the coefficients we are interested in. For some countries the hdi\_3% lower bound for the `b_oop` or `b_interaction` equals zero. This is compatible with the OOP effect on mortality being bounded away from zero as the coefficients can be correlated: say, low `b_oop` going together with high `b_interaction` leading to an overall strictly positive effect.

Hence, to understand the mortality effects of an increase in oop, we use equation (8) with the posterior distributions substituted in for all the parameters. This gives us Figure 2.

#### *C.4 derivation of the effect of oop on mortality*

As we assume that **TooExp** has a logit-normal distribution, the derivative of the expression in Lemma 1 with respect to  $OOP\bar{x}$  is given by

$$\frac{dT_{ooExp}}{d(OOP\bar{x})} = T_{ooExp}(1 - T_{ooExp})(b_{oop,c} + b_{interaction,c}Poverty_{2t})$$

In the simulation we work with a 500 euro increase in oop:  $d(OOP\bar{x}) = 500$ . That is, we multiply the fraction of healthcare expenditure paid oop with average healthcare expenditure per head. This gives us oop in euro terms. We assume that the increase in **TooExp** translates one-for-one in an increase in **Unmet**. Hence, the change in mortality is given by:

$$\frac{dm_{ga2t}}{m_{ga2t}} = \beta_{unmet}T_{ooExp}(1 - T_{ooExp})500(b_{oop,c} + b_{interaction,c}Poverty_{2t})$$

This is the increase in deaths per one dead. In Figure 2 we multiply this expression by 1000: number of deaths per 1000 dead.

Table 5: Summary statistics for estimated coefficients

	mean	sd	hdi_3%	hdi_97%	ess_bulk	r_hat
beta_unmet	0.09	0.02	0.06	0.12	1638.00	1.00
beta_lagged_log_mortality	0.54	0.00	0.53	0.54	2048.00	1.00
beta_poverty	0.01	0.01	0.00	0.01	2161.00	1.00
b_oop[Bulgaria]	0.51	0.43	0.00	1.32	2180.00	1.00
b_oop[Croatia]	1.94	1.49	0.00	4.66	1539.00	1.00
b_oop[Denmark]	3.13	0.52	2.14	4.08	701.00	1.00
b_oop[Finland]	0.04	0.04	0.00	0.10	2376.00	1.00
b_oop[Greece]	20.35	0.65	19.11	21.53	939.00	1.00
b_oop[Hungary]	0.12	0.12	0.00	0.32	3139.00	1.00
b_oop[Ireland]	6.97	0.84	5.42	8.58	1505.00	1.00
b_oop[Lithuania]	3.60	1.67	0.01	6.15	694.00	1.01
b_oop[Norway]	0.02	0.02	0.00	0.07	2658.00	1.00
b_oop[Romania]	12.67	1.72	9.45	15.83	806.00	1.00
b_oop[Slovakia]	2.15	1.18	0.03	4.10	861.00	1.01
b_oop[Slovenia]	0.33	0.32	0.00	0.92	2681.00	1.00
b_oop[Sweden]	0.87	0.30	0.32	1.45	469.00	1.00
b_oop[Switzerland]	0.02	0.02	0.00	0.05	2071.00	1.00
b_interaction[Bulgaria]	30.02	2.10	26.40	34.25	1291.00	1.00
b_interaction[Croatia]	2.75	2.01	0.00	6.28	2685.00	1.00
b_interaction[Denmark]	45.43	3.38	38.50	51.30	1972.00	1.00
b_interaction[Finland]	0.70	0.65	0.00	1.92	3853.00	1.00
b_interaction[Greece]	3.91	2.40	0.01	8.03	1871.00	1.00
b_interaction[Hungary]	65.61	2.73	60.53	70.82	1175.00	1.00
b_interaction[Ireland]	3.55	2.29	0.00	7.58	2360.00	1.00
b_interaction[Lithuania]	3.01	2.16	0.00	6.87	3090.00	1.00
b_interaction[Norway]	32.19	3.09	26.39	37.98	2068.00	1.00
b_interaction[Romania]	22.72	3.28	16.50	28.80	2241.00	1.00
b_interaction[Slovakia]	1.19	1.08	0.00	3.17	2880.00	1.00
b_interaction[Slovenia]	2.27	1.80	0.00	5.46	2425.00	1.00
b_interaction[Sweden]	13.32	2.97	7.84	18.96	2914.00	1.00
b_interaction[Switzerland]	24.05	1.67	20.77	26.99	2601.00	1.00
sd_prior_b	3.17	0.08	3.03	3.32	1507.00	1.00
sd_prior_beta	0.22	0.05	0.14	0.31	3745.00	1.00
$\sigma$	0.94	0.00	0.94	0.95	3649.00	1.00