

Dynamic effects of health insurance reform

Jan Boone*

September 11, 2025

Abstract

We estimate a dynamic structural model to assess the long-term health effects of demand-side cost-sharing. Using Eurostat regional data for 11 European countries (2008–2018), we track health transitions across cohorts defined by age, gender, income, and region. Our model shows that increased out-of-pocket payments reduce recovery rates, especially for low-income individuals. The resulting loss in quality-adjusted life years (qaly's) is significantly larger than what static estimates imply –up to twenty times higher for vulnerable groups. These findings highlight the importance of accounting for long-term dynamics in policy evaluation and underscore the unequal burden of cost-sharing on population health.

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

Healthcare policy decisions often unfold over long time horizons, yet much of the empirical evidence on health insurance reform focuses on short-run outcomes. In particular, studies of demand-side cost-sharing typically rely on difference-in-differences (DiD) designs that estimate health effects over a narrow window around the implementation of a policy. While these approaches yield credible short-run causal estimates, they are poorly suited to capturing the cumulative, long-term consequences of forgone care, especially when mortality is used as the outcome of interest.

This paper takes a dynamic approach to analyzing the health effects of cost-sharing. We develop and estimate a structural model of health transitions, mortality, and insurance generosity using regional-level Eurostat data from 11 European countries over the 2008–2018 period. Our framework follows cohorts by age, gender, region, and income, distinguishing between individuals in high and low health states, and modeling transitions using a Markov process. Importantly, we allow treatment probabilities –and thus recovery rates– to depend on out-of-pocket spending and unmet medical needs. This enables us to quantify both short- and long-run effects of cost-sharing on mortality.

We show that increases in out-of-pocket costs lead to reductions in health, disproportionately affecting low-income populations. Crucially, the dynamic effects are an order of magnitude larger than static DiD estimates would suggest. Our results highlight a key policy evaluation trade-off: methods that yield clean identification in the short run may vastly understate the true long-run burden of cost-sharing, particularly for the most vulnerable groups.

Our estimated model fits the data well and replicates a number of stylized facts. Women have higher health status than men and people on high income higher than people in poverty. The probability of falling ill increases with age. Although there is uncertainty about the parameter estimates, there is an unambiguous ordering of countries in terms of expected qaly’s for 45 year olds; over the 2008-2018 period Switzerland has the highest expected qaly’s and Hungary, Bulgaria and Lithuania the lowest.

We contribute to the literature on the health effects of demand-side cost-sharing by emphasizing dynamic consequences and the disproportionate impact on low-income individuals.

Several studies have established causal effects of health insurance on mortality using individual-level U.S. data and policy variation from the Affordable Care Act

(ACA). For instance, the Medicaid expansions have been shown to reduce mortality using DiD approaches (Borgschulte and Vogler 2020; Miller, Johnson, and Wherry 2021). However, these studies focus on short-run effects –typically two to four years post-reform– and do not capture long-run health dynamics.

Other studies also link higher out-of-pocket (oop) costs to reduced healthcare use and higher mortality (Chandra, Flack, and Obermeyer 2021), but again only in a static or within-year framework. Similarly, Goldin, Lurie, and McCubbin (2020) find mortality effects in an RCT encouraging ACA compliance, but focus on a two-year window and individuals aged 45–64.

Our contribution differs in three main ways. First, we use European data, where insurance systems are more homogeneous and nearly universal –over 99% of the population is typically covered (see the entry on Health insurance coverage with Government/compulsory health insurance from the OECD Data Explorer). This allows for a clearer link between policy variation and health outcomes at the population level.

Second, we estimate a dynamic structural model tracking health status transitions over time. This enables us to compare long-run effects of cost-sharing with short-run DiD-style estimates. Our results show dynamic effects for low-income groups are magnitudes larger than static ones. We illustrate this by using the estimation method in Finkelstein and McKnight (2008) to explain the difference in detail.

Third, by using EU-SILC data on unmet medical needs, we explore a concrete mechanism: individuals forgoing care due to cost, especially at low incomes. This aligns with previous findings on liquidity constraints and care deferral (Gross, Layton, and Prinz 2020; Nyman 2003). While high-income individuals may also forgo treatment, it is less driven by financial barriers (Brot-Goldberg et al. 2017; Chandra, Flack, and Obermeyer 2021). Our results reflect this, showing much smaller effects in that group.

The next section presents a dynamic model of health status and mortality. Then we describe the Eurostat data that we use. We explain the empirical model that we estimate. Estimation results are presented and we conclude with a discussion of the policy implications. The appendix contains 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.¹ This is an advantage of

¹See the github repository: <https://github.com/janboone/dynamic-effects-of-health-insurance-reform>.

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.

2 THEORY

The idea of our dynamic model is to follow a cohort aged a at time t across calendar years to age $a + 1$ and year $t + 1$. We capture this dynamic with a straightforward Markov model. The model that we estimate is illustrated in Figure 1. We follow individuals from age 45 to age 85. Let's denote the age at which we start following an individual by $a = 0$ at time $t = 0$. If this agent is healthy, there is a probability π that the agent falls ill and moves to state ill at age $a = 1$ in period $t = 1$. With probability $1 - \pi$ the agent stays healthy also at age $a = 1$. For an agent in the unhealthy state, there is a probability σ that she is cured and becomes healthy at age $a = 1$. With probability δ she dies and with probability $1 - \sigma - \delta$ she remains in the unhealthy state at $a = 1$. Finally, death is an absorbing state.

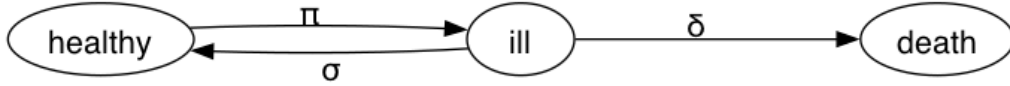


Figure 1: Markov chain model

Instead of deriving the steady state outcome of the model, we estimate parameters using the underlying difference equation. Focusing just on age for the moment to ease notation, the model implies the following difference equation. Let ι_a denote the fraction of people aged a who are ill and $1 - \iota_a$ the fraction who are healthy, then ι_a evolves over age (and time) as:

$$\iota_{a+1} = ((1 - \sigma - \delta)\iota_a + \pi_a(1 - \iota_a))/(1 - \delta\iota_a) \quad (1)$$

and $\delta\iota_a$ people die at age a . In words, the fraction of people ill aged $a + 1$ equals the fraction ill aged a who do neither recover to the healthy state nor die plus the fraction of healthy people who fall ill. We normalize by $1 - \delta\iota_a$ to ensure that the fractions ill and healthy add up to one.

We differentiate the fraction ill, ι_a , by region/year/gender/income. If we add all the sub/superscripts we get ι_{atr}^{ij} for age a in year t and region r among gender $i \in \{f, m\}$ and low/high income $j \in \{l, h\}$. Let α_{tr} denote the fraction of people on low income in region r and year t which is available in our data. We observe in

our data mortality μ_{atr}^i , the fraction of females/males i who die at age a in year t in region r . In terms of our model this is given by

$$\mu_{atr}^i = \alpha_{tr}\delta\iota_{atr}^{il} + (1 - \alpha_{tr})\delta\iota_{atr}^{ih} \quad (2)$$

As we do not observe ι directly in our data, we have to estimate the starting points of the fraction of people initially ill, ι_0 , in order to solve the difference equations over time. Our estimates of ι_0 vary by age/gender/region/income. The reason ι_0 varies by age is illustrated in Table 1. Differentiated by gender/region/income we follow the cohorts aged 45-74 over time until they are aged 55-84. The data section below motivates this choice. Hence we estimate the initial values ι_0 for the ages 45-74 in the year 2008. Then we apply equation (1) to determine the development of ι_a over time as cohorts get older.

Table 1: We follow cohorts aged 45-74 over time from 2008 until they are 55-84 in 2018.

year	age											
2008	45	46	47	48	...	68	69	70	71	72	73	74
2009	46	47	48	49	...	69	70	71	72	73	74	75
2010	47	48	49	50	...	70	71	72	73	74	75	76
2011	48	49	50	51	...	71	72	73	74	75	76	77
2012	49	50	51	52	...	72	73	74	75	76	77	78
2013	50	51	52	53	...	73	74	75	76	77	78	79
2014	51	52	53	54	...	74	75	76	77	78	79	80
2015	52	53	54	55	...	75	76	77	78	79	80	81
2016	53	54	55	56	...	76	77	78	79	80	81	82
2017	54	55	56	57	...	77	78	79	80	81	82	83
2018	55	56	57	58	...	78	79	80	81	82	83	84

The remainder of the section explains how π, δ, σ vary with age/year/gender/region/income. First, the probability of falling ill π varies by age/gender/country/income. We know that older people tend to be less healthy than young people, women tend to live longer than men, longevity varies by country and people on low income tend to have lower health status than people on high income. We allow π to vary along these dimensions to capture these stylized facts.

Second, once people are in state “ill” we assume that their probability of death δ is the same across age/gender/income. We do allow δ_c to vary by country as some

healthcare systems may be better in prolonging life for the ill than systems in other countries.

A final source of variation is the probability σ with which people recover to full health. This probability σ is determined by the probability of treatment and the quality of treatment conditional on being treated. We have data on unmet medical needs. Reasons for unmet medical needs, that is reasons why people forgo treatment include waiting lists, hospital too far away, afraid of treatment. The motivation that we focus on is that the treatment is skipped because it is too expensive. Let $v \in [0, 1]$ denote the fraction of people who have unmet medical needs. We write this as the sum of the fraction of people who indicate that treatment is too expensive, τ , and the fraction of people who give other reasons, v_0 .

$$v = v_0 + \tau$$

We expect that τ is increasing in the fraction of healthcare expenditures that people pay out-of-pocket, *oop*, where *oop* is measured as expenditure paid out-of-pocket over total healthcare expenditure. We interpret this as capturing the generosity of a country's health insurance system. We use a linear approximation of this relation:

$$\tau = \zeta oop$$

If healthcare is completely free at point-of-service, $oop = 0$ and $\tau = 0$: with free healthcare, no one forgoes treatment because it is too expensive. At the other extreme, no health insurance at all: everything is paid out-of-pocket: $oop = 1$. Then a fraction $\zeta \in [0, 1]$ will indicate they forgo treatment because it is too expensive.

We allow ζ to depend on income with $\zeta^l \geq \zeta^h$. We expect people on low income to react more strongly to an increase in *oop* than people on high income. Also people on low income are more likely to avoid treatment for other reasons: $v_0^l \geq v_0^h$. To illustrate, people on low income may be less skilled, say, in navigating waiting lists or find it more difficult to travel to a hospital that is further away (or even abroad).

If someone gets treatment, the probability of recovering to full health is given by λ . Hence, we write the probability of recovery σ as

$$\sigma = \sigma_0 + \lambda(1 - v) \tag{3}$$

where σ_0 is the probability that someone recovers without treatment. The parameter λ captures the quality of care which we allow to vary by region and year in ways explained below.

We consider two applications of our model. First, as a summary statistic of health we calculate the expected quality adjusted life-years (qaly's) of 45 year olds till they are 85. We normalize the value of a life-year in full health to 1.0 and set the qaly for a year being ill equal to $\theta \in \langle 0, 1 \rangle$. The value of a “life-year when dead” is normalized to 0.0. To make this qaly calculation, we keep track of the size of the groups healthy and ill as follows:

$$N_{a+1}^{healthy} = (1 - \pi_a)N_a^{healthy} + \sigma N_a^{ill}$$

and

$$N_{a+1}^{ill} = \pi_a N_a^{healthy} + (1 - \sigma - \delta)N_a^{ill}$$

where we start with 45 year olds at $a = 0$ and calculate up to age A (corresponding to 85 year olds). In contrast to equation (1) we do not normalize $N_a^{ill} + N_a^{healthy} = 1$; indeed, the point is that we lose life-years as age increases. As above $N_a^{healthy}, N_a^{ill}, \pi_a, \sigma, \delta$ vary with gender/region/income.

We then calculate expected quality adjusted life-years from age 1 to A as the undiscounted sum of qaly's:²

$$Q = \sum_{a=1}^A N_a^{healthy} + \theta N_a^{ill} \quad (4)$$

Second, we consider the effect on qaly's of an increase in demand-side cost-sharing $\Delta oop > 0$. For this, we calculate qaly's in the baseline outcome, Q_0 and compare it with the outcome where the fraction of people with unmet medical needs increases with $\Delta v = \zeta \Delta oop$. We denote the qaly's in this outcome by Q_1 . The loss of qaly's due to increased demand-side cost-sharing is given by:

$$\Delta Q = Q_0 - Q_1$$

In words, suppose there was a reform in year $t = 0$ which increased oop . After the reform we follow the cohort aged $a = 0$ over time till age $a = A$ and calculate their qaly's, Q_1 . We compare these qaly's with the counterfactual where oop was not changed, Q_0 . We denote this the dynamic comparison.

We compare this dynamic analysis with the following static one. Instead of following an age cohort over A calendar years, we consider A ages one period after the reform and denote their qaly's by \tilde{Q}_1 . That is, the equation for \tilde{Q} is comparable

²We use an undiscounted sum here to make it easier to compare to the static version \tilde{Q} discussed below.

to (4) except that in equation (4) both age a and calendar year t vary over the summation. With \tilde{Q} calendar year is fixed at $t = 1$ (with the reform happening in calendar year $t = 0$). We then compare \tilde{Q}_1 with the qaly's these A age groups would have received under the counterfactual where oop is unchanged, \tilde{Q}_0 .

To illustrate with Table 1: we start with ages 45-74 at $t = 0$. We have the reform and calculate \tilde{Q}_1 for the second row in the table (ages 46-75). Then we compare this to qaly's for the second row under the counterfactual of no change in oop . We argue that this captures the estimate made by an event study like a DiD analysis, say where the reform was introduced in one region but not in another region. The DiD approach would then control for year fixed effects.

This type of analysis can identify the causal effect convincingly but is usually applied for a couple of years only. For instance, because in later years the reform is introduced in other regions as well. Or because over time other factors may change causing a bias in the estimates based on the comparison of the regions.

Theoretically, it is plausible that the dynamic estimate exceeds the static one. In a model where stocks (of healthy and ill) determine the state of the system and its development over time, ignoring these dynamic effects is likely to cause an under-estimation of the effect on qaly's or mortality. With our estimated model we can quantify the difference between the static and dynamic estimates.

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 11 countries and 65 regions for the 11 years 2008-2018 and ages 45-84 for women and men. The years 2008-2018 were chosen because, at the time of the analysis, for most variables and regions data was available from 2008 onward till 2018. The age range was chosen with the idea that mortality starts to increase from 45 years onward and number of observations per region drop significantly from age 85 onward. Finally, NUTS 2 regions were selected where there was at least one observation for unmet medical needs over the years and where the number of deaths during the year does not exceed the population size at the start of the year.

Table 2 shows the summary statistics for our variables. We briefly discuss the main variables, the appendix provides more detail. We have more than 40k observations for population and deaths per age/gender/year/region category.³ We have

³The max. number of observations that we could have is: 65 (regions) * 11 (years) * 30 (ages)

missing observations and explain below how we deal with these in our Bayesian setting.

The average population size per region-age-gender category is about 6000 and the average number of deaths 90. Median population size per category equals 5300 and median number of deaths 62. 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 17.3% for some region and age combination. The material deprivation measure (denoted **deprivation**) comes from the EU statistics on income and living conditions (EU-SILC) survey. It refers to the enforced inability to pay unexpected expenses, afford adequate heating of the home, durable goods like a washing machine etc. This variable captures the idea that people may feel forced to forgo (valuable) treatment because it is too expensive.⁴ The mean deprivation rate across regions and years equals 10%. Some regions have (almost) no people in deprivation, while in others 55% of people live in deprivation. The median rate is around 3%.

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. The mean fraction of people forgoing treatment because it is too expensive equals 1.5%, while the mean fraction of people with unmet medical needs (because it is too expensive or other reasons) equals 6.5%. In some regions the fraction of people forgoing treatment because it is too expensive goes up to almost 20% and the fraction of people with unmet medical needs to almost 30%.

For the model application considering the effect of demand-side cost-sharing on health and mortality, healthcare quality could be a collider. Indeed, if government resources are reduced, there could be both an increase in out-of-pocket payments and a reduction in healthcare quality (by reducing investments in technology and/or

* 2 (genders) = 42,900 which we have in first three rows. For the next 5 rows in the table we have data for at max. 11 years and 65 regions which corresponds to 715 observations and for out-of-pocket we have at max. 11 years and 11 countries corresponding to 121.

⁴Eurostat also provides the variable “at-risk-of-poverty” per NUTS 2 region. 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. But the link between this poverty measure and forgoing treatment because it is too expensive is weaker. Further, there are more missing observations for this variable than for the deprivation variable.

reducing the number of physicians). This reduction in quality would also affect health and mortality thereby confounding the healthcare demand effect of out-of-pocket spending. We use two variables to control for the quality of care. First, infant mortality is a well known measure for the quality of a healthcare system. Moreover, infant mortality is not directly related to our measure of mortality which starts at age 45. On average there are almost four dead infants (younger than one year of age at death) per 1000 live births. But for some regions this almost reaches 16 per 1000 live births. The second quality measure is the number of physicians per 100k inhabitants. On average there are 380 physicians per 100k inhabitants but this varies between regions from 180 to 800. We view this as an approximate measure of resources available for healthcare in a region. Finally, we use the fraction of healthcare expenditure paid out-of-pocket as a signal of how generous health insurance is. This captures country wide policy. On average people pay a bit more than 20% out-of-pocket but this varies from 12% in some countries to more than 40% in others.

Table 2: Summary statistics main variables

	count	mean	std	min	median	max
population	42900	5961.0	3788.7	574.0	5286.0	30491.0
deaths	42900	88.9	83.7	0	62.0	867.0
mortality (%)	42900	1.8	1.9	0	1.2	17.3
deprivation (%)	481	9.9	13.3	0	2.8	55.2
too expensive (%)	527	1.5	2.6	0	0.5	18.4
unmet (%)	527	6.4	4.6	0.7	5.1	28.1
infant mortality (‰)	705	3.8	2	0.8	3.4	15.8
physicians per 100k inhab.	594	379.4	108.6	180	363.3	807.9
out-of-pocket (%)	55	22.8	9.5	12.0	18.9	43.2

4 ESTIMATION

We estimate the model using Bayesian methods. Here we present the equations of the model. The online appendix contains the details on the choice of priors and the `pymc` code of the model. Further, we explain how the Bayesian algorithm deals with missing observations.

We present the model in reverse order. We observe the number of deaths m

and the population size n per age/gender/region/year dimensions. We model the probability of death μ and m has a Binomial distribution:

$$m_{atr}^i = \text{Binomial}(n_{atr}^i, \mu_{atr}^i) \quad (5)$$

where μ_{atr}^i is given by equation (2). Poverty α_{tr} per region and time period is observed in the data using the material deprivation measure discussed in the previous section. The fraction of people ill ι_{atr}^{ij} varies by age a , gender i , time t , region r and income level j . This fraction ι follows from difference equation (1). Before we explain how this equation is estimated, recall that the estimate of δ_c is allowed to vary by country c .

To solve differential equation (1), we start by estimating the initial condition ι_{a0r}^{ij} for ages $a = 0 - 30$ (i.e. ages 45-74) in year $t = 0$ (2008) per region r , gender i and income category j .

Next, we need to estimate the probability of falling ill π_{ar}^{ij} which we assume to be additive in age, region, gender and income fixed effects in log-odds space. That is we add together the log-odds effects for age, gender, region and income and apply the inverse logit function $e^x/(1 + e^x)$ to the sum of these log-odds. This gives the probabilities $\pi_{ar}^{ij} \in \langle 0, 1 \rangle$.

The probability $\sigma = \sigma_0 + \lambda(1 - v)$ of recovering to full health in equation (3) is estimated as follows. We estimate the probability $\sigma_0 \in \langle 0, 1 \rangle$ of recovering to health without medical intervention. Further, we observe the fraction of people who indicate that they have unmet medical needs v_{tr} at the year/region level. For the ill who do get treatment, $1 - v$, there is a probability λ of full recovery. The log-odds of λ are the sum of three effects: (i) year fixed effects, capturing that over time the quality of care improves due to technological progress, (ii) infant mortality and (iii) number of physicians per 100k inhabitants. The latter two capture the variation across time and region in quality of care. Again we use the inverse logit function to turn the log-odds into a probability λ .

We expect both the probability of unmet medical needs because it is too expensive, τ , and for other reasons, v_o , to be bigger for low incomes than for high incomes. If your income is higher, you are less likely to skip a treatment because it is too expensive: $\tau^j = \zeta^j oop$ with $\zeta^l \geq \zeta^h$ and the fraction of medical expenditure paid out-of-pocket oop_{tc} varies with country and time. Further, also with the other reasons for unmet medical needs (like waiting lists) we expect $v_o^l \geq v_o^h$. Hence, we find that

$$v_{tr} = \alpha_{tr}(\zeta^l oop_{tc} + v_{otr}^l) + (1 - \alpha_{tr})(\zeta^h oop_{tc} + v_{otr}^h)$$

where we observe $\tau_{tr} = (\alpha_{tr}\zeta^l + (1 - \alpha_{tr})\zeta^h)oop_{tc}$ and $v_{otr} = \alpha_{tr}v_{otr}^l + (1 - \alpha_{tr})v_{otr}^h$.

Hence we find

$$\sigma_{tr}^j = \sigma_0 + \lambda_{tr}(1 - \zeta^j oop_{tc} - u_{otr}^j)$$

Given that we estimate the initial conditions ι_0 and the parameters π_a, σ, δ we can solve differential equation (1) to find ι_{atr}^i and thus calculate the probability of death μ_{atr}^i in equation (2) as input in the Binomial distribution (5).

One of the conceptual advantages of a Bayesian model is that there is no clear distinction between data and parameters. Both are basically distributions. This comes to the fore when dealing with missing values which are prevalent in our data as shown in Table 2. The Markov Chain Monte Carlo (MCMC) algorithm generates the posterior distribution of parameters by drawing samples from this distribution. For each parameter we will have 4 chains of 2000 draws, that is 8000 draws from the posterior distribution. When an observation is missing, the algorithm will randomly draw the observation from a pre-defined distribution. This implies that each draw is different signalling the uncertainty that surrounds the missing observation. This is better than imputing the value of the observation as the latter suggests that we are as certain about the imputed value as we are about an observed value; which is obviously not the case.

For variables based on summation, it is natural to assume a normal distribution for missing observations. The expectation (standard deviation) of this distribution is then given by the average value (standard deviation) of the observations that we do have for this region or country.

To improve the efficiency of the algorithm in solving the difference equation while estimating the parameters, the data is not in (two dimensional) dataframe format. Instead we work with multi-dimensional tensors. To illustrate, the tensors with the population and deaths data are 4-dimensional tensors with coordinates age/gender/year/region. The difference equation then iterates over age and calendar year using pytensor's `scan` function. See the online appendix for details.

4.1 Identification strategy and limitations

Our identification strategy is grounded in a structural modeling approach rather than exogenous quasi-experimental variation. We estimate a dynamic Markov model of health status transitions using cohort-level Eurostat data, where the key policy variable, *oop*, varies across countries and over time. We link these variations in cost-sharing to observed differences in recovery rates and mortality, conditional on

region, age, gender, income, time fixed effects and variables capturing healthcare quality.

The strength of this approach lies in its ability to simulate long-run effects by modeling underlying health dynamics. Unlike DiD estimators, which compare limited time points before and after a reform, our framework accumulates the effects of delayed or forgone treatment over multiple years. It also allows us to examine heterogeneity by income and gender and to propagate parameter uncertainty through Bayesian inference.

However, this comes with trade-offs. Our identification of the effects of cost-sharing relies on structural assumptions, notably the specification of the Markov transition probabilities. Further, *oop* variation is not derived from a policy discontinuity or random assignment. As such, our estimates could be biased by omitted variables that simultaneously influence health and insurance *oop* (e.g. austerity-induced cuts to both healthcare quality and insurance generosity). We address this concern through robustness checks: (i) the inclusion of unemployment and education to capture lifestyle differences between regions and GDP per capita and imaging device availability as additional controls for healthcare quality and (ii) a placebo test using traffic mortality which should not have the dynamic structure of mortality that we analyse but can be negatively affected by fiscal austerity.

While these checks support the plausibility of our identification, we acknowledge that the approach does not deliver the same level of internal validity as natural experiments. Our findings should therefore be interpreted as credible simulations under a structural model, rather than precise causal point estimates in the experimental sense.

5 RESULTS

In this section we present the results of the estimation of the model. Before presenting the outcome of our estimation, we present some (graphical) checks of our model.

5.1 *Model fit*

A first check of the model is to see whether it fits the data. Figure 2 shows for all our observations the observed number of deaths on the horizontal axis and the predicted number on the vertical axis. Further, for each prediction we add a dashed

vertical line representing the 95% prediction interval (of the posterior predictive distribution). The predictions clearly follow the 45-degree line and almost all 95% prediction intervals intersect the 45-degree line. In this sense the model fit seems quite good.

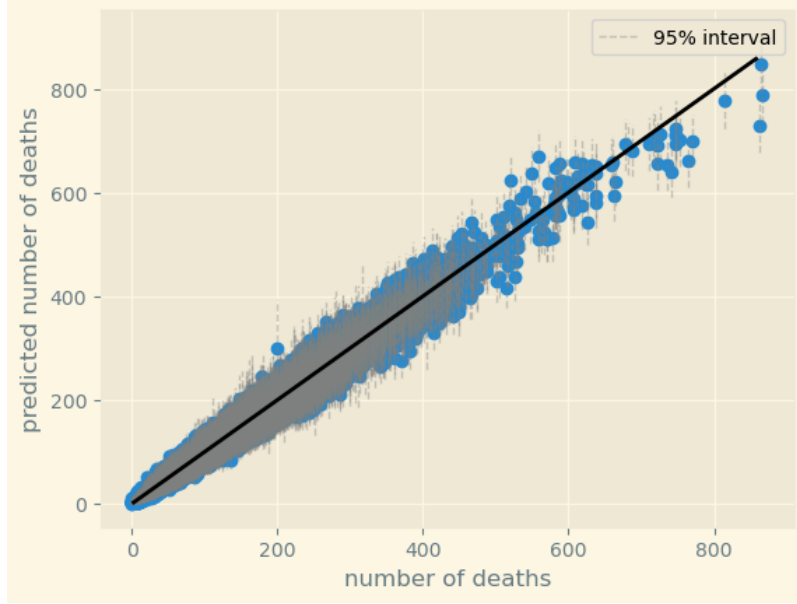


Figure 2: Observed vs predicted number of deaths (per age/gender/region/year) together with the 95% prediction intervals.

Figure 7 in the appendix presents the trace plots and Table 8 \hat{r} values for estimated parameters. The \hat{r} values are close to one and the traceplots satisfy the following three criteria. First, the plots in the right column of the figure are 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 is 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 one after the other. Finally, the four chains cover the same regions. This is most easily checked in the left column of the trace plot. The three features are satisfied in the figure. All this indicates that the MCMC algorithm manages to sample the posterior distribution correctly.

To gain an intuition on how the model works, Figure 3 shows a number of estimated outcomes. The top-left panel shows the probability of falling ill, π , as a function of age for women and men.⁵ This probability is quite low around age 45 and increases to 40% for men in their eighties. As shown in the panel below this

⁵These probabilities are calculated by averaging π_{ar}^{ij} over region r and income j .

one, the probability of death –conditional on being ill– is above 10% per year. In other words, falling ill in our model does not refer to a regular cough or breaking a leg. The observation that the average probability of falling ill is higher for men than for women is in line with the observation that women tend to live longer than men. The 95% posterior probability interval shows that there is quite some variation in the probability of falling ill at older ages but that uncertainty is small till age 75. In order not to crowd the figure we do not report the 95% intervals for males but they are of similar magnitude.

The top-right panel shows the country offsets (in log-odds terms) for four countries. There is a clear dichotomy between Switzerland and Norway on the one hand and Lithuania and Hungary on the other. But even between these countries there is hardly overlap between the distributions. Of these four countries, people in Switzerland are least likely to fall ill (at any age) and in Hungary most likely.

Similarly, the probability of dying conditional on being ill, δ_c , varies by country. Of the four countries presented, it is lowest for Denmark and Finland and highest for Hungary and Bulgaria.

Finally, the bottom-right panel shows the fraction of people ill for four categories: people aged 55 and 60 with low/high income. The fraction of people who are ill is higher for low than for high incomes, as one would expect. Also the fraction increases with age. The posterior distributions for these fractions do not overlap.

5.2 Analysis

The analysis in this section makes the following points. First, the expected qaly's for a 45 year old vary a lot between countries and within countries between gender and income categories. The concept of qaly's is much discussed in the literature; here we simply use it as a summary measure of the fractions of people in good health, poor health and deceased. Second, when we increase demand-side cost-sharing by ten percentage points ($\Delta oop = 0.1$), there is a clear but modest reduction in qaly's. Third, the effect is almost zero for people on high income but far higher for people on low income. In this sense, effect heterogeneity is high and considering the average effect (across incomes) is not informative. Finally, instead of comparing qaly's lost over, say, a thirty year time period (for which one needs a dynamic model), we can compare the loss in qaly's for the same age groups from one year before the increase in *oop* to one year after. The latter is close to a (static) measure used in most event study analyses. We show that the dynamic effect is an order of magnitude bigger

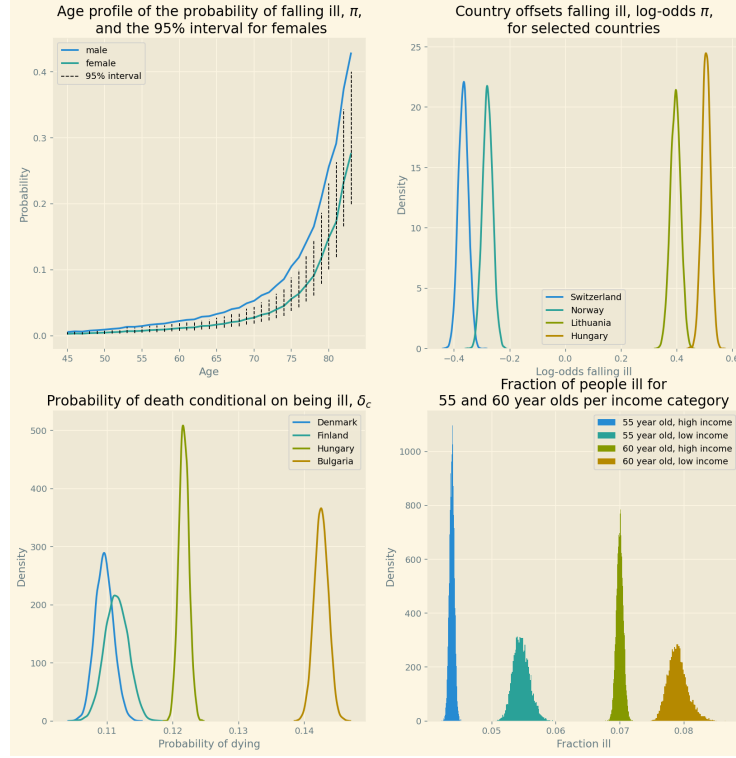


Figure 3: Summary of selected features of the model. Top left panel shows the age profile for π for women and men. Top right panel shows the country offsets in the log-odds of π and bottom left panel the probability of death δ_c for selected countries. Bottom right panel shows the fraction of two age groups who are in the low health state for low and high incomes.

than the estimated static effect.

Figure 4 summarizes the two applications that we analyse with the model. First, since our estimation method identifies the fraction of people who have low health status, we can derive qaly's using equation (4). In the figure we use $\theta = 0.6$: the quality of life for someone who is ill (and has a 10% probability of dying per year) is 60% of the quality of life for someone in full health. This is an arbitrary choice and obviously different choices can be made here. Different values for θ changes the precise numerical values but not the main (qualitative) conclusions of the analysis; as we show in the robustness analysis.

The figure was made as follows. For each group (female/male, low/high income) of 45 year olds we create a population of 1000 individuals. At 45 we use our estimates of ι_0 to split the groups into low/high health status. With our estimated parameters π, σ, δ we follow these individuals as they age over time till they are 85. Hence, for each group there are at max. 40,000 life-years (in full health) to be had over this

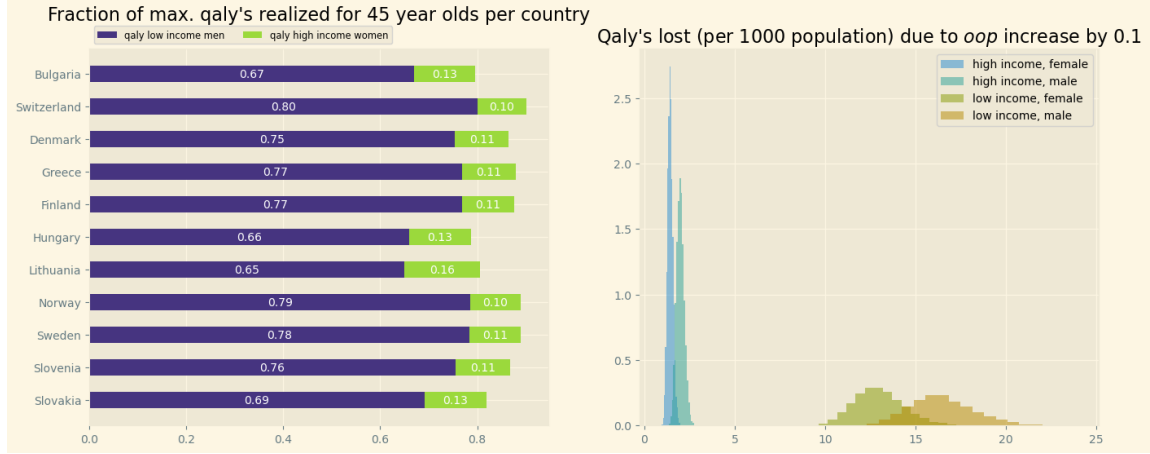


Figure 4: Summarizing two applications of the model: left panel shows the fraction of the maximum qaly's realized by country and the right panel shows qaly's lost due to 0.1 increase in *oop*.

period. For each subgroup gender/income we calculate the realized quality adjusted life-years and divide this by 40,000. The left panel of Figure 4 shows the qaly's realized for low income men and for high income women. The panel shows that in, say Switzerland, low income males get 80% of the max. number of qaly's and high income females get 10% of 40,000 qaly's more. That is, the latter get 12.5% ($0.10/0.80$) more qaly's than the former. This gender/income gap ranges all the way up to 20% in Hungary, Bulgaria and Lithuania.

The panel on the right shows what happens if *oop* is increased by 0.1: people pay an additional 10% of their healthcare expenditure out-of-pocket for each country in our data. It turns out that the average effect is around 3.0:⁶ among 1000 people, a region/gender/income group loses 3 life-years for 45-85 year olds due to $\Delta oop = 0.1$. This effect is not huge (people do not just pass away because cost-sharing is increased), but it is clearly bounded away from 0.

What is more surprising is the effect heterogeneity for high and low incomes. The effect is almost ten times bigger for people on low income. People on high income hardly react in terms of healthcare expenditure to the increase in *oop* as they can afford it anyway. But for low incomes, the increase reduces their demand for treatment. This raises unmet medical needs because it is too expensive and hence, for this group, reduces the probability σ of recovering to the healthy state. This has two effects on their qaly's. First, the quality of life in the ill state is

⁶This overall effect is not reported separately in the figure.

lower than in the healthy state. Second, staying longer in the ill state, increases the probability of death (compared to being in the healthy state).

Finally, we compare the dynamic (across years) qaly effect derived above with a year-on-year effect for the same ages. To avoid making additional assumptions we do this for the ages 45-74 in the first row of Table 1. This is the cohort of ages that we use to estimate the model. Hence we have a well defined ι_0 fraction of people that are ill per age/gender/region/income category in the year before the reform. Then we use the equations for \tilde{Q}_0, \tilde{Q}_1 defined in the theory section to find the static effect on qaly's. We compare this to the dynamic estimate in equation (4) for the same ages till $A = 75$.

Table 3: Dynamic vs. static estimates of qaly's lost due to $\Delta oop = 0.1$.

	male	female	male	female
estimate	low income	low income	high income	high income
$Q_0 - Q_1$	8.01	5.23	0.82	0.46
$\tilde{Q}_0 - \tilde{Q}_1$	0.31	0.24	0.12	0.06

As shown in Table 3, the dynamic low income estimates are lower than in Figure 4 (right panel): the dynamic estimates across ages till 75 are lower than over the full range till 85.

Comparing static and dynamic estimates shows that for this type of health model the dynamic estimates of qaly's lost $Q_0 - Q_1$ are a factor twenty higher for low incomes than the static estimates $\tilde{Q}_0 - \tilde{Q}_1$. For high incomes the factor is around seven. As noted in section 2, we expect the dynamic effect to be bigger than the static effect because the former takes the effect on the stock of ill/healthy into account. However, the size of the difference is larger than one might have expected.

This suggests that for health models where an underlying (latent) health variable is important, short-term estimates (as often with a DiD analysis) severely underestimate the overall effect of a policy change.

One can see the following trade-off. A DiD approach leads to a cleaner causal identification. It often allows for a convincing way to keep other factors constant. In contrast, identification in our model (which allows for a dynamic analysis) is based on underlying theory but suffers from potential missing variable bias. We show that for reasonable parameter values, the static estimate seriously under-estimates the effect compared to a dynamic one taking account of the cumulative effects over a life-time. In particular, we simulate the data using our dynamic model as true data

generating process. Then we use static estimators to see what effects they identify.

Figure 5 illustrates how a static estimate like DiD ignores the dynamics of health. The figure is an abstract representation of the DiD estimate of Finkelstein and McKnight (2008) analysing the introduction of Medicare in 1966 providing universal public health insurance to individuals aged 65 and over. To find the health effect of Medicare, the paper compares the average mortality rate of 65-74 year olds with the rate for 55-64 year olds both before and after the introduction of Medicare; the years 1952-1975.

The logic of comparing age category 55-64 with 65-74 is the underlying assumption that the two categories do not differ except for the introduction of Medicare. This assumption would be harder to motivate when comparing 55-64 with 82 year olds. The idea of the paper is that if Medicare has substantial health effects one expects that the mortality rate for 65-74 year olds falls compared to 55-64 year olds after 1966. The paper does not find such an effect and the figure illustrates one mechanism why this can be the case.

To place this analysis in the context of our paper, we simulate the effect of a 10% point increase in oop ($\Delta oop = 0.1$) in year 0 for people aged 65 and over. Hence, in our model this implies that mortality will increase for the 65+ category (not decrease as in the Medicare example). We can simulate both the reform and the counterfactual and hence can make the comparison within age category; no need to compare the effect to 55-64 year olds. For this comparison between the DiD and dynamic estimates we only need the estimated parameters to be reasonable, not perfectly accurate. As shown above, this assumption seems to be satisfied with our estimated coefficients.

We do the simulations for low income men in one region in Bulgaria aged 65 and above who face $\Delta oop = 0.1$ in year 0. On the horizontal axis we plot age for this group. The vertical axis shows the years after the introduction in year 0. The color intensity of a point indicates the number of years an agent has faced this higher oop. In the first year everyone aged 65 and above faced the higher out-of-pocket for one year (bottom row in the figure). In the second year, people aged 66 and over faced this for two years and the 65 year olds only for one year etc.

The horizontal black line in the left panel indicates the comparison group used in the paper (65-74) for year 10 (last year in the Finkelstein and McKnight (2008) data). The figure illustrates how –from a dynamic point of view– the effect gets diluted as the 65 year olds in this group have only experienced the reform for one year, the 66 year olds for two years etc. In contrast the blue diagonal line gives the

true dynamic effect when we follow 65 year olds in year 0 till they become 83 with 19 years of increased oop. The diagonal includes points with a darkness intensity that the horizontal line does not capture for any of the years 0-18. In this sense, the static estimate under-estimates the true dynamic effect.

The right panel shows the consequences of this. The vertical axis shows the increase in mortality by age normalized on the average mortality effect for 65-74 year olds. Hence, the static estimate for 65-74 year olds in this figure is one by construction. But due to the dynamics of health in our model, the effect is first below 1 (till age 69) and then increases to six times the static estimate for 80 year olds. In order to identify this (far bigger) effect, one has to estimate a dynamic model capturing the age effects instead of relying on the DiD estimate. It will depend on the application whether a static precise causal identification is preferable or an estimate taking the dynamics of the problem into account but lacking a convincing causal foundation.

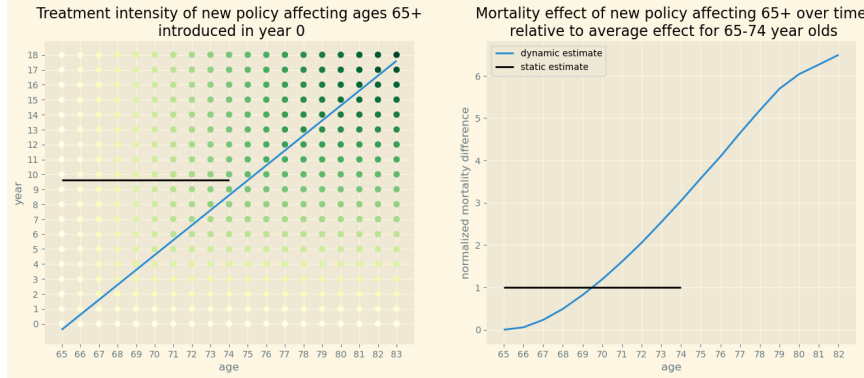


Figure 5: DiD vs. dynamic estimates of a 10% point increase in out-of-pocket expenditure for low income men aged 65 and over. The left panel shows the intensity of the effect across ages and years after the reform. The right panel shows the effect when following the 65 year old cohort till they are 83.

6 EXTENSIONS

In this section, we present two robustness checks to validate our baseline results. First, we assess potential confounding from broader macroeconomic conditions and healthcare quality. As a placebo test, we examine whether oop payments predict traffic mortality –a health outcome unlikely to be influenced by cost-sharing. We also augment the baseline model with GDP per capita. If the estimated effect of

the variable `TooExp` (unmet medical need due to cost) were to vanish with GDP included, it would imply that cost-sharing is merely a proxy for recession-induced fiscal constraints. In that case, higher mortality would stem from reduced healthcare quality rather than forgone treatment. To further address this, we include an additional healthcare quality proxy –the number of imaging devices per capita– alongside infant mortality and physician density. We also add education and unemployment as variables capturing lifestyle choices in a region.

Finally, we test the sensitivity of our results to the choice of θ , the relative value of a life-year in poor health compared to full health, used in our qaly computations.

6.1 Data

For the extensions we use the variables in Table 4 in addition to the variables in Table 2. We have data on the number of people that die in traffic accidents. The unit is number of deaths per 1 million inhabitants of the region. We have data on GDP per capita (purchasing power standard) and the percentage of people unemployed in the region. We have the percentage of people (aged 25-64) with tertiary education in the region. We only have this from 2013 onward. Finally, we have data on the number of devices for medical imaging (MRI’s, PET scanners etc.) per 100k inhabitants at the country level.⁷

Table 4: Summary statistics variables for model extensions

	count	mean	std	min	median	max
traffic mortality per 1m inhab.	679	66.2	41	6	58	254
GDP per capita	629	23829.3	11335.1	6600	21300	59500
unemployment (%)	679	9.7	6.5	1.8	7.7	31.6
tertiary education (%)	378	31.4	9.9	16.4	31.4	54.7
number imaging devices per 100k inhab.	121	40.7	47	0	27.8	205.7

On average 66 people per 1 million inhabitants die in traffics accidents per year. This ranges from six in one region/year combination to a maximum of 254 deaths. GDP per capita is on average 24k but ranges per region from 6600 to almost 60k. Unemployment is on average almost 10% and varies from less than 2% to more than

⁷For the first three variables we have eleven years of data for 65 regions which would mean 715 observations, but there are some missing values. Data on tertiary education only starts from 2013 and has max. 390 observations. Finally, for the imaging devices there are 121 observations and no missing values.

30%. Almost one third of the working population has tertiary education on average. This varies from 16% to more than half of the population. Finally, the number of imaging devices is on average 40 per 100k inhabitants but varies from no devices in one year for a country to more than 200.

6.2 *Extension of the model*

In the extension of the model, we adjust two parameters of the baseline model. First, in the equation for π we also include education and unemployment in the following way:

$$\pi e^{a_1 \text{education} + a_2 \text{unemployment}}$$

That is, we multiply the expression for π in the baseline model by a multiplier depending on education and unemployment. With more people having finished tertiary education one may expect more healthy lifestyle choices. Unemployment with the implication of lower income may lead to stress and/or worse lifestyle choices as fresh vegetables and fruit and, say, a gym membership are expensive. With the multiplier we capture that the effect of education and unemployment is small if the probability of falling ill is small to start with, say at age 45. At age 60 when the probability of falling ill is higher, the effect of unemployment and education is higher.

Second, in the equation for the probability of recovery conditional on treatment, λ , we add GDP per capita and the number of imaging devices as additional variables capturing healthcare quality. These variables we add in the same way as infant mortality and number of physicians in the baseline specification of λ .

With the estimated extended model we can do the same analysis as with the baseline model. Our main result is that the dynamic estimates are an order of magnitude higher than the static estimates. By comparing Tables 5 and 3 we see that this result is robust to the addition of the variables. The mean values of qaly's lost for the different categories are similar for the base and extended models. To illustrate that the differences in average qaly's lost are not "significant", Figure 6 shows the posterior distributions for qaly's lost for the female, low income category. Indeed, these distributions are almost the same for the base and extended models. The same is true for the other gender/income categories.

Table 5: Dynamic vs. static estimates of qaly’s lost due to $\Delta oop = 0.1$ in the extended model.

	male	female	male	female
estimate	low income	low income	high income	high income
$Q_0 - Q_1$	7.07	4.63	0.74	0.41
$\tilde{Q}_0 - \tilde{Q}_1$	0.28	0.23	0.11	0.05

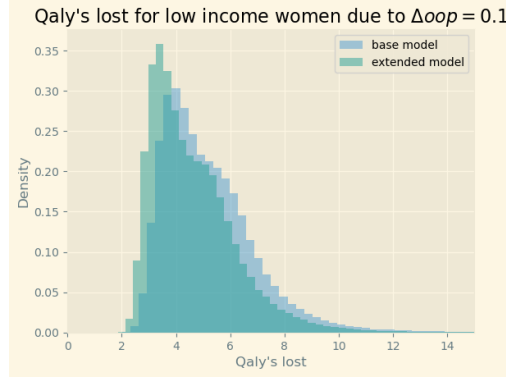


Figure 6: Comparison of the posterior distribution of qaly’s lost for the base and extended models

6.3 Placebo treatment

As a final check on our proposed causal mechanism, we add a placebo treatment. Instead of analyzing mortality linked to health, we analyze mortality due to traffic accidents. In the latter case there is neither the dynamic element of a health stock developing over time nor a role for out-of-pocket expenditure causing forgone treatments which may cause mortality. The details of the model are in the appendix.

The idea is that if out-of-pocket expenditure has a clear positive effect on traffic mortality, an alternative causal mechanism presents itself: the government’s fiscal space has narrowed (e.g. due to a recession) which simultaneously causes the government to increase demand-side cost-sharing and reduce investments in healthcare. The reduction in these investments, reduces healthcare quality and this is the cause of increased mortality. An increase in traffic accidents is then caused by reduced investment in road safety.

We do not find such a clear relation between our *oop* variable and traffic mortality. In particular, we find a posterior mean for the effect of *oop* on traffic mortality of 0.35 with a (posterior) standard deviation of 0.60 which is almost twice as high. This suggests that the path from an increase in *oop* to forgone treatments because they

are too expensive (for low incomes) to reduced health status and higher mortality is plausible.

6.4 Different qaly values

When evaluating the effect of an increase in oop, we use $\theta = 0.6$ in the qaly calculation –see equation (4). Table 6 illustrates that our main results are not sensitive to this choice. The table reports qaly’s lost (due to $\Delta oop = 0.1$) for $\theta = 0.3$ and 0.8. The middle rows equal Table 3 for ease of reference.

For each value of θ we see that the effects are far bigger for low than for high incomes and the dynamic estimates are order of magnitude bigger than the static ones. Further, part of the effect on qaly’s lost is –next to higher mortality– that more people are in the low health state compared to the baseline situation with lower oop. When higher weight θ is given to the low health state, the effect of Δoop is lower: the qaly penalty for being in the low state is reduced. Indeed, all entries in the table fall with θ .

Although the size of the effect varies, the main results are robust to the choice of θ .

Table 6: Dynamic vs. static estimates of qaly’s lost due to $\Delta oop = 0.1$ for different values of θ .

		male	female	male	female
θ	estimate	low income	low income	high income	high income
0.3	$Q_0 - Q_1$	9.56	6.26	0.98	0.56
	$\tilde{Q}_0 - \tilde{Q}_1$	0.48	0.39	0.18	0.09
0.6	$Q_0 - Q_1$	8.01	5.23	0.82	0.46
	$\tilde{Q}_0 - \tilde{Q}_1$	0.31	0.24	0.12	0.06
0.8	$Q_0 - Q_1$	6.97	4.54	0.72	0.39
	$\tilde{Q}_0 - \tilde{Q}_1$	0.20	0.16	0.07	0.04

7 DISCUSSION AND POLICY IMPLICATIONS

This paper develops and estimates a dynamic structural model to assess the long-term effects of health insurance generosity on population health. By modeling transitions between health states and mortality for cohorts defined by age, gender, in-

come and region, we are able to simulate the cumulative impact of demand-side cost-sharing on health outcomes mortality and quality-adjusted life years (qaly's).

Our results yield three key insights. First, increases in the fraction of out-of-pocket spending reduce population health, with particularly large effects for people on low-income. Second, the magnitude of the long-term health cost is substantially greater than what would be inferred from short-run, static estimates such as difference-in-differences. For low-incomes dynamic effects are up to twenty times larger than their static counterparts. Third, widely used methods for causal identification may deliver internally valid but incomplete assessments of health policy, particularly by ignoring the dynamics of health stocks.

These findings have broader relevance beyond the context of healthcare. They underscore the difficulties of evaluating policies whose outcomes develop gradually over time –such as education reforms influencing human capital, climate policies impacting environmental degradation and innovation policies shaping the accumulation of knowledge. Our framework offers a template for incorporating such dynamic effects into empirical work in these fields.

Future research can build on this work in several directions. Incorporating individual-level microdata could help to refine the heterogeneity estimates and to explore alternative functional forms for treatment response in the transition probabilities. Further, gathering more data on healthcare quality and/or explicitly modeling the government's policy on demand-side cost-sharing in this framework can reduce confounder bias.

8 BIBLIOGRAPHY

- Borgschulte, Mark, and Jacob Vogler. 2020. “Did the Aca Medicaid Expansion Save Lives?” *Journal of Health Economics* 72 (July):102333. <https://doi.org/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): 1261–1318. <https://doi.org/10.1093/qje/qjx013>.
- Chandra, Amitabh, Evan Flack, and Ziad Obermeyer. 2021. “The Health Costs of Cost-Sharing,” February. <https://doi.org/10.3386/w28439>.
- Finkelstein, Amy, and Robin McKnight. 2008. “What Did Medicare Do? the Initial Impact of Medicare on Mortality and out of Pocket Medical Spending.” *Journal*

- of Public Economics* 92 (7): 1644–68. <https://doi.org/10.1016/j.jpubeco.2007.10.005>.
- 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): 1–49. <https://doi.org/10.1093/qje/qjaa029>.
- Gross, Tal, Timothy Layton, and Daniel Prinz. 2020. “The Liquidity Sensitivity of Healthcare Consumption: Evidence from Social Security Payments,” October. <https://doi.org/10.3386/w27977>.
- 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. <https://doi.org/10.1093/qje/qjab004>.
- Nyman, J.A. 2003. *The Theory of Demand for Health Insurance*. Stanford University Press.

A DATA

All our variables come from Eurostat. Table 7 shows the dimensions over which our variables vary: country, NUTS 2, calendar year, age and sex. We also present the DOI and 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.⁸

Table 7: Variables and the dimensions over which they vary.

variable	country	NUTS 2	time	age	sex	reference
population		x	x	x	x	https://doi.org/10.2908/DEM
deaths		x	x	x	x	https://doi.org/10.2908/DEM
deprivation		x	x			https://doi.org/10.2908/TGSC
too expensive		x	x			https://doi.org/10.2908/HLTH
unmet		x	x			https://doi.org/10.2908/HLTH
infant mortality		x	x			https://doi.org/10.2908/DEM
number of physicians		x	x			https://doi.org/10.2908/HLTH
out-of-pocket	x		x			https://doi.org/10.2908/HLTH
traffic mortality		x	x			https://doi.org/10.2908/TRAI
GDP per capita		x	x			https://doi.org/10.2908/NAM
unemployment		x	x			https://doi.org/10.2908/LFST
tertiary education		x	x			https://doi.org/10.2908/TGSC
devices for medical imaging	x		x			https://doi.org/10.2908/HLTH

The variables on 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: “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

⁸This file can be found in the github repository: <https://github.com/janboone/dynamic-effects-of-health-insurance-reform>.

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.

Infant mortality measures the number of deaths of infants per 1000 live births in a year at the NUTS 2 region level.

We measure the number of physicians per 100k inhabitants per NUTS 2 region and calendar year. It gives an idea of the resources spent on healthcare in a region. Together with infant mortality and unmet medical needs it gives an idea of healthcare quality in a region.

We characterize how generous a health insurance system is using the variable OOP in our analysis. This variable is derived from data on health care expenditure by financing scheme at the country level. For our OOP measure we focus on household out-of-pocket payment (**out-of-pocket**) as share of total current health expenditure.

In the model extensions we also use GPD per capita and devices for medical imaging as signals of healthcare quality. Unemployment and education capture regional deviations from average health status across age and gender.

Finally, we use traffic mortality in our placebo treatment where changes in out-of-pocket payments are not expected to change mortality.

B ESTIMATION

Table 8 shows the details of the parameters in the model that do not have too many indices. To illustrate we do not present the results for `logodds_age_π` as this would entail 40 rows in the table.

The value of `r_hat` is equal or close to one for all variables. This suggests that the NUTS algorithm converged in drawing samples from the posterior distribution. The probability of falling ill is clearly lower for women than for men. The increase in the fraction of people forgoing treatment because it is too expensive, ζ , is higher for low incomes than for high incomes. The ν parameter of the Beta distribution is sometimes referred to as sample size. For the distribution of the fraction of people reporting unmet medical needs due to reasons other than the treatment being too expensive, ν_{U_o} is above 100. For the fraction of people reporting unmet medical needs (for any reason) ν_U is close to 300. The variable ε captures the

extent to which unmet medical needs for other reasons than too expensive originates with low incomes compared to high incomes. Although the posterior expectation of ε is positive, zero is part of the 94% credibility interval. The uncertainty of this parameter, including that ε equals zero, is propagated through our results and simulated outcomes. Finally, λ_1 equals the (negative) effect of healthcare quality as measured by infant mortality on the log odds of λ , the probability of recovery conditional on being treated. When measuring quality with the number of physicians per 100k inhabitants, λ_2 captures the effect on the log odds of λ .

$$\log(\lambda/(1-\lambda)) = \lambda_0 - \lambda_1 M_{infant} + \lambda_2 N_{physician}$$

where λ_0 varies by year to capture technical progress, M_{infant} measures infant mortality and $N_{physician}$ the number of physicians. The table shows that λ_1 is positive in expectation but can be zero. For λ_2 there is clearer evidence that the effect of physicians on recovery is positive.

Table 8: Estimated parameters of the main model

	mean	sd	hdi_3%	hdi_97%	ess_bulk	r_hat
logodds_sex_pi[F]	-0.19	0.03	-0.25	-0.13	349.00	1.01
logodds_sex_pi[M]	0.50	0.03	0.44	0.56	354.00	1.01
$\zeta[y_l]$	0.27	0.01	0.25	0.30	8429.00	1.00
$\zeta[y_h]$	0.03	0.00	0.03	0.04	6048.00	1.00
nu_U	299.81	18.94	264.86	336.40	7950.00	1.00
nu_U_o	109.35	6.59	97.78	122.34	5584.00	1.00
ε	0.07	0.07	0.00	0.19	10331.00	1.00
lambda.1	0.01	0.01	0.00	0.02	6389.00	1.00
lambda.2	0.06	0.01	0.04	0.07	1704.00	1.01

Figure 7 shows the traceplots for the main parameters of the model. We check the following three characteristics of the plots. First, the plot is 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 is 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 one after the other. Finally, the four chains cover the same regions. All three features are satisfied for the coefficients in the figure.



Figure 7: Traceplot for the main model