

**Cutbacks on health prevention policies:
penny wise, pound foolish?**
Evidence from a health production function

ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Department of Economics

MSc Economics & Business

Master Specialization Policy Economics

Supervisor: Dr. Pilar García Gómez

Name: Francesc López Seguí

Student number: 383416

E-mail: francesc.lopez (at) cmail.cat

Date: August 22, 2014

Abstract

The recent economic slowdown affecting the developed countries has motivated a significant decrease in the governmental prevention and public health expenses, items which already accounted for a very residual part of the health budgets. The main aim of this Master's Dissertation is to provide evidence in favor of the role of prevention in public health: using the National Child Development Study's longitudinal dataset, we estimate the effect of a range of variables related to prevention on health outcomes in a health production function. Grounded on Grossman's (1972) theoretical approach, the specification highlights smoking and obesity as the most important areas for policymaking enhancement, results which are in line with current concerns (Mokdad et al., 2004; WHO, 2014) and existing literature (Fielding, 1985; Swallen et al., 2005). Breastfeeding, a less conventional policy, is also shown to be strongly related to general health outcomes. In sum, no strong evidence is provided to unconditionally defend preventive policies.

I wish to express thanks to Dr. Pilar García Gómez (EUR) for his encouragement and support as well as for his extensive comments and suggestions. Also appreciated is the assistance rendered by Dr. Guillem López Casasnovas (UPF) and the UK Data Service. Finally, thanks are due to Júlia López Seguí: this work strongly benefitted from her spellchecking. All remaining errors are my own.

Governments, under pressure to protect funding for acute care, are cutting other expenditures such as public health and prevention programs. In 2010, on average across EU countries, only 3% of health budget was allocated to areas such as immunization, smoking, alcohol drinking, nutrition and physical activity.

The report notes that spending on prevention now can be much more cost-effective than treating diseases in the future.

Health at a Glance, OECD 2012

Contents

1. Introduction.....	5
2. Background	8
2.1 A definition for prevention	8
2.2 What do we know about the efficiency of prevention?.....	9
2.3 The role of Health Production Functions	11
2.4 Summary of main evidence available.....	14
3. Data and descriptive statistics.....	15
3.1 The sample	15
3.2 The variables	15
3.3 Empirical model	20
4. Results	22
5. Conclusions.....	30
6. References.....	31

1. Introduction

One of the major articles suggesting the potential of prevention policies was published in 2010 by the 2000 Nobel Prize winner James Heckman. It calculates the rate of return of the randomized control trial *Perry PreSchool Program*, a high-quality project targeted to children aged 3 and 4 from disadvantaged backgrounds which studied the effect of the program on crime rates, income and educational achievement, proposing estimates ranging between 7% and 10%. The publication is an emblem of the effectiveness of early-life policies to avert ulterior academic failure among others. On the author's words:

The real question is how to use the available funds wisely.

The best evidence supports the policy prescription: invest in the very young.

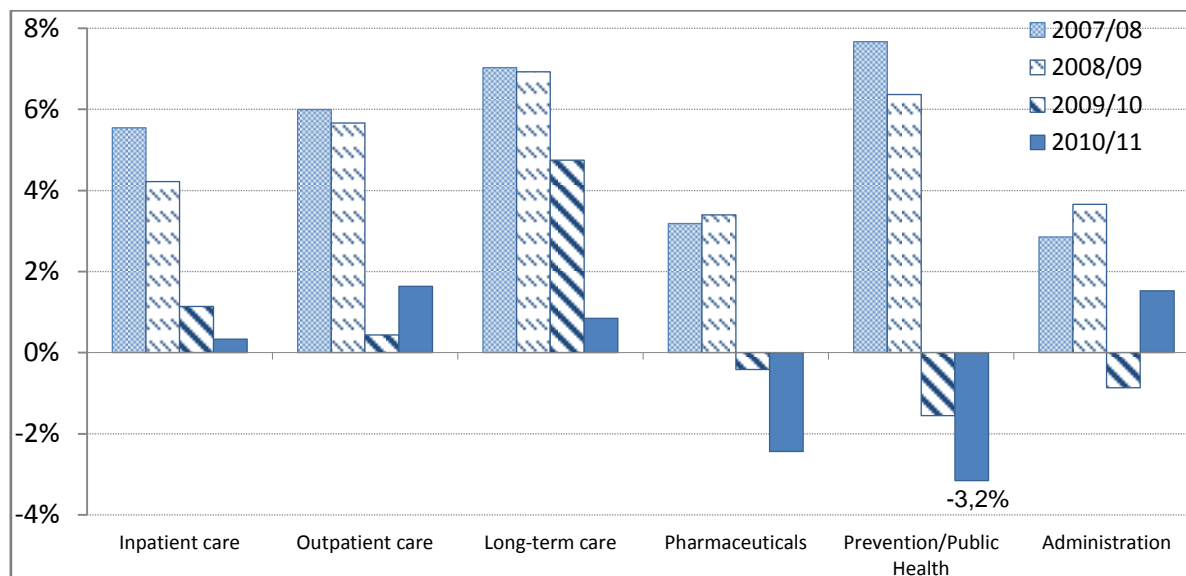
Should this preventive approach be extended to the field of public health policies? Smoking, the misuse of alcohol, a poor diet and physical inactivity are responsible for nearly one million deaths annually in the US, almost 40% of total mortality (Mokdad et al., 2004). In their paper, the authors used these results, together with rising health care costs and aging population, to "persuasively" argue the necessity to set up a more preventive orientation in the American health care system.

This sort of insights contrast with the current expectations placed on preventive measures, which play a residual role in the developed countries' directives. The world economic crisis starting 2008 implied a strong deterioration of the developed countries' governmental finances (Kurt, 2012) which further constrained the scope of the public sector policymaking. Among its spending, health care policies consume around 14% of the national budget, a slightly lower (higher) fraction for low (high) income countries (WHO, 2009), and is predicted to be increasingly important (Breyer et al., 2006).

Every country employs differently this portion of the budget. Nevertheless, a common trend can be identified in the context of the OECD countries: Figure 1 shows the changes in the distribution of health expenditures during the first four years of the economic slowdown, illustrating a reduction on prevention and

public health expenses¹. This last item already represents a marginal part of the overall health expenditure (3% on average in the EU-27 countries).

Figure 1: Average growth by main function of the public expenditure on health, OECD countries, 2008-2011



Source: OECD Health Data 2013

Suspecting that this type of policy might be underutilized, the main aim of this Master's dissertation is to provide evidence in favor of the role of prevention in public health. We have tried to do so by studying the effect of a set of variables related to prevention on health outcomes in a health production function, grounded on Grossman's theoretical benchmark (1972) while using a longitudinal dataset. This is not a common approach as existing literature normally employs cost-benefit analyses, which do not provide strong evidence in favor of preventive policies.

The intertemporal health production function approach is shown to be relatively good at modeling overall health outcomes and highlights smoking and obesity as the two most important areas for policymaking enhancement, results which are in line with current concerns (WHO, 2014) and existing literature (Fielding, 1985; Swallen et al., 2005). Breastfeeding, a less conventional policy, is also

¹ Note that this graph is only motivational. For example, the per capita (not available) figure would have been much more representative. Additionally, regulatory policies (i.e. the "sugar tax") might play an important role when complementing preventive health policymaking and explain why so few resources are devoted to this budget item.

shown to be strongly related to general health outcomes. In sum, in line with the rest of the literature, no strong evidence is provided to unconditionally defend preventive policies. Additionally, the control variable gender is systematically associated with specific diseases (positively associated with cancer, due to a high prevalence of breast cancer among women at the observed age, and negatively related to diabetes and high blood pressure, pathologies that typically show a higher prevalence in men), results which are in line with existing evidence.

The rest of the paper is organized as follows. Section 2 provides some background of the existing evidence on the effectiveness of prevention policies and the health production function approach. Section 3 describes the British National Child Development Study sample, the selected variables and the empirical model. Section 4 comments and discusses on the results of the models and section 5 closes the dissertation with a conclusion and the proposed policy implications.

2. Background

2.1 A definition for prevention

This section aims to review the existing evidence of the efficiency of preventive policies, trying to answer the following question: is the rate of return to investment of a prevention policy higher than treatment expenses? In other words, is *prevention cheaper than the cure*? A preliminary definition might refine what, from now onwards, should be understood as prevention policies.

A New Perspective on the Health of Canadians (also known as *The Lalonde Report*) was published in 1974 and became a groundbreaking paper in the field of public health. After his epidemiological study of Canadians' death factors, the at that time Canadian Minister of National Health and Welfare Mark Lalonde proposed a theoretical model of four broad elements affecting health: human biology, environment, lifestyle and health care organization. The report emphasized the need to look beyond the traditional idea of public health by acknowledging the role of a wide set of variables affecting it.

In this extensive scope of health determinants, multiple policies could be considered prevention. Screening and vaccinations programs seem clear examples -but there could also be less conventional policies such as subsidies to fruit. The concept should definitely be delimited.

The WHO Health Promotion Glossary (1998) defines *disease prevention* as the “measures not only to prevent the occurrence of disease, such as risk factor reduction, but also to arrest its progress and reduce its consequences once established”. The lexicon further differentiates it: “primary prevention is directed towards preventing the initial occurrence of a disorder; secondary and tertiary prevention seek to arrest or retard existing disease and its effects through early detection and appropriate treatment”.

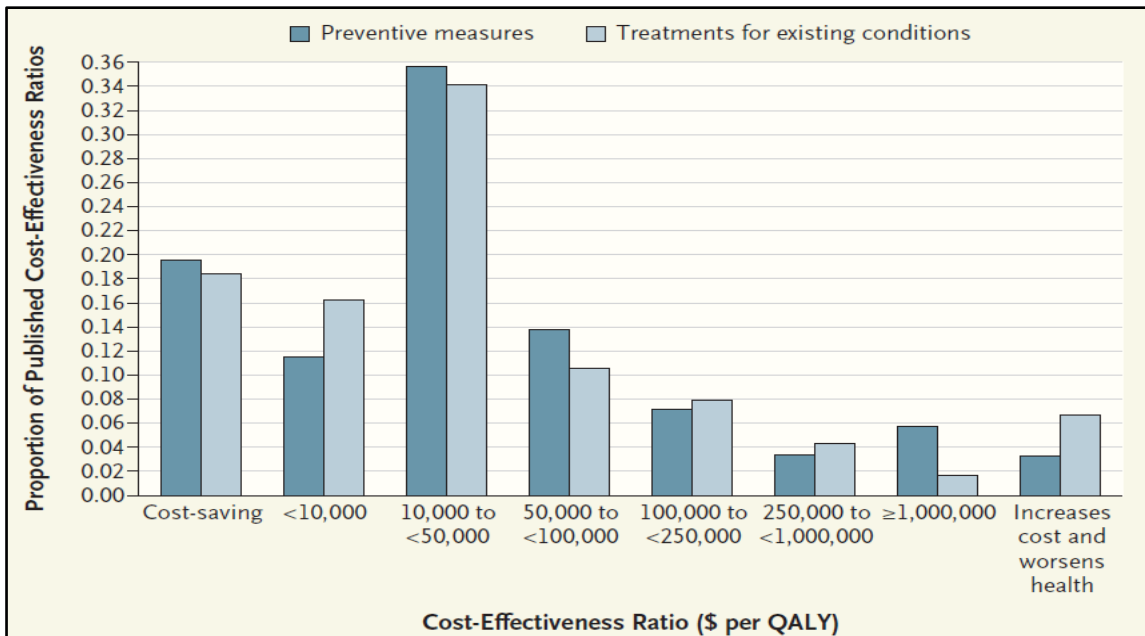
Two ideas can be drawn from this benchmark description. Firstly, we can note that some policies will match with multiple types of prevention: for example, weight control programs are meant to decrease the prevalence of diabetes but also to reduce its adverse effects once the patient already suffers the illness.

Secondly, it is likely to be the case that each of the papers studying the cost-effectiveness of prevention uses a different definition; for example, Cohen et al. (2008) consider the previously defined secondary and tertiary prevention as *treatment* in order to compare the policies' effectiveness².

2.2 What do we know about the efficiency of prevention?

Cohen et al. (2008) provide an intensive review of the evidence of the most relevant articles published in the period 2000-2005. Among 599 articles, the authors select 279 out of 1500 ratios as “preventive policies” following the WHO definition of primary prevention (interventions that try to avert disease) and then compare them to the remaining 1221 “treatment” ratios (policies performed once the individual is already ill).

Figure 2: Distribution of cost-effectiveness ratios for preventive measures and treatments for existing conditions



Source: Cohen et al. (2008)

Data are from the Tufts–New England Medical Center Cost-Effectiveness Registry

Figure 2 shows their main results. The bar graph classifies the 1500 interventions according to their effectiveness separating treatment and

² We will generically call “acute care” or “treatment” anything which is not prevention even when aware that this definition is not medically rigorous.

prevention policies. For example, the first two bars (on the left) show that approximately 19% and 18% of preventive and treatment policies respectively are cost saving. Instead, the two last bars (on the right) show that approximately 3% and 5% of prevention and treatment policies respectively are “nonsense” (increase cost and worsen health). Three main conclusions can be drawn from the results.

Firstly, it cannot be said that prevention is always cheaper than a treatment. In fact, even with this clear bar plot it is impossible to assess it: on the one hand, more preventive measures are shown to be cost-saving; on the other hand, treatment policies are more numerous in the subset of low cost-effectiveness policies. In the author’s words “sweeping statements about the cost-saving potential of prevention [...] are overreaching. Studies have concluded that preventing illness can in some cases save money but in other cases can add to health care costs”. In the one hand, the article identifies certain valuable policies as advising adults to quit smoking, screening for colorectal cancer and providing influenza vaccines. In the other hand, the effectiveness of other policies are less straightforward: drugs that treat high cholesterol are only useful when the targeted population is at high risk of coronary heart sickness, and the usefulness of cancer screening greatly depends on the regularity of the screening and the degree of cancer risk in the cohort.

Secondly, prevention is not always a good idea. According to the traditional \$50,000 threshold used in cost-benefit analyses³, approximately 40% of the policies captured by the article are not worth the money.

Thirdly and lastly, there are a relatively big proportion of policies (both prevention and treatment) that are cost-saving or that show a very low cost-effectiveness (cost relatively few money for each QALY achieved). These are the policies that policymakers should identify and execute firstly.

The previous results are in line with Hackl et al. (2012). These authors calculate the effectiveness of a general health screening program (a bundle of prevention products), finding cost saving potential for a subsample of young participants.

³ £30,000 per QALY is a commonly accepted decision rule in the UK (National Institute for Clinical Excellence, NICE)

Both articles suggest that policies should be well-targeted, which is the same idea that can be drawn from Russell (2007), which states that “it is impossible to generalize about prevention interventions [...], (they) are good investments when used selectively -targeted at those people who benefit most from them- but not such good investments when used for more broadly defined groups”. This will be one of the lessons of this first section.

2.3 The role of Health Production Functions

A Health Production Function (HPF henceforth) will be the main tool used in the empirical part to capture evidence for or against the usage of health prevention policies. The benchmark in the arena is Grossman (1972), who developed a theoretical model proposing health as a fundamental commodity (in the sense that we can assume it to be an argument of the utility function without any ethical judgment) specified as

$$H = f(X) \tag{1}$$

where H stands for the individual health output and X is a vector of inputs of the production function, which he suggested to include anything like nutrient intake, income, consumption of public goods, time devoted to health care, education, genetics and personal and community endowments (such as the environment).

Grossman (1972) conceives the individual as both a consumer and a producer of health, which is understood as a stock which depreciates over time, as if it was a sort of capital. The model considers health as both a consumption good (yields direct satisfaction) and an investment good (yielding satisfaction indirectly through increased productivity).

This individual needs to balance the (costly) resources devoted to health investment against other alternatives, trading-off the marginal costs (defined in the model as the depreciation rate plus the interest rate) and benefits⁴ (the rate

⁴ Note that health care is usually not acquired directly from providers and out-of-pocket expenses are lower than market prices. Our individual balances the marginal costs and benefits of this *out of pocket* expenses such that there is a gap between the “effective” and “real” demand, named “ex post moral hazard”.

of return of this good in both market and non-market sectors) hence determining an optimal level of health demand. For example, the depreciation rate increases with age becoming more and more costly to attain a certain health level. In turn, age reduces the marginal benefit of the investment. Based on these insights, Grossman later presented a model predicting the effect of variations in prices of related goods on macro magnitudes as employment, wages and productivity.

Health production functions with aggregate data

The first found evidence applying Grossman's theoretical approach was developed by Auster et al. (1969), analyzing the USA states with the following Cobb-Douglas HPF

$$H_i = CZ_i^\alpha X_i^\beta M_i^\gamma e^{\delta D_i} e^{u_i} \quad (2)$$

where H_i is the mortality rate corrected for the demographic composition in state i ; Z , X , M and D are vector of economic, consumption, medical and organization inputs respectively; and u_i is the error term. The equation is linearized by applying logs and estimated by 2SLS and OLS. The results of their pioneer specification (not shown) indicate very few significant variables (in the preferred specification, only four out of twelve) though already signals the importance of environmental variables as unfavorable nutrition, lack of exercise and psychological tensions when compared to medical care.

Cochrane et al. (1978) estimated a similar model with aggregate data for the OCDE countries. Using extensive information on many dietary, economic, demographic and health factors, they explore the effect of a main input of interest, namely the health services in these countries (doctors, nurses, beds, pediatricians, etc.) across different age groups. None of these service factors are consistently related to mortality but the article remarks some interesting (though non-significant) relations. Firstly, income is suggested to be negatively related to mortality independently of the observed age. Secondly, certain moments of life (the perinatal period, for example) are shown to be more susceptible to be affected by the proposed variables.

The two previous surveyed articles share two important characteristics: firstly, they lead to similar conclusions, defending the importance of environmental and lifestyle variables vis-à-vis medical care. Secondly, they both use aggregate data. We will therefore complement this review with micro-based HPF studies.

Health production functions with individual data

Contemporary to Grossman's seminal paper, Belloc and Breslow (1972) studied the relation of physical health status and health practices based on a sample of adult residents in the Alameda County, California. In their paper, "good practices" as regularity of meals, physical activity, absence of smoking and drinking are shown to be associated with good health. Ulterior papers have updated the data, interest variables and methods, adding other remarkable insights. In a well-known article called "Should you eat breakfast? Estimates from health production functions", Kenkel (1995) provocatively suggests that you should not do so.

Although the aim of the formerly mentioned papers is similar to our purpose, they differ in one ingredient which is the intertemporal approach. All the previous present a static methodology, where the input and output variables are coeval. The first breaking evidence is by the Cebu Study Team (1991). In this paper, a team of researchers from the USA and the Philippines studied the effect of a range of variables on diarrhea, febrile respiratory infection and weight of a sample of 3,000 mother-infant pairs in the Filipino city of Cebu on a very limited time scope (five years). The results show a significant relation between household factors and health outcomes.

We can conclude that the HPF approach has been extensively used to measure the relation between health inputs and outputs. More precisely, the literature signals the environmental and lifestyle factors as important ingredients to estimate an individual's health status. Conversely, we have found no models specifically interested in the effect of prevention policies and a single one exploiting the intertemporal approach using longitudinal data. Hence, we consider that there is not any benchmark available of this specific method ever applied on our topic.

2.4 Summary of main evidence available

The presented evidence of the effectiveness of preventive compared to treatment policies, strongly based on the meta-analysis by Cohen et al. (2008), suggests that the academy has indeed been interested in measuring the consequences of both policy alternatives. The literature concludes that policymakers should pursue options that move toward greater use of the proven prevention instruments and for certain targeted individuals as prevention is not *always* “cheaper” than cure.

Despite the significant problems intrinsic in cost-benefit analyses (besides the discussion of what is (not) prevention, scholars recognize the variety of items included in cost-benefit analyses, hurting their comparability, see Vilma et al., 2000), these have been the main discussion tools up to now and HPFs have not played any role in the debate yet. Additionally, we have found that HPF and longitudinal data are not commonly combined. Both facts enrich the exploratory perspective of the rest of this dissertation.

3. Data and descriptive statistics

3.1 The sample

The National Child Development Study (NCDS) is one of the most world renowned longitudinal cohort studies. The project targeted all the 18,558 individuals born in Great Britain in a single week of 1958 to collect broad information on physical and educational development, economic circumstances, employment, health and social participation among others along their lives. Since then there have been ten follow-ups, the latest being in 2013 (at age 55 for the cohort members). The data is available through the Centre for Longitudinal Studies (CLS).

This paper focuses on the effect of a set of variables observed at ages 0-16 on 7 health outcomes observed in the last follow-up available (the ninth wave, at age 50). Due to attrition, defined as the pattern of loss of individual records over time, which might be partly caused by deaths, the individuals studied are progressively reduced throughout the years such that our final selected sample consists of nearly half of the initial sample (9,790 citizens). We retake this potential problem in the discussion closing this section.

3.2 The variables

Outcome variables

We will indistinctively call outcome, dependent or interest variables the following group of 7 indicators that capture different health dimensions. Figure 3 provides a graphical description.

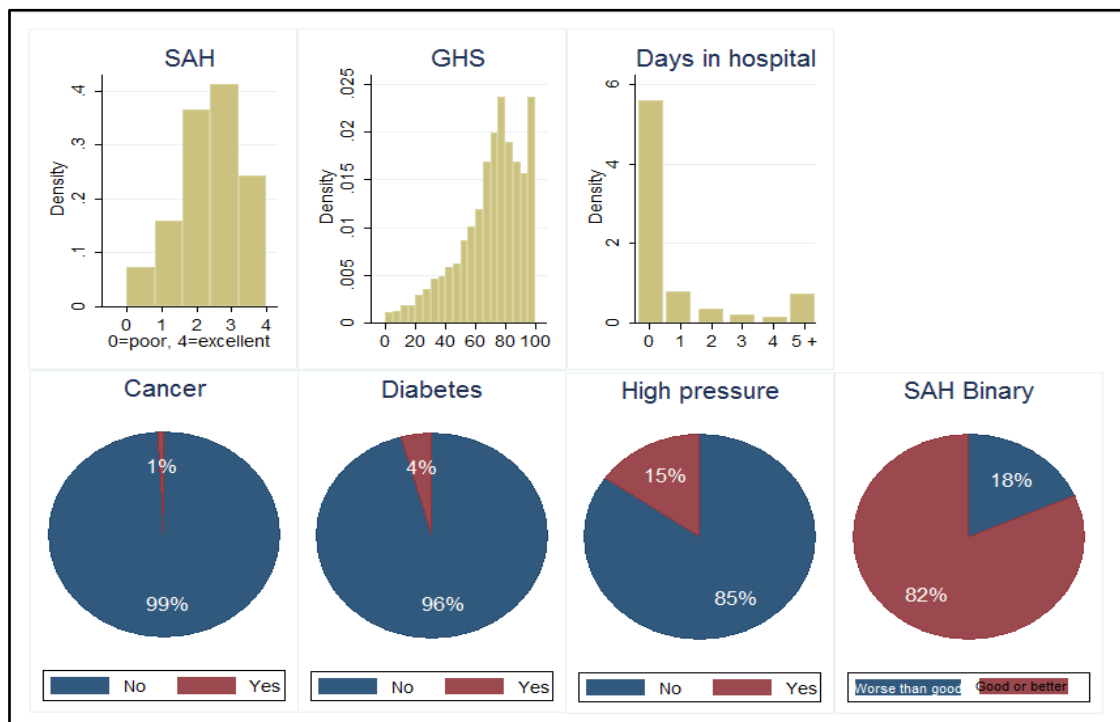
The variable “Self-Assessed Health” (SAH) is shown to be a good predictor of health deterioration and mortality and thus is widely used in the related literature (Idler and Kasl, 1995; Lundberg, 1996; Idler and Benyamini, 1997; and Crossley, 2002). The SAH results range 0-4 where 0 stands for “Poor” while 4 means “Excellent” and is shown to be skewed to the right, meaning that, in general, our 9,790 sample individuals enjoy a good health status.

“SAH binary” takes the value of 1 if the individual reports his health status to be at least “Good” and 0 otherwise (note that this is only a simplification of the previous SAH variable). Complementary, the overall well-being level is measured by the synthetic measure “General Health Status” (GHS): in line with Goldberg’s General Health Score (1978), it contains information on the overall health condition and ranges from 0 to 100. This variable often used in national health surveys and its graph bar is also shown to be asymmetric to the right.

To continue, the following are complementing the three previous non-disease specific variables: measures on cancer, diabetes and cardiovascular health “High Blood Pressure”. We include them motivated by the fact they are in the top ten causes of death, specially affecting high-income countries (WHO, 2013). These binary variables are also outcomes of interest in the health production functions in the previously referred model by Kenkel (1995) and show that the diseases are suffered by 1%, 4% and 15% of the sample respectively.

Finally, “Days in hospital” is a count variable indicating the amount of days the individual stayed in hospital since the previous interview (at age 46) and aims to proxy health expenditures. In consonance with the previous SAH and GHS variables, the respective bar plot is skewed to the left.

Figure 3: Summary of dependent variables



The previous outcomes will be explained by the following independent variables, which have been classified depending on whether they are related to the main research question (the “prevention variables”) or other variables that help avoiding the potential omitted variable bias (the “control variables”).

Detailed considerations concerning what does constitute a pure preventive factor are beyond the scope of our dissertation. For example, the birth weight could be seen as a control variable, capturing health status, but also part from a prevention policy program aimed to reinforce pregnancy feeding⁵.

Prevention variables

In our model we use nine “Prevention variables”, the majority of them being dummies, observed at age 16. Table 1 provides some summary statistics.

Firstly, “overweight” indicates whether an individual’s Body Mass Index (BMI) at age 16 is above the threshold set by the WHO⁶ and is shown to affect 7% of the sample. Secondly, smoking and alcohol drinking (affecting 33% and 96% of the sample⁷), are among the most important nongenetic (modifiable) factors causing death in the developed countries (Mokdad et al., 2004). Thirdly, sleeping quality (Kenkel, 1995; Taveras et al., 2008) is added -8% of the sample report having trouble when sleeping. Finally, we observe whether the cohort member was breastfed⁸ (72% of the sample), and whether the number of antenatal visits were above or below the WHO recommended threshold⁹

⁵ For example, Kramer (1998) insists in the *quality* -rather than the *quantity*- of maternal nutrition as the key factor in order to achieve the correct fetal growth rate. In these lines, further feeding monitoring during pregnancy could be defended as part of a prevention program.

⁶ The WHO considers an individual as overweight when their BMI is above 25.

⁷ There is no found consensus on the optimal acceptable of alcohol consumption. Consequently, an individual is considered to “drink” if reported to consume any amount. Hence, only 4% of the sample reports not drinking at all. Results do not vary significantly when changing this threshold.

⁸ Kramer et al. (2008) shows strong evidence that prolonged breastfeeding improves the infant’s cognitive development. A meta-analysis by Harder et al. (2005) concludes that the duration of breastfeeding decreases overweight risk.

⁹ WHO stresses the effectiveness of antenatal care so as to prevent, detect and treat health problems and recommends at least 4 visits, according to its Global Health Observatory http://www.who.int/gho/urban_health/services/antenatal_care_text/en/ (retrieved 18 August 2014). Surprisingly, 70% of our sample did not achieve the recommended threshold, probably because it belongs to a more modern concern.

Additional non-dummy variables are included to capture other health-related aspects: the degree of sport practice at age 16 (2.4 activities per individual per week on average, strongly related to the obesity campaigns), and the amount of vaccines administered and visits to the GP the year when the cohort is 16 (2.2 and 4.4 on average respectively).

Table 1: Summary statistics

Output variables (7)							
Variable				Mean	Std. Dev.	Missings	
Self-Assessed Health				2.50	1.11	57	
General Health Status				68.37	21.98	1003	
SAH binary				0.82	0.38	57	
Days in Hospital				1.92	11.55	6	
Diabetes				0.04	0.20	31	
High pressure				0.15	0.15	31	
Cancer				0.01	0.01	31	
Prevention variables (9)				Control variables (12)			
Variable	Mean	Std. Dev.	Missings	Variable	Mean	Std. Dev.	Missings
Overweight	0.07	0.25	2823	Pregnancy was abnormal	0.26	0.44	517
Number of visits to GP	4.42	7.82	940	Monoparental Family	0.03	0.16	1175
Antenatal visits below threshold	0.70	0.46	644	Financial hardship at age 7	0.05	0.25	2059
Sleeping difficulty	0.08	0.27	2532	Financial hardship at age 11	0.08	0.29	1672
Smoking	0.33	0.47	2367	Financial hardship at age 16	0.06	0.28	2677
Drinking	0.96	0.22	2350	Birth weight below 2.5kg threshold	0.04	0.22	511
Breastfed during childhood	0.72	0.46	1275	Father went to school	0.25	0.43	1490
Number of vaccines taken	2.21	1.17	0	Number of illness	0.83	0.95	1547
Num. of weekly sport activities	2.40	1.62	0	Number of times in hospital	0.41	0.65	1695
(Dummy for vaccines missing)	0.17	0.49	3963*	Number of syndromes	7.90	8.42	1099
(Dummy for sports missing)	0.16	0.49	3935*	Female	0.51	0.49	1
*Number of missings previous to the creation of the dummies				Socioeconomic status: blue-collar	0.66	0.47	1466

Note: Number of observations is constant (9,790).

Control variables

The previous prevention measures are complemented with twelve controls, which might be policy relevant but are not the main variables in our research question, their aim being to capture the effect of important factors likely to affect health and be correlated with the preventive variables but without a straightforward scope for policymaking.

Social conditions during youth take into account whether the cohort member is grown in a monoparental family (“Monoparental”, affecting 3% of our sample), the parental level of education (approximated by the binary “Father went to

school”, affecting just 25% of the sample¹⁰) and socioeconomic group (approximated by “Father is blue-collar”, 66% of sample’s parents).

Along the lines of the literature on health inputs, who defend the effect of childhood conditions and family context on health status in later life (Currie and Stabile, 2003, Case et al., 2005; Lindeboom et al., 2009; Trannoy et al., 2010), three dummies called “Financial Hardship” stand for whether the parents reported having economic difficulties the year when the cohort member was 7, 11 and 16. Table 1 shows that this dummy equals to one for 5%, 8% and 6% of the sample respectively. “Birth Weight Below Threshold”, a measure for the health status at birth, estimates whether the cohort member’s mass did not exceed 2.5 kg when born (affecting 4% of the sample)¹¹.

Additionally, we include the following indicators. On the one hand, individuals’ gender (“Female”, nearly half of the sample) and the quality of pregnancy (approximated by “Abnormal Pregnancy”, affecting close to one quarter of the cases); on the other hand, the number of illness (nearly one per individual) and syndromes (roughly eight per individual). Finally, the amount of times been in hospital since previous interview (around 0.4 per individual on average) aims to capture the effect of possible chronic conditions.

Additional remarks

Two other observations are worth a mention. Firstly, none of the variables presented above has more than 30% of missing values except for two cases (sport practice and the number of vaccines, which are above 40% each). Complementary dummies were created to capture the effect of this significant concentration of absent values.

Secondly, we might suspect that some of our variables are significantly correlated among each other’s and thus we might suffer from collinearity. The clearest case is the three dummies capturing the effect of budgetary problems.

¹⁰ This result should not be surprising when taking into account that our sample individuals’ progenitors could be born around 1925.

¹¹ Numerous articles signal birth weight as an important determinant of later-life conditions (Gluckman et al., 2008). The 2.5 kg threshold was adopted in 1950 by the WHO.

Nevertheless, as shown in the correlation matrix presented in Table 2, they do not seem to be strongly related.

Table 2: Correlation matrix of financial hardship variables

	Hardship at age 7	Hardship at age 11	Hardship at age 16
Hardship at age 7	1	0.26	0.24
Hardship at age 11	0.26	1	0.30
Hardship at age 16	0.24	0.30	1

It could have also been possible with the three variables measuring the number of syndromes, illness and times been in hospital since last interview. The following table refutes this possibility.

Table 3: Correlation matrix of medical variables

	N° of syndromes	N° of illness	Times been in hospital
Number of syndromes	1	0.03	0.06
Number of illness	0.03	1	0.12
Times been in hospital	0.06	0.12	1

3.3 Empirical model

This paper will be using different Health Production Function (HPF) specifications conditional on the type of outcome variable. Firstly, “SAH” will be estimated via an Ordered Probit (the dependent variable has more than two categories and they can be sorted). Secondly, a Probit regression will be performed to model our four dummies “Cancer”, “Diabetes”, “High Blood Pressure” and “SAH Binary”. Thirdly, a standard-error robust regression will be the benchmark to estimate the “GHS” variable. Finally, a Negative Binomial regression will be used for the count variable “Days Hospital”.

Our reference benchmark is Grossman’s theoretical model (1972)¹² where the previous X vector in equation 1 is now split into P and C standing for the set of nine prevention elements and twelve control variables respectively:

$$Health\ outcome_{io} = f(P_y, C_y) \tag{3}$$

such that, for example, the baseline OLS model would be specified as

$$General\ Health\ Status_{io} = \sum_1^9 P_{iy} \beta + \sum_1^{12} C_{iy} \gamma + \varepsilon_i \tag{4}$$

¹² Recap section 2.3 for further details on Grossman’s HPF theoretical model.

Recall that even though endogeneity seems to be intrinsic in any HPF (as it would be very optimistic to state that any model controls for all relevant factors), this specification may avoid reverse causation by clearly distinguishing the timing between the outcome (subscript *o* stands for *old*) and independent variables (subscript *y* stands for *young*). This is to say that even if it is clear that current lifestyle is a crucial determinant of current health, including it would imply simultaneity across the variables.

Results of our specifications are shown in the following section.

4. Results

Table 4 presents the estimates of the referred seven specifications.

A general overview

To get a general idea of the overall performance of the variables, we will distinguish them depending on whether they have a significance level of at least 10% in most of the specifications: only 3 out of the 9 prevention proposed variables (smoking, being overweight and breastfeeding) have a significant impact on the majority of health outcomes, giving evidence of the low power of the chosen variables to predict them. Again, only 3 out of 12 control variables seem to have explanatory power.

If we take a look at the results of the models one by one, less pessimistic conclusions can be drawn. On the one hand, specifications 1-3 and 7 (the “general health outcomes”) show strong relations as nearly half of the independent variables are significant. On the other hand, the *health-specific* outcome variables, measuring the prevalence of diseases (regressions 4-6), seem to have less predictive power when using the same set of variables.

Moreover, the sample shows a relatively low prevalence for these diseases (1%, 4% and 15% for cancer, diabetes and high blood pressure respectively). Therefore, the focus of interest of further analysis will move to the significant relations found in models 1-3 and 7, excluding models 4-6.

Some preventive activities are related to some health variables but are not to others: for example, drinking is shown to be related to diabetes while might not be that determinant in estimating another disease or a general health measure. Unfortunately, the existing related evidence has shown not deep enough to provide this kind of insights. We can thus state that specific factors seem to be determinant of specific diseases¹³ while environmental and lifestyle variables are more likely to predict general health outcomes (note that outcome variables

¹³ This might be the reason why these illnesses are themselves individually modelled in more specific literature.

Table 4: Models' results with estimated coefficients¹⁴

Variable	(1) SAH		(2) GHS		(3) SAH binary		(4) Diabetes		(5) Cancer		(6) High pressure		(7) Days in hospital	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Overweight	-0.31***	0.07	-6.40***	1.63	-0.33***	0.09	0.74***	0.11			0.37***	0.09	0.48**	0.20
Smoking	-0.28***	0.04	-4.06***	0.84	-0.24***	0.05	-0.08	0.09	0.00	0.16	0.09	0.05	0.23**	0.11
Drinking	0.12	0.09	2.12	1.91	0.10	0.13	-0.39**	0.16	-0.37	0.27	-0.21	0.12	0.31	0.26
Number of vaccines	0.04	0.02	0.84	0.59	0.03	0.04	-0.00	0.06	-0.11	0.09	0.02	0.04	0.04	0.08
Sports	0.01	0.01	0.95**	0.39	0.05**	0.02	0.00	0.04	0.07	0.08	0.02	0.02	-0.05	0.05
Was breastfed	0.14***	0.04	2.03**	0.88	0.17***	0.06	0.04	0.09	0.04	0.17	-0.09	0.06	-0.14	0.12
Number of visits to GP	-0.01***	0.00	-0.16***	0.05	-0.01**	0.00	0.00	0.00	-0.00	0.01	0.00	0.00	0.01*	0.00
Antenatal visits below threshold	-0.00	0.04	-0.57	0.84	-0.02	0.06	-0.11	0.08	0.20	0.18	-0.01	0.06	0.15	0.12
Sleeping difficulty	-0.11	0.07	-3.26**	1.50	-0.22**	0.09	0.05	0.14	0.19	0.23	0.06	0.09	0.03	0.19
Father went to school	0.12***	0.04	2.55***	0.92	0.14**	0.07	-0.06	0.10	-0.15	0.19	-0.10	0.07	-0.62***	0.13
Female	0.03	0.04	1.671*	0.875	-0.03	0.06	-0.22**	0.09	0.61***	0.19	-0.22***	0.06	0.30**	0.12
Father is blue-collar	-0.11**	0.04	-1.41	0.88	-0.02	0.06	0.03	0.09	0.09	0.17	0.11*	0.06	-0.02	0.12
Number of illness	-0.05***	0.02	-0.99**	0.41	-0.03	0.02	0.05	0.04	0.03	0.07	0.04*	0.02	-0.05	0.05
Monoparental family	-0.13	0.65	2.72	11.76	-0.51	0.48			1.44**	0.59	-0.20	0.59	0.43	1.04
Number of times in hospital	0.00	0.03	-0.31	0.60	0.02	0.04	-0.01	0.06	0.16	0.10	-0.01	0.04	0.05	0.08
Number of syndromes	-0.01***	0.00	-0.13**	0.05	-0.01***	0.00	0.01**	0.00	0.00	0.01	0.00	0.00	0.01	0.00
Pregnancy was abnormal	-0.04	0.04	-0.21	0.90	-0.04	0.06	-0.02	0.09	-0.40*	0.21	-0.00	0.06	-0.00	0.12
Financial hardship age 7	-0.09	0.11	-3.21	2.51	-0.22*	0.12	0.18	0.18	0.33	0.29	0.13	0.13	0.51*	0.27
Financial hardship age 11	-0.12	0.08	0.33	1.80	-0.35***	0.10	-0.04	0.16	-0.33	0.37	-0.06	0.11	0.89***	0.22
Financial hardship age 16	-0.10	0.08	-0.31	1.84	-0.04	0.11	-0.06	0.17	-0.19	0.38	0.11	0.11	-0.32	0.23
Birth weight below threshold	0.01	0.10	1.50	1.95	-0.09	0.13	-0.18	0.23			0.14	0.13	0.33	0.27
Vaccines value was missing	0.03	0.07	0.56	1.59	0.01	0.107	0.10	0.16	-0.73**	0.36	0.10	0.11	0.22	0.21
Sport value was missing	-0.05	0.07	1.85	1.58	-0.01	0.106	0.18	0.16	0.41	0.29	0.08	0.11	0.24	0.21
Number of observations	3228		2963		3228		3224		2872		3232		3237	

¹⁴ Note: *** p<0.01, ** p<0.05, * p<0.1. Variables of interest are highlighted in purple while the rest act as controls. Model 1 is estimated via an Ordered Probit. Model 2 uses a standard-error robust regression. A Probit regression is performed in models 3-6 while a Negative Binomial specification is used in 7.

such as GHS and SAH do somehow include the specific disease effect). This is why our model does not show many significant relations in models 4-6.

An interesting exception is the gender effect, which is found significant across all three models, showing that being a female is positively related to having cancer and negatively related to cardiovascular diseases and diabetes.

The corresponding marginal effects¹⁵ (not displayed for models 4-6) show that being female yields a probability decrease of 0.017 of being diabetic; a probability decrease of 0.05 of suffering from high blood pressure and a probability increase of 0.01 of suffering cancer, keeping the rest of the variables constant at their mean level. These results are in line with existing evidence, which shows that for the specific age group (age 50 approximately) women have a high prevalence of breast cancer while diabetes and high blood pressure are pathologies that typically show a higher prevalence in men.

Interpretation of the figures

Firstly, a certain pattern should be noted. As it would be expected, the sign of the coefficient reverses, the majority of the cases, when moving from models 1-3 (“SAH”, “SAH binary” and “GHS”, the health status measures) to model 7 (“Days in Hospital”, the health utilization measure), such that a factor positively affecting your health (namely “Father went to school”) negatively affects the propensity to go to the hospital. Analogously, the amount of syndromes shown when young negatively affect the health outcomes of the future whereas has a positive effect on your propensity to go to the hospital.

Coefficients from Table 4, model 2 do have a direct interpretation (recall that the GHS scores range 0-100). Therefore, for example, being overweight, smoking, having trouble to sleep well and each symptom suffered when young decrease this score by 6.40, 4.06, 3.26 and 0.13 respectively. Instead, parental education increases your score by 2.55 points.

While the sign and significance level keeps representative, the figures for the models 1, 3 and 7 cannot be directly interpreted. Their marginal effects are

¹⁵ These refer to the marginal effects at the mean from now onwards

shown in Table 5: figures in model 1 should be interpreted as the marginal probability induced by the independent variable to report a certain health status (in our case, the chances of the health status being “good”, an arbitrary outcome for the dependent variable). Therefore, the figures say that, for example, being overweight or smoking reduces your chances of being “good” by 4.5 and 3.5 percentage points respectively.

Table 5: Marginal effects

Variable	Marginal effects model 1: SAH (ordered Probit) ¹⁶	Marginal effects model 3: SAH Binary (Probit)	IRR model 7: Days in hospital (Negative Binomial)
Overweight	-0.045***	-0.089***	1.63**
Smoking	-0.035***	-0.060***	1.26**
Drinking	0.016	0.024	1.38
Number of vaccines	0.004	0.007	1.05
Sports	0.001	0.013**	0.95
Was breastfed	0.017***	0.042***	0.87
Number of visits to GP	0.000***	-0.001**	1.01*
Antenatal visits below threshold	-0.000	-0.006	1.17
Sleeping difficulty	-0.013	-0.005**	1.04
Father went to school	0.013***	-0.005**	0.53***
Female	0.003	-0.005	1.35**
Father is blue-collar	-0.012**	-0.005	0.98
Number of illness	-0.006***	-0.151	0.95
Monoparental family	-0.016	-0.152	1.55
Number of times in hospital	0.000	0.006	1.06
Number of syndromes	-0.000***	-0.003***	1.01*
Pregnancy was abnormal	-0.004	-0.009	1.00***
Financial hardship age 7	-0.010	-0.058	1.68
Financial hardship age 11	-0.015	-0.094***	2.44
Financial hardship age 16	-0.012	-0.011	0.73
Birth weight below threshold	0.001	-0.024	1.39

Note: Recall *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Variables of interest are highlighted in purple.

Marginal effects in model 3 are to be interpreted as the probability of moving from outcome 0 (individual’s health status is “worse than good”) onto outcome 1 (individual’s health is “good or better”) keeping all other variables constant at their mean level. For example, being overweight, having financial trouble at ages 7 and 11 and having sleeping difficulties decreases your chances of transiting from the first to the second outcome by 9, 6, 9 and 6 percentage points respectively.

¹⁶ These are the coefficients for the benchmark case of outcome being good (H=3).

Finally, model 7 was estimated by a Negative Binomial specification. In this case, the coefficients presented in Table 4 are semi-elasticities -measure the relative change in the conditional mean of a marginal change in the independent variable, *ceteris paribus*. Nevertheless, in the cases where the independent variable is discrete (all the significant independent variables, in this case), this coefficient does not make sense on itself and the ratio between the conditional expectations of the independent variable with and without the effect of the dummy, all else equal, needs to be calculated instead via the following equation:

$$\frac{E(y_i|x_{ik}=1,x')}{E(y_i|x_{ik}=0,x')} = \exp\beta_k \quad (5)$$

The alternative to this cumbersome calculation is regressing the model again deriving the incidence rate ratios (IRR), which allow for a direct interpretation. The results are also shown in Table 5. Their interpretation come as following: in the case of the variable capturing the gender effect, the IRR is 1.35. This ratio compares females to males, when the other factors are held invariable. Therefore, females are expected to have a rate 1.35 times greater for “Days in Hospital”, compared to males, all the rest equal. In the case of the variable capturing the effect of the parental education, the coefficient shows that “educated” parental background individuals are expected to have a rate of 0.53 for the dependent variable compared to the “not educated” individuals.

The majority of the prevention variables are observed during adolescence (age 16). Three of them seem important because of their size and because they systematically show a significant effect across our preferred models: the BMI measure, breastfeeding and smoking, in comparison with performing sports, the sleeping quality and the number of visits to the GP. Except for breastfeeding, these results are in line with current concerns (Mokdad et al., 2004; WHO, 2014) and existing literature (Fielding, 1985; Swallen et al., 2005). The results bring immediately to the conclusion that policymaking in these fields should be further enhanced. Vaccination policies, drinking and the amount of antenatal visits do not seem to play any role, in contrast to the evidence found in Cohen et al. (2008) and current policymaking concerns (WHO, 2014).

Among the proposed control variables related to the individual's socioeconomic status, parental education seems to be the only important factor affecting health, in contrast with the budgetary constraints and the parental structure. The variables estimating the health status when young suggest that the amount of illness and syndromes do have some predictive power while "neonatal" variables (birth weight and pregnancies' quality) seem not to be related at all to the health outcomes.

Lastly, it should be clarified the reason why Table 4 has some missing values. It is due to multicollinearity: due to the lack of variability, two dependent variables are highly correlated such that one can "predict" the other. In this case, the estimates would be wrong such that the regression automatically omits the correlated variable and the phenomenon does not decrease the reliability of the whole specification. Even when the regression does not automatically drop the affected variable, multicollinearity could also be present in a lower degree in the models if the variability is not high enough.

Summary results

In sum, two conclusions close this part of the paper. Firstly, the model is a better predictor of overall-health outcome variables (SAH, SAH binary and GHS) than specific maladies. These results suggest that estimating this type of diseases involves observing additional specific factors which our estimation does not capture.

Secondly, the results identify three variables which seem to be relevant in terms of preventive policymaking: in the one hand, the habit of smoking and being overweight, all of them are identified as potential future threats to individual's health. These results are in line with the current guidelines (WHO, 2014) and suggest this field of policymaking should be further enhanced. In the other hand, breastfeeding, a less conventional policy, is also shown to be strongly related to general health outcomes

The following subsections aim to comment on the potential threats to the validity of the previously presented results.

Attrition

As mentioned in section 3.1, the quality of the dataset needs to be a main concern. We have to check whether the individuals still observed at age 50 (9,790) are significantly different from the initial sample (18,558) so as to make sure the final cohort is still representative of the whole population. If not, the loss of observations through time will be non-random, related to some of the explanatory variables and thus will limit the interpretation of our results. This is the case: the mean-comparison tests show that 22 out of the 23 independent variables which have been used in the paper and observed during ages 0-16 are significantly different for those individuals (not) followed at later ages.

Another question would be: for a certain outcome variable, are its missing observations significantly different regarding the independent variables? Performed mean-comparison tests show that this is the case for approximately 20% of the variables: for example, 7 out of 23 variables which have been used to predict the prevalence of diabetes are significantly different across individuals depending on whether diabetes is afterwards measured or not. We conclude that the quality of the dataset has been progressively eroded as information on individuals was lost.

Some published papers using the same dataset practically ignore the problem: Rosa Dias (2010), for example, performs a variable addition test to show that the self-assessed health (his main outcome variable) is free from the previous concern and cites previous articles (Case et al., 2005 and Lindeboom et al., 2006) to argue that variables associated with socioeconomic status and unemployment, in this specific dataset, have been proved to be unrelated to attrition. Tubeauf (2012) recognizes the problem too but is bind to ignore it.

The problem could be embedded with a multiple imputation strategy, a general approach to handling missing data as proposed in Goldstein (2009), which particularly refers to the National Child Development Study. On the one hand, the whole data still has a reasonable size (9,790 individuals). Thus it seems that weighting the dataset, even when renouncing to part of the observations, might improve rather than undermine the analysis and improve its external validity. On

the other hand, Table 4 shows that, in practice, the regressions are performed with only around 3,000 observations. There is a clear trade-off between internal and external validity that has to be considered. Additionally, the amount of observations at age 50 could be enlarged if the dataset allowed to distinguish people who died after age 16 such that, in models 1-3, the effect of the variables when young could be estimated with a health status equal to 0.

Endogeneity

Based on our results and additional analysis of the related literature, the most relevant variables should be further sorted, meaning that more efforts should be put in finding relevant variables. For example, Kenkel (1995) already included the stress level as an important factor to explain the health outcomes; additional chronic conditions (as in Cebu Study Team, 1991) and supplementary information on maternal habits during pregnancy (smoking, sugar levels, and excessive weight gain) might be relevant.

Nevertheless, since we work with a good although limited dataset, which focuses on micro data, not everything can be observed. In any case, the underlying drawback in increasing the amount of explanatory variables is the potential effect of collinearity.

Sensitivity analysis

Finally, further analysis of the data could focus on different aspects so as to assess whether the conclusions can be maintained or not: firstly, the effect along age cohorts; secondly, distinguishing the explanatory power of the prevention variables with and without the controls; thirdly, including additional outcome variables which further explore the effect of the specification on measures which proxy, for example, health expenses; and finally, including interaction variables.

5. Conclusions

This dissertation has aimed to throw light on the compared effectiveness of health prevention policies by creating a health production function which relates variables at ages 0-16 to seven health outcomes at age 50.

The existing literature about the relative cost-effectiveness of prevention and treatment investments shows that it cannot be stated that preventive policies are more efficient investments. Evidence has already identified certain very few cost-effective -and even cost-saving- policies of both types which should be, in any case, emphasized. In other words: not only preventive policies need to be enhanced but also other type of measures that have been proved to be very valuable.

In our specifications, the following variables are identified as important factors determining health and with potential for preventive policymaking, namely the smoking addiction and being overweight. This is in line with current concerns (Mokdad et al., 2004; WHO, 2014) and with existing literature (Fielding, 1985; Swallen et al., 2005). Hence, current policymaking seems to run in the right direction.

Additionally, breastfeeding is also systematically found to be strongly related to an individual's health status. This result adds to the debate on the importance of this variable and opens a new scope for more discussion. Finally, the significance of the gender effect across the health-specific variables is an interesting unexpected result that needs to be further explored –maybe altogether with a more medical setup.

The paper has added some evidence to the literature of health production functions -more specifically, with the added value of the longitudinal data approach- and the compared effect of variables subject to preventive policymaking –often related to early-life and environmental and lifestyle policies. Solving the technical drawbacks, further research should deepen in the topic, exploring the robustness of the previous results across ages with a wider set of health outcome variables.

6. References

Auster, R. , Irving L. and Deborah S. (1969) "The production of health: an exploratory study." *Journal of Human Resources* 411-436.

Barker, D. J. (2004) "The developmental origins of adult disease." *Journal of the American College of Nutrition* Dec, 23 (6 suppl.): 588S-595S.

Baumol, W. J. (1968) "On the social rate of discount." *The American Economic Review* 788-802.

Braithwaite, R. S. et al. (2008) "What does the value of modern medicine say about the \$50,000 per quality-adjusted life-year decision rule?" *Journal of Medical care* 46.4: 349-356.

Breyer, F. and Stefan F. (2006) "Life expectancy and health care expenditures: a new calculation for Germany using the costs of dying." *Health Policy* 75.2: 178-186.

Case, A. , Fertig A. and Paxson C. (2005) "The lasting impact of childhood health and circumstance." *Journal of Health Economics* 24.2: 365-389.

Cebu Study Team (1992) "A child health production function estimated from longitudinal data." *Journal of Development Economics* 38.2: 323-351.

Centre for Longitudinal Studies, Institute of Education (2013). "National Child Development Study. General Information.". P³G observatory. <http://www.p3gobservatory.org:8080/catalogue.htm?studyId=108>

Cochrane, A. , Leger, A. S. St. and Moore, F. (1978) "Health service 'input' and mortality 'output' in developed countries." *Journal of Epidemiology and Community Health* 32.3: 200-205.

Cohen J. T. et al (2008) "Does preventive care save money? Health Economics and the presidential candidates." *New England Journal of Medicine* 358: 661-663.

Crossley, T. F. and Kennedy, S. (2002) "The reliability of self-assessed health status." *Journal of Health Economics* 21.4: 643-658.

Fielding, J. E. (1985) "Smoking: health effects and control (1)." *The New England journal of medicine* 313.8: 491.

Forget, E. L. et al. (2008) "Variations in lifetime healthcare costs across a population." *Healthcare Policy* 4.1: e148.

Fuchs, V. R. (1982) "Time preference and health: an exploratory study." *Economic Aspects of Health* University of Chicago Press 93-120.

Gigerenzer, G, et al. (2007) "Helping doctors and patients make sense of health statistics." *Psychological science in the public interest* 8.2: 53-96.

Gigerenzer, G. , Mata, J. and Frank, R. (2009) "Public knowledge of benefits of breast and prostate cancer screening in Europe." *Journal of the National Cancer Institute* 101.17: 1216-1220.

Goldberg, D. P. (1978) "Manual of the general health questionnaire." *National Foundation for Educational Research (NFER)*, 1978.

Goldstein, H. (2009) "Handling attrition and non-response in longitudinal data." *Longitudinal and Life Course Studies* 1.1.

Grossman, M. (1972) "On the concept of health capital and the demand for health." *The Journal of Political Economy* 80.2 (1972): 223.

Grossman, M. (1972) "The demand for health: a theoretical and empirical investigation." *NBER Books*. <http://ideas.repec.org/b/nbr/nberbk/gros72-1.html>

Grossman, M. (2004) "The demand for health, 30 years later: a very personal retrospective and prospective reflection." *Journal of health economics* 23.4: 629-636.

Heckman, J. J. et al. (2010) "The rate of return to the HighScope Perry Preschool Program." *Journal of Public Economics* 94.1 (2010): 114-128.

Idler, E. L. and Kasl S. V. (1995) "Self-ratings of health: do they also predict change in functional ability?" *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 50.6: S344-S353.

Idler, E. L. and Benyamini, Y. (1997) "Self-rated health and mortality: a review of twenty-seven community studies." *Journal of health and social behavior*: 21-37.

Jusot, F. , Or, Z. and Sirven, N. (2012) "Variations in preventive care utilization in Europe." *European Journal of Ageing* 9.1: 15-25.

Kenkel, D. S. (1994) "The demand for preventive medical care." *Applied Economics* 26.4: 313-325.

Kenkel, D. S. (1995) "Should you eat breakfast? Estimates from health production functions." *Health Economics* 4.1: 15-29.

King, J. T. , et al. (2005) "Willingness to pay for a quality-adjusted life year: implications for societal health care resource allocation." *Medical Decision Making* 25.6: 667-677.

Kramer, M. S. (1998) "Maternal nutrition, pregnancy outcome and public health policy." *Canadian Medical Association Journal* 159.6: 663.

Kurt, S. et al. (2012) "The effect of global financial crisis on budget deficits in European countries: panel data analysis." *Istanbul University Econometrics & Statistics e-Journal* 17.1.

Lange, F. (2011) "The role of education in complex health decisions: evidence from cancer screening." *Journal of health economics* 30.1: 43-54.

Lundberg, O. and Manderbacka, K (1996). "Assessing reliability of a measure of self-rated health." *Scandinavian Journal of Public Health* 24.3: 218-224.

Maciosek M. V. et al (2006) "Priorities among effective clinical preventive services: results of a systematic review and analysis." *American Journal of Preventive Medicine* 31:52-61.

Mokdad, A. H. , et al (2004) "Actual causes of death in the United States, 2000." *Journal of American Medical Association* 291.10: 1238-1245.

OECD (2012) "Health at a Glance: Europe 2012" OECD Publishing.
<http://dx.doi.org/10.1787/9789264183896-en>

OECD (2013) "OECD Health statistics 2013" OECD Publishing.
<http://www.oecd.org/health/health-systems/oecdhealthdata.htm>

Pebley, A. R. (1984) "Intervention projects and the study of socioeconomic determinants of mortality." *Population and Development Review* 281-305.

Picone, G. , Sloan, F. and Taylor, D. (2004) "Effects of risk and time preference and expected longevity on demand for medical tests." *Journal of Risk and Uncertainty* 28.1: 39-53.

Robert Wood Johnson Foundation (2013) "Return on Investments in Public Health: Saving lives and money." Policy Highlight Brief.
http://www.rwjf.org/content/dam/farm/reports/issue_briefs/2013/rwjf72446

Rosa Dias, P. (2010) "Modelling opportunity in health under partial observability of circumstances." *Health Economics* 19.3: 252-264.

Russell, L. B. (2009) "Prevention's potential for slowing the growth of medical spending." *Health Affairs* 28.1: 42-45.

Swallen, K. C. et al. (2005) "Overweight, obesity, and health-related quality of life among adolescents: the National Longitudinal Study of Adolescent Health." *Pediatrics* 115.2: 340-347.

Tubeuf, S. et al. (2012) "Mediating role of education and lifestyles in the relationship between early-life conditions and health: evidence from the 1958 British Cohort." *Health economics* 21.S1: 129-150.

Van Den Berg, G. J. et al. (2006) "Economic conditions early in life and individual mortality." *The American Economic Review*: 290-302.

Weinstein, M. C. (2008) "How much are Americans willing to pay for a quality-adjusted life year?" *Medical care* 46.4: 343-345.

World Health Organization (1998) "Health Promotion Glossary". Health education and Health Promotion Unit (HEP).
<http://www.who.int/healthpromotion/about/HPR%20Glossary%201998.pdf?ua=1>

WHO (2009) "World Health Statistics: Health Expenditure" WHO Publishing
http://www.who.int/whosis/whostat/EN_WHS09_Table7.pdf

WHO (2013) "The top 10 causes of death" Fact sheet 310, WHO Publishing
<http://www.who.int/mediacentre/factsheets/fs310/en/>

WHO (2014) "World Health Statistics 2014" WHO Publishing
http://www.who.int/gho/publications/world_health_statistics/2014/en/