Essays on the economics of health

Paul Andrés Rodríguez Lesmes

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Economics
University College London

December 7, 2016

I, Paul Andrés Rodríguez Lesmes, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

This thesis presents three empirical studies related to the economics of health interventions. All of them use data from England and are related to preventive care.

The first study estimates the potential impact of early diagnosis programmes on medication, subjective health and lifestyle. By taking advantage of the survey design of the *English Longitudinal Study of Ageing* (ELSA), a regression discontinuity design based on the blood pressure of the respondents allows for estimates free of selection bias due to screening. There is evidence of a temporal increase in the use of medication as a treatment for the condition, and induced lifestyle changes.

The second study proposes a structural dynamic life-cycle model for studying the economic value of the adoption of medical innovations. It allows for both cost-benefit and cost-effectiveness calculations, by considering long-run gains on productivity and on welfare derived from adjusting savings and labour supply throughout life. In particular, the case of a medication that reduces the odds of developing cardiovascular diseases, namely statins, is considered. Using data from ELSA, it is possible to calculate the value of such treatment, and to consider counterfactual policy scenarios.

The last study proposes an empirical test for determining whether rewarded tasks are cost complements or substitutes in a pay for performance scheme with kinks on linear task-specific reward functions. The test is based on the insensitivity of effort exerted on a particular task to variations in the price of competing tasks for agents who are bunched near the kink. As a case study, we consider the case of the *Quality and Outcomes Framework*, which is a pay for performance scheme for family doctors in the UK. We found no evidence of effort-diversion as a result of the changes introduced in 2011.

Acknowledgements

Firstly, I wish to thank my mother, father, brothers and sisters for their support during my career. I would also like to thank my beloved wife, for enduring all of these years.

I would also like to thank my supervisors, Marcos Vera-Hernández and Áureo De Paula for their valuable contribution to these projects.

I also acknowledge the valuable discussions with James Banks, Gabriella Conti, David Cutler, Mariacristina De Nardi, Eric French, Elaine Kelly, Jennifer Mindell, Raun van Oojin, Svetlana Pashchenko, Matilde Pinto Machado, Suphanit Piyapromdee, Ponpoje Porapakkarm and Luigi Siciliani. Also contributions from attendants and discussants from the Royal Economic Society Symposium of Junior Researchers and the PhD Meetings 2016, Netspar International Pension Workshop 2016, the International PhD Economics Conference in Leicester 2015, the Students' Work in Progress Seminar at UCL 2016, the American-European Health Economics Study Group workshop 2016 at Tor Vergata University, the pre-doctoral NBER-IFS Value of Medical Research network meetings, and faculty seminars at Universidad ICESI, Groningen University and Universidad del Rosario.

Finally, I would like to express my greatest gratitude to Banco de la República, COLFUTURO, ICETEX, and the NBER-IFS Value of Medical Research network, institutions that sponsored my PhD studies. Also Professors David Bardey, Darwin Cortés, Juan Gallego, Luis F. Gamboa, Darío Maldonado, Juan F. Vargas and Juanita Villaveces for their help and encouragement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Authors like Becker (2007) and Murphy and Topel (2010) argue that the health revolution of the 20th century has been among its more important developments, if not the most significant. Not only has life expectancy doubled, but quality of health has also improved sharply. Behind this revolution has been a notorious advance of medical science as well as the introduction of public policies that have allowed for the adoption of crucial innovations. Such improvements have consequences for economic behaviour which are also discussed by Becker (2007). Following Grossman (1972), health can be interpreted as a form of human capital, meaning that individuals invest in preserving it not only due to its commodity value but also because it determines their total time available for either labour or leisure. This motivates economists to understand how current quantity (mortality) and quality (morbidity) greatly affect all individuals' current and future choices, and how governments can shape such decisions via policies. This thesis is based on these general questions.

This thesis is organised in three main chapters following this introduction. While there are common elements across them, like the general background of interventions, datasets and bibliography, each chapter is self-contained. Each one has its own introduction, dataset section, results, conclusions and appendix, with few references between them. They explore policies related to health from different perspectives and use very disparate methodological approaches. The first measures the impact of a policy in terms of individuals' response to it. The second goes beyond the impact, and is about calculating the value of health technologies in order to allow a policy maker to decide whether to implement them or not.

And finally, the last one is a general empirical test designed to understand the functioning of pay-for-performance schemes, a standard approach for adopting health technologies.

The first study is devoted to the role of information treatments in health investments. Regular health checks are commonly proposed as a means of slowing the progress of several non-transmittable diseases; specifically, they allow for the early diagnosis of conditions that increase the odds of serious health complications. This is the case for high blood pressure or hypertension, which has no evident symptoms but increases the odds of developing a cardiovascular event such as a heart attack or a stroke. Early diagnosis means that patients can receive early treatment. However, the identification of the potential impact of receiving health information is challenging as individuals' demand for preventive care is likely to be related to other health investments and with their beliefs about their morbidity progression. Chapter 2 estimates the potential impact of early diagnosis programmes on medication, objective and subjective health measures and lifestyle.

In order to deal with potential selection bias due to screening, I employ a feature of the *English Longitudinal Study of Ageing* (ELSA) that motivates a regression discontinuity design based on the blood pressure of respondents. ELSA was designed following the *Health Survey for England*, which has a strict protocol for assessing objective measures of health approximately every four years, starting in 2004. The survey, which includes information such as blood pressure or cholesterol levels, is collected from respondents by nurses. One of the key elements of the protocol is the standardised feedback that is given to respondents about such biomarkers. For the specific case of blood pressure, respondents above the NHS high blood pressure thresholds are told about the potential risk of being hypertense. It is also suggested to them that they visit their GP in order to get a proper assessment, as the survey measure does not follow diagnosis protocols. This survey design allows for identification of the effect of information on health beliefs and behaviour based on a regression discontinuity design.

As a result of the exercise, there is evidence of a temporal increase in the probability of being prescribed medication for blood pressure (4.41 pp) as a treat-

ment for the condition, which has almost doubled the proportion of people on medication for these levels of blood pressure. At the same time, there has been a permanent reduction in alcohol intake frequency of 8.4 pp and an increase in fruit consumption. However, there is also evidence of higher smoking intensity (5 cigarettes per week) and a higher probability of being obese (11 pp) for those above the threshold. Moreover, no clear effects on either objective or subjective health were found after four years of the intervention. These results suggest that this type of information-based interventions might have a strong impact on the demand for preventive care treatments, with permanent positive effects on behaviour at the same time.

While chapter 2 considers one potential preventive intervention, chapter 3 is about how to value these policies. In general terms it contributes to the literature that estimates the value improvements on life quantity and quality. There is plenty of literature on the topics of cost-benefit and cost-effectiveness analysis. There are established tools like the *quality adjust life years* (QALY), which is used for assessing the medical value of an intervention. It gives a fixed weight to potential health estates, based on the subjective measure of quality of life, in order to compute the gains of a treatment. Nevertheless, this does not take into account other potential benefits (or losses) derived from a given intervention. For instance, by how much would average labour income change if we managed to gain 1 QALY? These types of effects arise as a reduction in the number of people who suffer from work disabilities might increase labour participation. However, reducing the prospects of suffering such a type of disability might also decrease incentives for extra savings before retirement. An alternative to QALYs and cost-effectiveness analysis is to consider how much of their wealth are individuals willing to give up in order to enjoy the treatment. Willingness-to-pay calculation is the basis of cost-benefit analysis. This requires a counterfactual scenario that considers not only health but also financial variables. In Chapter 3, I contribute to this literature by introducing a structural dynamic life-cycle model for studying the economic value of the adoption of medical innovations. It allows for both cost-benefit and cost-effectiveness calculations, by considering long-run gains on QALYs, productivity, and welfare derived from adjusting savings and labour supply throughout life.

Such a tool considers how realised and expected health shocks modify trade-offs in the life-cycle, and its parameters are estimated using a rich longitudinal dataset that involves both health and financial information.

With this model, I consider the case of a medication, namely statins, which reduces the odds of developing cardiovascular diseases. Using data from ELSA between 2002/03 to 2012/13, the calculated value of this use of the drug is £79 billion. This is nearly 12% higher than considering the value derived from assuming a willingness-to-pay of £23.000 per QALY, a more standard valuation strategy. I also explore how the value depends on the effectiveness of the drug and of policies directed to its diffusion in the primary care system, as well as non-health related elements such as retirement age. It is also shown that one of the main drivers of the results is the implied willingness-to-pay for extending longevity. This concept, related to the value of a statistical life, is governed by the bequest motive formulation.

The prescription of statins is done at primary care level, and it is incentivised under the *Quality and Outcomes Framework* (QOF). This is a pay for performance scheme that rewards family doctors for accomplishing several goals. For instance, doctors are paid more for every increase in the percentage of diabetic patients who have their cholesterol controlled until this figure reaches 70%. This system of payment motivates Chapter 4. In this third and last project, a joint effort with Marcos Vera-Hernández, we develop a test for assessing a central property of this type of reward system: whether effort is diverted from one task into another after changing the reward of one the tasks. This is a central question as overall quality of care might be affected by modifying the incentives of some of the tasks. Such a type of response occurs when exerting effort in a particular task might result in an increase (substitutes) or decrease (complements) in the marginal cost of alternative efforts. While the empirical implications of these changes in a contract are straightforward, in practice it is difficult to isolate optimal responses from other concurrent unobserved shocks. This is a common challenge that arises due to the lack of adequate control groups, as incentive systems are typically rolled-out at the same time for an entire target group.

We contribute to the analysis of multitasking by using as a control group ob-

servations close to the presence of *kinks* in the reward function. In other words, agents who decided to exert effort just above the point in which there is a drastic change on the marginal benefit of effort. The relevance of such kinks in agents' choices can be assessed by their impact on the density function: it produces bunching near the kink point. This is because agents who would have decided to exert a higher effort without the kink, because of it choose to be exactly (or slightly above) at the kink point. Our test proceeds in two steps. First, we test whether or not the kink has affect on agents' choices, In other words, if there is bunching. Second, if that is the case, we test whether or not the response on effort to a change in the rewards between agents whose decisions are affected by the kink. We show that as agents at the kink are less likely to react to such a change in conditions, these individuals constitute a control group.

In terms of the QOF analysis, we show that changes introduced in 2010/11 revealed that tasks for which there was no price variation are not substitutes of those tasks. In fact, several indicators are complements.

Chapter 2

Early diagnosis of chronic conditions and lifestyle modification

2.1 Introduction

The rise in public expenditure due to an ageing population is partly due to diseases that could be prevented or delayed by modifying the habits of patients. One of the potential solutions is a preventive strategy based on the early treatment of individuals who are at risk of potential complications. This idea motives the strategy of periodical health checks on the population. Massive programmes such as NHS Health Checks in the UK, or some preventive care components of the Affordable Care Act in the US, point in that direction. For instance, the former invites people aged between 35 and 74 to routine check-ups for detecting signs of chronic conditions such as cardiovascular diseases. However, some authors like MacAuley (2012) consider that the impact of such policies might be even negative due to the misallocation of resources, over-diagnosis of certain conditions and to behavioural effects.

A first question about these programs is about their potential for inducing changes on demand for health care. A review by Krogsbøll et al. (2012) found that in general this type of programmes increased the number of individuals using anti-hypertensive drugs, but without conclusive effect on health benefits. In the specific case of the UK, there is no evidence so far on the benefits of the NHS Health Checks programme. Some studies like Artac et al. (2013) or Cochrane et al. (2012) provide descriptive evidence of the potential problems and benefits of the intervention in small areas of the country. Robson et al. (2016) suggests that

NHS Health Checks is related to an increase on the attendance to GP practices of individuals in risk of developing cardiovascular diseases (CVD). They have also observe increased prescription of medication for controlling high blood pressure (HBP) as well as for lowering cholesterol.

Second, there is an special concern of the effects of periodical health checks on risky behaviours. There is evidence that individuals might be sensible in terms of information related to their own health, consistent with the idea of rational addiction (Arcidiacono et al., 2007). Moreover, smokers' tend to be optimistic about their own mortality (Khwaja et al., 2007), and are updated with the onset of diagnosis of smoking-related diseases (Smith et al., 2001). However, treatment for mild conditions detected with the checks might induce risk compensation/offsetting behaviours. In other words, individuals could potentially increase their risky behaviour in response to improved prospects of future health due to medical treatment, or due to reassurance when they receive 'good news'. This is a common concern on areas like unsafe sexual activity and HIV treatment (Cassell et al., 2006). In order to understand this potential side-effect, it is required to analyse whether medical treatment and health behaviours are complements or substitutes in the context of a competing risks model. In principle, theory suggests complementarities between health investments as reducing one of the risks increases the marginal benefit of reducing the others (Becker, 2007; Dow et al., 1999). However, if lifestyle gains in reducing a disease-specific risk are offset by medical treatment, substitution effect might dominate (Kaestner et al., 2014). So far Kahn (1999) found that diabetics lifestyle improved over time without signs of medication, Fichera and Sutton (2011) suggests that statins were associated for lowering cholesterol with reductions on smoking in England. On the other hand, Kaestner et al. (2014) found an increase in obesity in response to the use of statins and no effect on smoking.

This chapter contributes to both the understanding of health advice effects and the analysis of complementarity or substitution between medical treatment and health behaviours. First, I am able to identify the medium and long run impact of informing individuals about the odds of being hypertensive, a condition that might increase the likelihood of developing cardiovascular diseases (CVDs).

Second, given this evidence, I am able to test if individuals modify their lifestyle and beliefs about their current and future health status in response to medical intervention.

My identification strategy to estimate the causal effect of receiving medical advice relies on the protocols of the *English Longitudinal Study of Aging* (ELSA) and the *Health Survey of England* (HSE). During the progress of the survey, a nurse takes the blood pressure of interviewees. In ELSA, those with a systolic/diastolic reading higher than 149/85 mmHg, are encouraged to visit their family doctor in order to have a proper screening test to confirm the findings. A similar procedure is in place in HSE with a 160/95 threshold for men aged 50 and over. As a result, we can compare individuals aged above fifty, not previously diagnosed with HBP, who are very similar in their health status but who differ only in having being advised or not to visit primary care services. This motivates a Regression Discontinuity Design (RDD) that identifies the impact for individuals who are close to the advice thresholds but who had not previously been diagnosed with any cardiovascular conditions.

A significant increase of 4.41 pp in the use of BP-lowering medication was found around to years after the intervention for those with a systolic BP slightly above the advice threshold compared to those below it. It almost doubles the proportion of individuals who are under such medication at this level of blood pressure. After 2 waves (approximately 4 years), the difference on prescription drops to 1 pp. and it is not statistically different form zero. This is in line with previous findings in the health checks literature that found an increase on medication use. Additionally, the advice caused a permanent decrease on alcohol intake of 8.4 pp and a positive impact on fruit portions per day. However, it also caused an increase on self-reported smoking intensity of 5 cigarettes per week, and of 11 pp on the probability to the obese (BMI>30). These findings suggest that improvements on fruits consumption and heavy drinking corresponds to a direct response to the threat of worse future health. Under Kaestner et al. (2014) framework, results for smoking and obesity indicate substitution between medical treatment and lifestyle. However, there is no evidence that such extra risky behaviour is in direct response to BP medication induced by the advice.

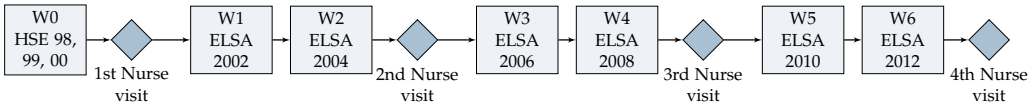
Results suggest that this type of information-based interventions might have strong impacts on demand for preventive care treatments, with permanent positive effects on behaviour at the same time. Moreover, impacts are stronger on men and on individuals will low risk of developing CVDs, showing that the policy might be effective for targeting this specific population.

This chapter is organized as follows. The introduction was the first part of this chapter. Section 2.2 presents the main details of the dataset and the sample employed and explains the health advice procedure by the survey nurses. Section 2.3 discusses the empirical strategy and Section 2.4 the main findings. Finally, Section 2.5 concludes.

2.2 Data

I use the *English Longitudinal Study of Ageing* (Marmot et al., 2013) for the years 2002-2012. It is a longitudinal study with a representative sample of those aged 50 and over in England. Its baseline was constructed using the *Health Survey for England* (NatCen and UCL, 2010) and it contains high-quality subjective and objective health information and detailed socio-economic information.[1]

Figure 2.1: Survey timing



Notes: The *English Longitudinal Study of Ageing* (ELSA) is based on the original sample from the *Health Survey for England* (HSE)

Additionally to the core interview, I use the biomarkers data collected in waves 0, 2, 4 and 6 (see Figure 2.1). All core individuals[2] who had a interview in person were eligible for the blood pressure measurements (BP), and depending on their health, to other measures.[3] After completing the questionnaire, respondents were asked for their approval to be visited by a nurse[4] in the following weeks. If they agreed, an appointment was made for between 2 to 4 weeks after

[1] More details can be found on the survey website:`http://www.ifs.org.uk/ELSA/about`

[2] ELSA collected information about partners, even if they were not part of the original HSE sample. These 'new' partners were not eligible for biomarkers measurements.

[3] For example, for blood samples eligibility depended on non-suffering a condition or being under a medication that implies that the test might compromise respondent's health.

[4] They are professional nurses trained by the researchers to take the measures following a strict protocol.

the interview. Diastolic and systolic blood pressure was derived by taking into account the last two of three measurements,[5] using an automated monitor under standardized conditions.[6]

As cooperation is a choice, the observed sample might be affected by selection. In particular, there is evidence which suggests that respondents are usually more likely to be more worried about their health and to engage in practices for preserving it (Heidi Guyer, 2010).

2.2.1 Descriptive information

For this exercise we consider only individuals for whom there are at least three valid BP measurements in at least one of the waves. Table 2.1 presents the descriptive information of this sample for each wave. In general, from Panel A, our sample is getting older during the observed time despite the refreshment samples that have been added since wave 3.[7] Though younger cohorts are more educated, the levels of education are represented in similar proportions across time as other characteristics as well as ethnicity, gender and marital status.

Panel B presents the evolution of self-reported health conditions. As our sample gets older, the prevalence of most diseases increases. The opposite occurs for lifestyle as observed in Panel C. There is a declining trend in the prevalences of smoking[8] (both in the extensive and intense margins) and alcohol intake.[9] Such a trend is not clear for the case of physical activity.[10] Two final measures on veg-

[5]The protocol discards the first measurement in order to minimize the *white coat syndrome*. Essentially, anxiety and stress produced by clinical settings temporally increases blood pressure but without being associated with cardiovascular risk (Pickering, 1996).

[6]People were asked to sit quietly 5 minutes before the measurement. Nurses were also instructed to delay the start of the measurements until at least half an hour after their arrival. Other conditions that might be relevant, such as ambient air temperature, was recorded. If the respondent had eaten, drunk, smoked or exercised in the last half an hour, his answers would be invalid.

[7]HSE 2002 to 2006 data is not used in some of the specifications due to the lack of information on hypertensive status.

[8]In ELSA, individuals are asked about smoking as part of the health module. If they report to be currently smoking, they are asked whether they use cigarettes and/or roll-ups. In both cases, they are ask about their consumption on weekdays and weekends separately: number of cigarettes and/or grams/ounces of tobacco. Around 23% of the smokers report to be roll-up consumers only, and I assumed 1 gram to be equivalent to 1 cigarette, and 1 ounce to be 28.35 cigarettes. The top 1% of these measures are excluded as they seem to be outliers. One important concern is variation on prices: Leicester and Levell (2012) and Czubek et al. (2010) have a good description on the evolution of real prices and consumption trends during the period. Relevant actions were in 1998 where the NHS quit was implemented and in 2007, when bans on smoking in public spaces were implemented.

[9]ELSA questions on alcohol intake is part of a self-completion module, and they vary from wave to wave. The present classification tries to capture the available information in a way that is comparable across waves.

[10]A recoded version of the level of physical activity derived by NatCen. These questions are part

etable and fruit intake are included in the ELSA, and as shown in the table, there is a substantial difference on how they were measured after wave five.[11] Such discrepancies are not problematic for the estimation of the sign of the impacts, as this compares variations within waves. However, the interpretation of the magnitudes is difficult as the estimates mixes both type of measures.

Panel D and E present subjective and objective measures of health. Evolution of objective health measures is not homogeneous. Some of them deteriorate on time: individuals are getting fatter (BMI and obesity), with higher levels of cholesterol; but their blood pressure is decreasing as the same time. First, binary variables for reporting to be in good and bad health are derived from standard likert scale type of question for self-rated health. ELSA also involves subjective probabilities on the chances to survive age 75; and the chances of suffering an event that limits ability to work. The former question is asked to individuals aged 60 and younger, and the later only to those who are currently working. Interesting, despite the an increasing proportion of individuals being diagnosed with hypertension or diabetes, all subjective health measures are on average increasing on time.

Finally, Panel F presents information on financial variables derived by the Institute of Fiscal Studies (IFS). Measures of income, savings and wealth, as well as labour supply, are included. Values in the top 1% of these variables are excluded as these values can be considered atypical for the rest of the distribution.

As this study aims to understand the effect of receiving advise about potential undiagnosed hypertension, the objective population has to be those who are in risk of such condition and are less likely to be tested for it. Falaschetti et al. (2014) documents that both systolic and diastolic BP increase with age until age 60 where the diastolic measure start to decrease systematically. The also show that by 2011, prevalence of hypertension was 28% for the age group 40-49, 40% for 50-59, and 60% for 60-69. Nevertheless, the authors documented an increase in awareness and management of the condition between 1994 (46%) and 2011 (71%). This is related

of the health module and involve both leisure and labour activities.

[11] These questions are part of the self-completion questionnaire. For waves 3 and 4, individuals have to record the total number of fruits/vegetables per item in a list, and then the number was added up in order to construct the measure. In contrast, waves 5 and 6 ask directly for the number of portions consumed per day.

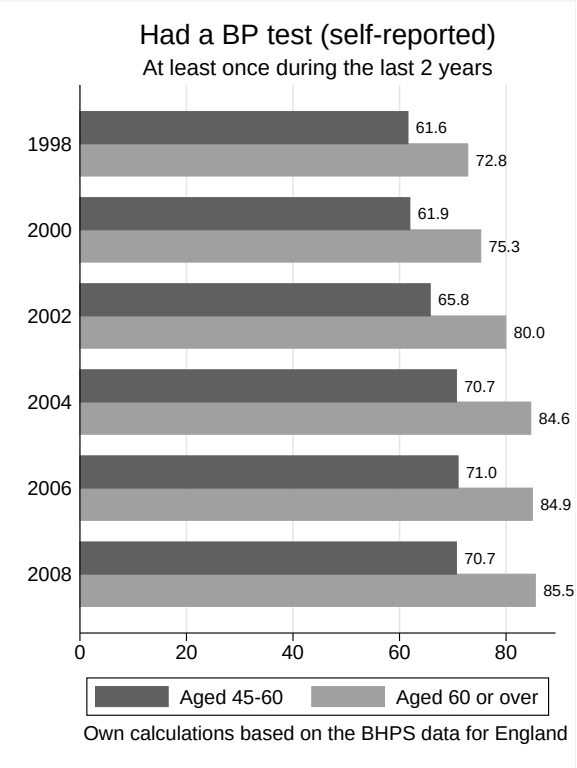
Table 2.1: Sample Means by Wave

Variables	Wave 0	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
Panel A. Socio-demographic Characteristics							
Age	60.6	63.5	65.8	67.9	66.3	67.3	69.0
Male	43.9%	43.9%	43.9%	44.5%	44.8%	44.8%	44.9%
Educ: No qualifications mentioned	38.9%	38.9%	36.6%	32.2%	30.8%	27.0%	27.8%
Educ: Some medium qualif.	36.3%	36.3%	36.9%	38.7%	39.2%	40.1%	39.8%
Educ: Some high level or above qualif.	24.9%	24.9%	26.5%	29.1%	30.0%	32.9%	32.5%
Non white ethnicity	29.0%	2.6%	8.1%	2.9%	6.3%	2.5%	4.6%
Married	71.2%	71.1%	63.4%	65.1%	61.0%	66.8%	61.3%
Panel B. Health Conditions							
Diagnosed HBP ever	15.3%	23.8%	47.1%	48.7%	48.8%	46.4%	52.6%
High Cholesterol, wave 2 onwards			18.8%	34.6%	35.7%	42.2%	46.6%
Diagnosed Diabetes ever	2.4%	5.9%	8.4%	10.4%	11.2%	11.4%	13.2%
Takes BP medication	11.4%	17.6%	32.0%	36.0%	32.8%	34.9%	35.7%
Takes Lipid-lowering medication				21.4%	22.6%	25.9%	28.0%
Diagnosed Major Cardiovascular Event ever	6.4%	13.2%	18.2%	18.0%	17.4%	15.2%	18.3%
Panel C. Lifestyle							
Current smoker	17.5%	16.3%	13.8%	10.1%	11.6%	10.4%	9.3%
Cigarettes per week (0 for non-smokers, includes rollups)	0.0	13.7	10.8	7.7	9.0	7.8	6.9
Alcohol twice a week or more	64.5%	59.3%	43.8%	42.7%	41.0%	40.9%	38.3%
Sedentary or low physical activity		29.9%	30.0%	30.1%	29.2%	29.1%	30.7%
Portions of vegetables per day				5.3	5.7	2.8	2.9
Portions of fruits per day				5.5	5.2	2.2	2.2
Panel D. Health Perceptions							
Self-reported good health	70.5%	71.7%	73.5%	68.8%	75.0%	75.8%	73.2%
Self-reported bad health	7.3%	24.2%	26.5%	31.2%	25.0%	24.2%	26.8%
SSP: Chances to live to age 75		65.5	65.4	67.1	67.8	68.9	68.0
What are the chances that your health will limit your ability to work before you		37.8	35.4	33.3	32.8	32.3	29.9
Panel E. Health Measures							
BMI: Body Mass Index (kg/m2)	27.4		27.9		28.3		28.2
Waist-to-height ratio (WHtR)			0.6		0.6		0.6
Overweight or above: BMI 25+	68.7%		72.6%		73.5%		72.8%
Obesity level 1 or above: BMI 30+	23.2%		28.7%		31.2%		30.5%
Blood HDL level (mmol/l)	1.5		1.5		1.6		1.7
Blood total cholesterol level (mmol/l)	5.9		5.9		5.6		5.5
Blood glucose level (mmol/L) - fasting samples only			5.0		4.9		5.4
(D) Valid Mean Systolic BP	138.3		135.1		132.6		132.5
(D) Valid Mean Diastolic BP	76.2		75.0		74.3		73.1
Panel F. Economic activity							
BU total weekly income (£ of May2005)		0.4	0.4	0.4	0.4	0.4	0.5
BU total savings (1000£ of May2005)		22.7	27.4	32.5	37.6	36.5	36.8
BU total net (non-pension) wealth (1000£ of May2005)		232.8	277.8	307.7	310.6	302.6	307.0
Hours of work all jobs (employed or self employed)		35.9	34.7	32.8	33.8	33.2	32.1
Working		41.0%	35.1%	30.7%	36.0%	32.7%	27.3%
Individuals	6572	6572	8538	5627	9059	7056	7308
Year	98-00	2002	2004	2006	2008	2010	2012

Source: own calculations using HSE 1998,99,00 for wave 0 and ELSA waves 1-6.

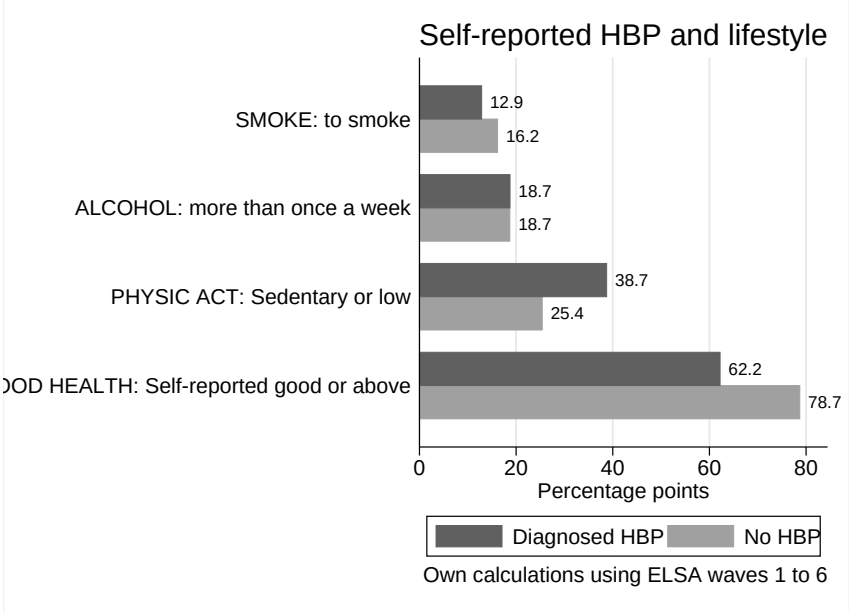
to the proportion of individuals who regularly had blood pressure tests. Between 1998 and 2008, data from the *British Household Panel Survey* (ISER, 2010) shows that there was an increase from 61% to 80% on the proportion of individuals aged 45 to 60 report having had their BP tested in the last two years (see Figure 2.2). The proportion is larger for the older group, going from 73% to 86%. As a result, despite improvements over time, while prevalence is higher in older individuals, testing is lower in the middle-age group. Hence, it is expected that this type of intervention would be useful for younger individuals.

Figure 2.2: Demand for BP screening tests



Another element of discussion is the relationship between lifestyle and diagnosis of hypertension. Figure 2.3 presents the correlation between habits and self-reported HBP that arises from the ELSA. It shows that, in general, individuals who report having been told by a doctor about being hypertense are less likely to smoke or to consume alcohol more than once a week, but at the same time are more likely to have a sedentary life.

Figure 2.3: HBP and lifestyle



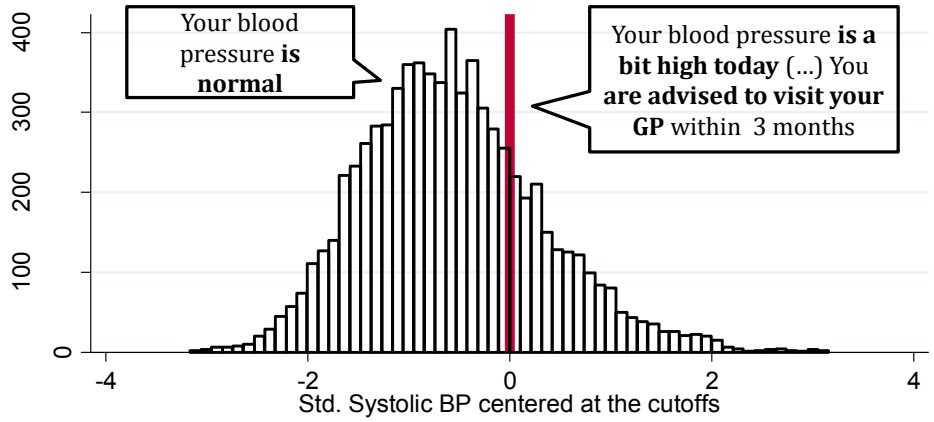
2.2.2 Health Advice Intervention

There is a particular characteristic of the ELSA that makes it ideal for our purposes. As previously indicated, nurses hired by ELSA visit the survey respondents two weeks after the survey interview and take their BP. According to the ELSA protocol, the nurses advise respondents to visit their family doctor (general practitioner, GP) if their BP measure is above a certain threshold (see below). This message might induce some individuals to visit their family doctor and get an adequate screening to stablish whether they suffer from hypertension.

Essentially, the advice varies with the last 2 out of 3 measurements of respondents' systolic/diastolic BP. In the ELSA, the thresholds are 140/85 mmHg for mildly raised blood pressure, 160/100 mmHg for moderately raised and 180/115 mmHg for considerably raised. Below 140/85 mmHg, the blood pressure was considered normal. In the HSE, the values were the same for women and men under 50, but changed for men aged 50 or over.[12] A respondent with mildly raised blood pressure was instructed to visit their GP in the next 3 months, for moderately raised it was 3 weeks, and for considerably raised, 5 days. These thresholds are similar to the official recommendation for systolic BP used by the NHS, where

[12] 160/95, 170/105 and 180/115 mmHg respectively

Figure 2.4: Systolic Blood Pressure Distribution and Nurses' Advice



hypertension is diagnosed with 140/90 mmHg (NICE, 2011). For diastolic BP the recommendation is quite conservative and we will see this reflected in the results. Figure 2.4 presents the strategy that will be followed in this chapter: the BP measures are standardized around the relevant mildly raised cut-off according to respondents' age, gender and year of the survey. For this analysis, an individual is treated if such a measure is greater or equal to 0, and is a control otherwise.

Nurses were clearly instructed to provide only the survey interpretation. Respondents were allowed to avoid feedback from the readings, or to allow the results to be sent to their GP.[13] That information could be left written in a "measurement record card" along with other biomarkers.[14] The suggestion given by the nurses was homogeneous as stated by the survey protocol. For instance, in the case of moderately raised blood pressure, they will tell the respondent:

> *Blood pressure can vary from day to day and throughout the day so that one high reading does not necessarily mean that you suffer from high blood pressure. You are advised to visit your GP within 2-3 weeks to have a further blood pressure reading to see whether this is a once-off finding or not.*

[13] Unfortunately,public available data does not report these choices.
[14] There were not any other comments or suggestions based on the biomarkers of the survey.

2.3 Empirical Strategy

The previously described nurse protocol motivates a Sharp Regression Discontinuity Design (RDD). The idea is to compare the value of the outcomes in the waves following the measurement, for those individuals that were just below and just above the threshold. By doing this, we are assuming that having the maximum standardised BP measurement slightly above or below the advice cut-off is essentially random once we take into account the trend. Formally, following Imbens and Lemieux (2008), the impact of nurse advice at wave t, $W = 1$, on outcome Y_{t+s} at wave $t + s$ ($s \in \{1,2\}$) is identified by the discontinuity in the conditional expectation of such outcome at the advice cut-off $BP = 0$:

$$\begin{aligned} \delta_0 &= E[Y_{i,t+s}(W = 1) - Y_{i,t+s}(W = 0)|BP_{i,t} = 0] \\ &= \lim_{BP \downarrow c} E[Y_{i,t+s}|BP_{i,t} = 0] - \lim_{BP \uparrow 0} E[Y_{i,t+s}|BP_{i,t} = 0] \end{aligned} \tag{2.1}$$

This strategy identifies the impact of the policy on the outcomes of a particular group of individuals. First, it tell us how individuals who might be considered to have mildly raised blood pressure would react to the diagnosis of such a condition. Second, it measures how people who comply with the advice react: that is, those who visit their GP as the nurse told them to, and who would not do so in the absence of the nurse advice.

Main results are presented based on the estimated parameter δ from Equation 2.2, which identifies δ_0 in Equation 2.1. Essentially, within a bandwidth of 1 standard deviation ($h = 1SD$) of the cut-off, a second order polynomial is fit at both sides of the cut-off in order to capture the observed relationship between prescriptions and blood pressure (see Figure 2.5, described in detail in the results section).

$$\begin{aligned} Y_{i,t+s} = \delta W_{it} + \alpha_0 + f_l(\alpha_l, BP_{i,t}|W_{it} = 0) + f_r(\alpha_r, BP_{i,t}|W_{it} = 1) \quad &, s \in \{1,2\} \\ f_x(\alpha_x, BP_{i,t}|W_{it} = 0) = \alpha_{x,1} BP_{i,t} + \alpha_{x,2} BP_{i,t}^2 \quad &, x \in \{l,r\} \\ \forall BP_{i,t} \in [-h,h], \quad h = 1 \end{aligned} \tag{2.2}$$

Given that δ_0 can be estimated under different bandwidths h and functions $f(\cdot)$, it is essential to test alternative specifications. Main tables will present results based on local linear regressions with rectangular and triangular weights.[15]

Balancing tests are carried out in order to test the validity of the main assumption. These test consist on running Equation 2.3 with $s = 0$. Such regression analysis assess if the discontinuities were in place before the nurse advice took place. Also, it is possible to determine if the effect is related to other pre-existing elements in the data. This is done by setting socio-demographic characteristics as left-hand side elements in the regression.

2.4 Results

2.4.1 Main Results

The intervention does increase the likelihood of being treated with medication for BP among those who report not being diagnosed with HBP at the moment of the nurse visit, around two years after they received the advice from the nurse.[16] Those who were above the systolic BP advice threshold were around 4.41 pp more likely to report that they were taking medication. This figure is significant at the 90% level and also large, as around 2% of the population with such BP levels take medication. For the diastolic BP the estimate is 0.05 pp, which is not significant at 90% level.

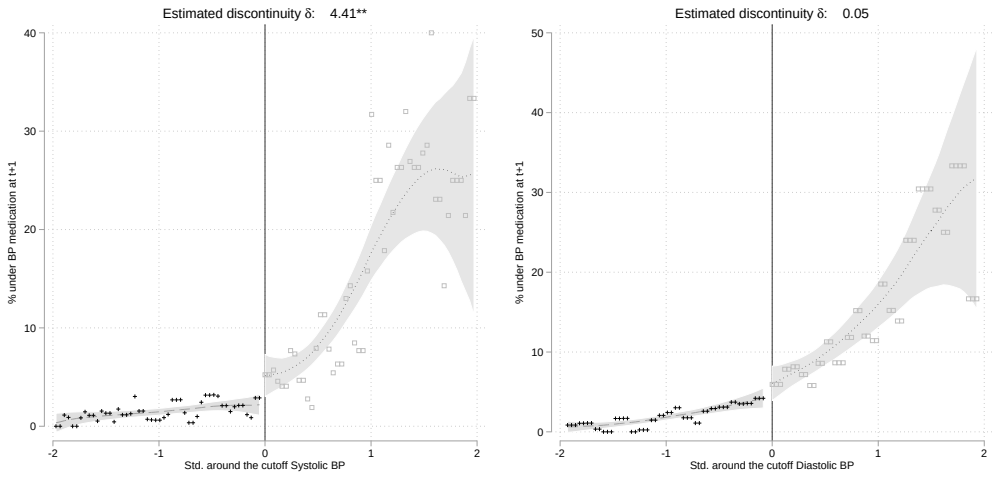
Figure 2.5 presents a graphic version of the RDD analysis. In both graphs, the horizontal axis shows the standardized BP measurement where 0 is the relevant cut-off. A smoothed average, using the triangular linear kernel, is represented by the dashed lines at both sides of the threshold. The goal of the strategy is to measure the jump between the dashed lines. The value reported in the graph corresponds to Equation 2.2, and which will be called the *Local Quadratic Rectangular* estimator in Tables below.

Tables 2.2 and 2.3 present the main results for one and two waves after the

[15] For further details see Appendix 2.B.

[16] The sample is selected in that way in order to avoid confounding factors. First, in general individuals above the threshold are more likely to report being diagnosed with HBP even before the nurses visited them. This is expected as the advice cut-off is equivalent to the common diagnosis threshold. Second, individuals in their fifties will benefit the most from the health checks, as they are less likely to demand primary health care in the first place as shown in Figure 2.2. Age is explored with more detail in Section 2.4.3. For more details see Appendix 2.A.

Figure 2.5: Nurse Advice and BP lowering medication at the following-wave



Sample: Individuals aged 60 or younger who reported not to be diagnosed with HBP or diabetes from the HSE-ELSA data.

Notes: Calculations using a quadratic function within 1 standard deviation of the cutoff. A 90% CI is presented. Significance level: *90%, ** 95%, *** 99%

nurse visit. In both of them, the rows present the outcome variables. Panel A shows the jump estimator for health conditions and medications; Panel B does so for lifestyle indicators; Panel C for health perceptions; and Panel D, only in Table 2.3, covers objective health measures. The first column is the mean of each dependent variable for those observations one standard deviation below the threshold. The other columns present different specifications for the trend between the outcome and systolic blood pressure. Last column, number 4, corresponds to the estimate of δ according to Equation 2.2. In the rows, standard errors are presented as well as the number of observations included. They differ according to output variable and method.[17] As a comparison between variables, the reader can check the common bandwidth of one standard deviation ($h = 1$). This sample size is used for the main results in Column 4.[18] Notice that some variables have fewer observations as they were not collected in every wave (ex. fruits and vegetables), or because they are conditional on some characteristic (ex. cigarettes per week for those who reported to be smokers at the wave of the nurse advice).

[17]See Appendix 2.B for more details on the optimal bandwidth for local linear regressions estimates presented in Columns 2 and 3.

[18]One standard deviation of systolic blood pressure is between 19 and 20 mmHg. Appendix 2.C shows that the BP medication estimates are robust to the bandwidth selection.

Before presenting results, tables 2.2 and 2.3 present the differential attrition below and above the threshold. One standard deviation below the threshold, average attrition is around 15% approximately two years after the measurement. This figure is nearly 25% after four years. Nevertheless, there is no observed systematic difference above and below the cut-off.

Table 2.2: RDD next wave (apx. 2 years) outcomes

RDD on systolic BP standarized around the nurse advice cut-off.

$$Y_{i,t+1} = \delta(BP^c_{i,t} \geq 0) + \alpha_0 + f_l(\alpha_l, BP^c_{i,t}|BP_i < 0) + f_r(\alpha_r, BP^c_{i,t}|BP_i \geq 0) + u_{i,t+1}|Age_{i,t} <= 64$$

Dependent Variable at $t+1$	(1) Mean 1SD Below	(2) Loc Linear Triangular h_1^*	(3) Loc Linear Rectang. h_2^*	(4) Loc Quad Rectang. $h = 1SD$
Missing this wave	15.17%	0.02		−1.24
N: 4000 ($h_1^* = 0.91$), ($h_2^* =$) , *4476* ($h = 1$)		(2.47)		(3.22)
Panel A. Health Conditions				
Diagnosed HBP ever	5.74%	2.57	5.52*	2.72
N: 3188 ($h_1^* = 0.87$), *2038* ($h_2^* = 0.53$) , *3772* ($h = 1$)		(2.32)	(2.95)	(2.95)
High Cholesterol, wave 2 onwards	24.43%	−1.41	−1.57	0.22
N: 2475 ($h_1^* = 1.41$), *1224* ($h_2^* = 0.66$) , *1888* ($h = 1$)		(3.93)	(5.93)	(6.08)
Diagnosed Diabetes ever	0.86%	−0.12	0.34	−0.21
N: 3585 ($h_1^* = 0.97$), *2828* ($h_2^* = 0.77$) , *3772* ($h = 1$)		(0.88)	(0.96)	(1.14)
Takes BP medication	1.94%	2.98*	4.69***	4.41**
N: 3373 ($h_1^* = 0.90$), *2252* ($h_2^* = 0.62$) , *3772* ($h = 1$)		(1.54)	(1.81)	(2.00)
Takes Lipid-lowering medication	9.32%	3.79	3.81	6.47
N: 1888 ($h_1^* = 1.01$), *1409* ($h_2^* = 0.74$) , *1888* ($h = 1$)		(3.14)	(3.94)	(4.17)
Diagnosed Major Cardiovascular Event ever (Stroke, Heart Failure, Infarction, An	2.46%	1.43	3.03**	1.29
N: 3017 ($h_1^* = 0.82$), *2829* ($h_2^* = 0.75$) , *3774* ($h = 1$)		(1.44)	(1.46)	(1.77)
Panel B. Lifestyle				
Current smoker	16.48%	3.49	4.14	2.47
N: 4775 ($h_1^* = 1.37$), *2817* ($h_2^* = 0.77$) , *3761* ($h = 1$)		(2.49)	(3.32)	(3.79)
Current smoker if smoker at t	83.30%	−1.60	−3.40	−2.22
N: 596 ($h_1^* = 0.85$), *423* ($h_2^* = 0.62$) , *697* ($h = 1$)		(5.81)	(6.82)	(7.57)
Cigaretes per week (0 for non-smokers)	10.87	4.38	5.68*	2.69
N: 3874 ($h_1^* = 1.12$), *3003* ($h_2^* = 0.87$) , *3557* ($h = 1$)		(2.66)	(3.19)	(3.68)
Cigaretes per week (0 for non-smokers, includes rollups)	13.04	5.32*	6.17*	5.12
N: 4156 ($h_1^* = 1.15$), *3287* ($h_2^* = 0.90$) , *3682* ($h = 1$)		(2.99)	(3.47)	(4.13)
Alcohol twice a week or more	55.39%	−7.19*	−10.57**	−8.41*
N: 3050 ($h_1^* = 0.87$), *2358* ($h_2^* = 0.62$) , *3621* ($h = 1$)		(3.97)	(4.62)	(5.09)

Continued on next page

Table 2.2: (Continued)

Dependent Variable at $t+1$	(1) Mean 1SD Below	(2) Loc Linear Triangular h_1^*	(3) Loc Linear Rectang. h_2^*	(4) Loc Quad Rectang. $h = 1SD$
Sedentary or low physical activity	17.47%	0.78	1.40	3.72
N: 3909 ($h_1^* = 1.05$), *3734* ($h_2^* = 1.00$) , *3734* ($h = 1$)		(2.68)	(2.80)	(3.75)
Portions of vegtables per day	3.94	0.02	0.26	-0.09
N: 1192 ($h_1^* = 0.69$), *1466* ($h_2^* = 0.86$) , *1734* ($h = 1$)		(0.46)	(0.47)	(0.56)
Portions of fruits per day	3.20	0.70**	0.51*	0.76*
N: 1646 ($h_1^* = 0.94$), *2146* ($h_2^* = 1.32$) , *1732* ($h = 1$)		(0.34)	(0.31)	(0.43)
Panel C. Health Perceptions				
Self-reported GOOD health	83.86%	0.29	-0.43	1.13
N: 2632 ($h_1^* = 0.72$), *2455* ($h_2^* = 0.62$) , *3755* ($h = 1$)		(3.30)	(3.50)	(3.83)
Self-reported bad health	14.64%	-0.12	1.31	-1.15
N: 3568 ($h_1^* = 0.94$), *3173* ($h_2^* = 0.85$) , *3755* ($h = 1$)		(2.73)	(2.84)	(3.63)
SSP: Chances to live to age 75	68.80	-0.57	-0.33	0.33
N: 2779 ($h_1^* = 0.73$), *1810* ($h_2^* = 0.52$) , *3707* ($h = 1$)		(1.79)	(2.24)	(2.10)
What are the chances that your health will limit your ability to work before you	34.88	0.97	-1.24	0.43
N: 2596 ($h_1^* = 0.99$), *1547* ($h_2^* = 0.62$) , *2596* ($h = 1$)		(2.41)	(2.95)	(3.27)

Sample: Respondents aged 60 and younger at the moment of the nurse advice, who were not diagnosed with HBP or being taking medication for lowering BP.
Notes: Column (1) presents the mean of each dependent variable for those observations one standard deviation below the threshold. Columns (2) to (4) present different specifications for the trend (function $f(\cdot)$) between the outcome and systolic blood pressure. Robust SE are presented in parenthesis. Significance: * 10%, ** 5%, *** 1%.

2.4.1.1 Health Conditions

The impact of advice on objective and subjective measures of health is presented in Panel A of Tables 2.2 and 2.3. First, self-reported diagnosis of HBP, diabetes, high cholesterol and other cardiovascular conditions or events.[19] Second, prescription of blood pressure medication and lipid-lowering medication[20] are analysed. It includes the main result: there is an increase on prescription of medication for lowering blood pressure of 4.41 pp approximately 2 years after the nurse advice (Table 2.2). Such estimate is based on the quadratic specification, but a slightly more conservative figure is presented by the local lineal average using triangular

[19]Diabetes, stroke, angina, heart attack - including myocardial infraction or coronary thrombosis -, congestive heart failure, a heart murmur, an abnormal heart rhythm, or any other heart trouble.
[20]Medication for lowering cholesterol use for prevention of cardiovascular diseases, mostly statins in the UK.

weights:[21] 3 pp significant at the 90% level. Apart from this result, there is no evidence of an increase on the odds of being diagnosed with high blood pressure, or diabetes. While in one of the specifications a positive impact is found for diagnosis of high blood pressure and of other cardiovascular diseases, such impacts are not robust to the specification. A notorious increase on medication for cholesterol is also reported: 6 pp relative to a prescription rate of 9.32% below the cut-off. Nevertheless, standard errors are large. One possible explanation is the reduced sample size: it is based on 1888 observations rather than the 3772 of other outcomes.

Approximately two years later (four after the nurse advice), below the cut-offs prevalence of detected hypertension increased from 6% to 11% (Table 2.3). Prescription of BP medication doubled from 2% to 4%. On the other hand, the difference at the cutoff decreased to 1 pp., and such figure is not statistically different from 0. In all other diagnosed conditions results are similar: there is no difference below and above the threshold after four years.

Apart from medication, family doctors normally give advice on lifestyle. Panel B of Tables 2.2 and 2.3 covers smoking, alcohol intake, physical activity and nutrition variables.

2.4.1.2 Lifestyle

As described before, ELSA covers carefully smoking on both the extensive and intensive margins. Two years after the nurse advice, 16% of the sample below the cut-off reports to be smoking (Table 2.2). However, there is a decreasing trend: 20% of former smokers have drop this behaviour. While the decrease seems to be larger above the cut-off by 2 pp, we cannot reject that such figure is different from 0. On the intensive margin, there is a difference of 5 cigarettes per week between those below and above the threshold. This estimate is similar if we consider a measure that includes both cigarette and roll-up smokers.[22] However, it cannot be rejected to be 0 under the second order polynomial specification.[23] Four years

[21] See Appendix 2.B for more details about this specification.

[22] Roll-ups are measured in tobacco ounces or grams, which is translated into 'cigarettes' in order to obtain a measure that avoids substitution between both types of smoking. While not including roll-ups avoids this measurement restrictions, it underestimates total smoking intensity.

[23] If we condition these intensity measures on being an smoker at the moment of the nurse advice, the impact on the roll-ups inclusive measure is estimated to be of 23 cigarettes, significant at 90%. It drops to 14 cigarettes if we do not consider roll-ups.

after the advice, 75% of the original smokers are still smoking (Table 2.3). The difference on intensity below and above the cut-off is estimated between 0.67 to 6.48 cigarettes, according to the measure definition and specification. Just as with the two-years estimate, the effect is not statistically different from zero for the quadratic specification.

With respect to alcohol intake, there is a clear reduction of 8.4 pp on the probability to report to be drinking at least two days a week. The impact is the same both two and four years after the nurse advice. This is a large impact on a very common lifestyle: more than half of the respondents below the cutoff have such alcohol intake frequency, a figure that drops to 4 in 10 two years later.

Finally, there are no effects on physical activity or vegetables consumption. There is evidence of a positive effect on fruit's intake, statistically different from zero two years after the nurse advice (Table 2.2). However, as explained in the data section, the exact amount of portions per day cannot be determined as the measure involves two different elicitation methods.

2.4.1.3 Health measures

Given that there is evidence that supports an effect on medication prescription and lifestyle choices, it is possible to expect an effect on both objective and subjective health.

With respect to perceived health, there is no evidence of an impact on reporting either good or bad health, subjective survival probabilities. Moreover, there is no effect on the reported chances to suffer a problem that limits ability to work, for those who were working at the time of the survey.[24]

With respect to objective health measures, there is evidence of an increase on the odds of being obese (BMI above 30) of 11 pp, relative to a prevalence below the cut-off of 27%. This impact is around 8 pp if we consider alternative specifications, and is not significant in all of them. This result is related to a positive but not robust to the specification effect on average BMI and waist-to-height ratio; both of them related with increased CVD-risk. On the other hand, there is no perceived difference on biomarkers as blood pressure or cholesterol despite the increase on medication.

[24]In Table 2.3, a significant difference of 4.3 pp was found for this outcome. However, the estimate becomes negative under the second order polynomial specification.

Table 2.3: RDD 2 waves (apx. 4 years) later

RDD on systolic BP standarized around the nurse advice cut-off.

$$Y_{i,t+2} = \delta(BP^c_{i,t} \geq 0) + \alpha_0 + f_l(\alpha_l, BP^c_{i,t}|BP_i < 0) + f_r(\alpha_r, BP^c_{i,t}|BP_i \geq 0) + u_{i,t+2}|Age_{i,t} <= 64$$

Dependent Variable at $t+2$	(1) Mean 1SD Below	(2) Loc Linear Triangular h_1^*	(3) Loc Linear Rectang. h_2^*	(4) Loc Quad Rectang. $h = 1SD$
Missing this wave	25.28%	3.52	3.13	3.78
N: 3778 ($h_1^* = 0.85$), *2943* ($h_2^* = 0.67$) , *4476* ($h = 1$)		(3.21)	(3.33)	(4.05)
Panel A. Health Conditions				
Diagnosed HBP ever	11.60%	2.15	3.40	−1.73
N: 4386 ($h_1^* = 1.42$), *3657* ($h_2^* = 1.11$) , *3355* ($h = 1$)		(2.69)	(2.73)	(4.19)
High Cholesterol, wave 2 onwards	22.23%	−2.68	−3.50	0.34
N: 3664 ($h_1^* = 1.12$), *2844* ($h_2^* = 0.88$) , *3363* ($h = 1$)		(3.20)	(3.29)	(4.61)
Diagnosed Diabetes ever	2.33%	0.61	0.48	1.07
N: 3869 ($h_1^* = 1.24$), *3143* ($h_2^* = 0.98$) , *3309* ($h = 1$)		(1.46)	(1.43)	(2.16)
Takes BP medication	4.36%	0.84	1.03	1.06
N: 4108 ($h_1^* = 1.30$), *3309* ($h_2^* = 1.02$) , *3309* ($h = 1$)		(1.96)	(2.02)	(2.91)
Takes Lipid-lowering medication	11.91%	1.96	1.84	3.76
N: 2520 ($h_1^* = 1.64$), *2129* ($h_2^* = 1.29$) , *1759* ($h = 1$)		(3.15)	(3.11)	(5.17)
Diagnosed Major Cardiovascular Event ever (Stroke, Heart Failure, Infarction, An	3.54%	2.24	2.55	2.24
N: 2812 ($h_1^* = 0.86$), *2332* ($h_2^* = 0.68$) , *3333* ($h = 1$)		(1.79)	(1.74)	(2.26)
Panel B. Lifestyle				
Current smoker	14.05%	3.00	3.43	3.13
N: 3143 ($h_1^* = 0.97$), *2474* ($h_2^* = 0.76$) , *3309* ($h = 1$)		(2.78)	(2.90)	(3.71)
Current smoker if smoker at t	75.93%	1.38	0.36	0.55
N: 525 ($h_1^* = 0.92$), *419* ($h_2^* = 0.72$) , *579* ($h = 1$)		(8.46)	(8.68)	(11.21)
Cigaretes per week (0 for non-smokers)	8.01	4.83**	5.38**	0.67
N: 4300 ($h_1^* = 1.52$), *3655* ($h_2^* = 1.20$) , *3132* ($h = 1$)		(2.19)	(2.20)	(3.35)
Cigaretes per week (0 for non-smokers, includes rollups)	10.39	6.20**	6.48**	3.08
N: 4765 ($h_1^* = 1.73$), *4144* ($h_2^* = 1.36$) , *3236* ($h = 1$)		(2.52)	(2.54)	(3.89)
Alcohol twice a week or more	42.82%	−9.00**	−9.42**	−11.19**
N: 2546 ($h_1^* = 0.87$), *2125* ($h_2^* = 0.68$) , *3028* ($h = 1$)		(4.32)	(4.39)	(5.53)
Sedentary or low physical activity	17.71%	2.66	3.70	2.48
N: 3134 ($h_1^* = 0.97$), *2467* ($h_2^* = 0.76$) , *3299* ($h = 1$)		(3.03)	(3.15)	(4.10)
Portions of vegtables per day	4.21	0.15	0.08	0.44
N: 2085 ($h_1^* = 1.44$), *1739* ($h_2^* = 1.13$) , *1590* ($h = 1$)		(0.45)	(0.48)	(0.59)
Portions of fruits per day	3.27	0.65	0.70*	0.60
N: 1033 ($h_1^* = 0.67$), *870* ($h_2^* = 0.52$) , *1595* ($h = 1$)		(0.41)	(0.40)	(0.45)

Continued on next page

Table 2.3: (Continued)

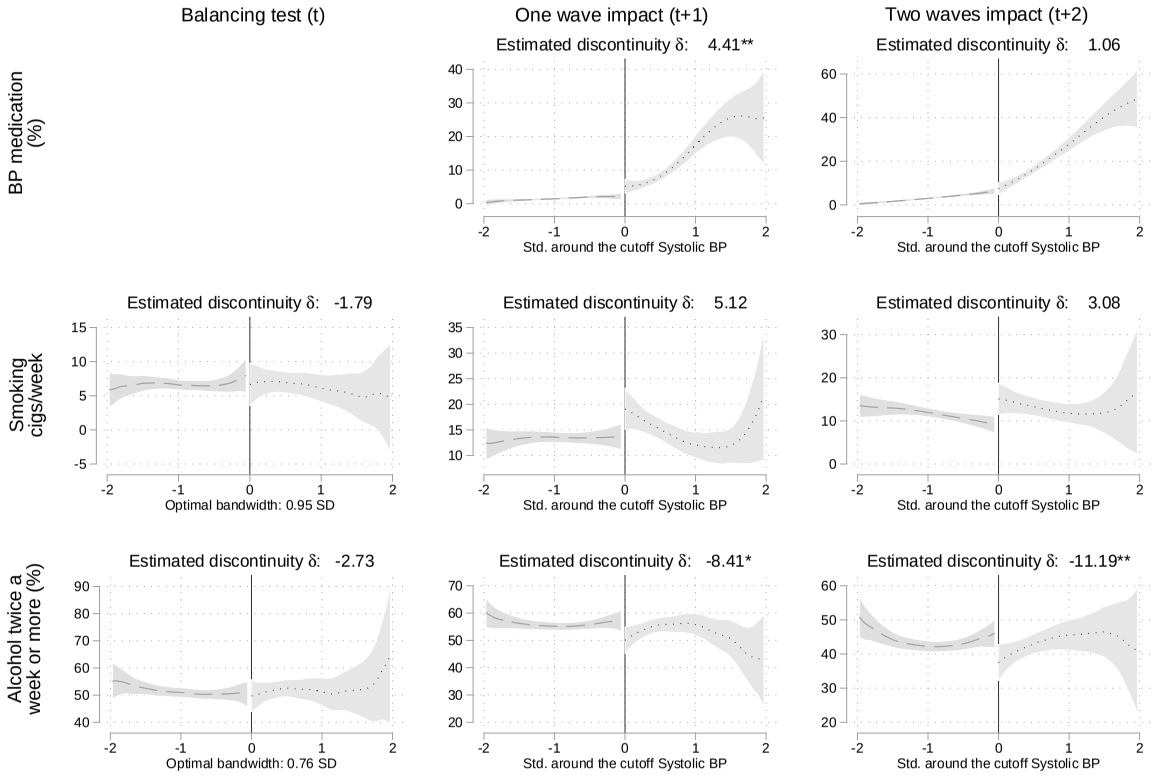
Dependent Variable at $t+2$	(1) Mean 1SD Below	(2) Loc Linear Triangular h_1^*	(3) Loc Linear Rectang. h_2^*	(4) Loc Quad Rectang. $h = 1SD$
Panel C. Health Perceptions				
Self-reported GOOD health	84.56%	2.42	3.17	3.58
N: 2299 ($h_1^* = 0.69$), *1771* ($h_2^* = 0.54$) , *3284* ($h = 1$)		(3.76)	(3.82)	(4.24)
Self-reported bad health	15.44%	−2.42	−3.17	−3.58
N: 2299 ($h_1^* = 0.69$), *1771* ($h_2^* = 0.54$) , *3284* ($h = 1$)		(3.76)	(3.82)	(4.24)
SSP: Chances to live to age 75	68.48	0.88	1.26	1.69
N: 2549 ($h_1^* = 0.80$), *2083* ($h_2^* = 0.63$) , *3202* ($h = 1$)		(1.90)	(1.95)	(2.34)
What are the chances that your health will limit your ability to work before you	31.08	3.39	4.32**	−0.73
N: 3046 ($h_1^* = 1.84$), *2629* ($h_2^* = 1.44$) , *2018* ($h = 1$)		(2.13)	(2.17)	(3.58)
Panel D. Health Measures				
BMI: Body Mass Index (kg/m2)	27.86	0.59	0.74	0.97*
N: 2092 ($h_1^* = 0.76$), *1679* ($h_2^* = 0.59$) , *2782* ($h = 1$)		(0.47)	(0.49)	(0.57)
Waist-to-height ratio (WHtR)	0.56	0.01	0.01	0.02*
N: 1956 ($h_1^* = 0.69$), *1531* ($h_2^* = 0.54$) , *2778* ($h = 1$)		(0.01)	(0.01)	(0.01)
Overweight or above: BMI 25+	70.72%	1.62	0.18	5.04
N: 2505 ($h_1^* = 0.91$), *1958* ($h_2^* = 0.72$) , *2782* ($h = 1$)		(3.80)	(4.00)	(5.03)
Obesity level 1 or above: BMI 30+	27.60%	8.14*	7.67	11.20**
N: 1958 ($h_1^* = 0.70$), *1532* ($h_2^* = 0.55$) , *2782* ($h = 1$)		(4.85)	(5.03)	(5.56)
Blood HDL level (mmol/l)	1.63	−0.03	−0.03	−0.02
N: 2120 ($h_1^* = 0.92$), *1654* ($h_2^* = 0.72$) , *2365* ($h = 1$)		(0.04)	(0.04)	(0.05)
Blood total cholesterol level (mmol/l)	6.01	−0.09	−0.12	−0.09
N: 2008 ($h_1^* = 0.88$), *1655* ($h_2^* = 0.69$) , *2367* ($h = 1$)		(0.12)	(0.12)	(0.16)
Blood glucose level (mmol/L) - fasting samples only	4.93	−0.04	−0.04	−0.02
N: 2420 ($h_1^* = 1.88$), *2149* ($h_2^* = 1.47$) , *1614* ($h = 1$)		(0.07)	(0.07)	(0.10)
(D) Valid Mean Systolic BP	128.85	−1.79	−1.27	−2.34
N: 1824 ($h_1^* = 0.72$), *1424* ($h_2^* = 0.57$) , *2601* ($h = 1$)		(1.61)	(1.68)	(1.86)
(D) Valid Mean Diastolic BP	76.22	−0.27	0.12	−0.32
N: 2722 ($h_1^* = 1.05$), *2086* ($h_2^* = 0.82$) , *2601* ($h = 1$)		(0.83)	(0.87)	(1.16)

Sample: Respondents aged 60 and younger at the moment of the nurse advice, who were not diagnosed with HBP or being taking medication for lowering BP.
Notes: Column (1) presents the mean of each dependent variable for those observations one standard deviation below the threshold. Columns (2) to (4) present different specifications for the trend (function $f(\cdot)$) between the outcome and systolic blood pressure. Robust SE are presented in parenthesis. Significance: * 10%, ** 5%, *** 1%.

Figure 2.6 presents a visual summary of the main results. Figure 2.5 style of graph is used for a set of outcomes in different moments of time. The first row presents BP prescriptions, the second smoking intensity, and the last one alcohol-

intake. Columns refer to the moment of measurement of each of these outcomes. The first one is contemporary to the measurement and the advice. This is done in order to verify that the discontinuity occurs after the intervention. These are balancing tests which are part of the robustness checks that are detailed in the next section. The second and third columns of the graph correspond to the estimates in column 2 of Tables 2.2 and 2.3 but with a fix bandwidth of 0.053.

Figure 2.6: Impacts on BP medication, smoking and alcohol intake



Notes: Horizontal axis: Calculations using a quadratic function within 1 standard deviation of the cutoff. A 90% CI is presented. Cigarettes per week includes roll-ups. Significance level: *90%, ** 95%, *** 99%. Sample: individuals aged 60 or less with no prior diagnosis of HBP or diabetes.

2.4.2 Complementarity of medication and health behaviours

Given that the advice caused an increase on medication use for reducing BP without increasing detection rate of HBP, the advice might be interpreted as an exogenous variation on medication use. Table 2.4 considers this exercise it in order to examine if lifestyle are substitutes or complements of BP detection as in (Kaestner et al., 2014). In particular, above we found higher rates of smoking and on obesity incidence among those who were advised to visit their GP. Using a Wald estimate, we can determine if such lifestyle change was a response to higher medication usage.[25] There is no evidence of this effect for any of the lifestyle variables that were reported to be affected by the policy. However, the large standard errors indicate that the sample might not be large enough to detect this behavioural consequence.

Table 2.4: Wald estimates: BP medication and lifestyle

Wald estimates based on the impact of the nurse advice cut-off on the dependent variable (numerator) and BP medication (denominator).

Dependent Variable at $t+1$	(1) Mean 1SD Below	(2) Loc Linear Triangular h_1^*	(3) Loc Linear Rectang. h_2^*	(4) Loc Quad Rectang. $h = 1SD$
Panel A. Lifestyle after one wave (apx. two years)				
Cigaretes per week (0 for non-smokers)	10.87	220.146	269.329	62.36
N: 3874 ($h_1^* = 1.12$), *3003* ($h_2^* = 0.87$) , *3557* ($h = 1$)		(211.324)	(262.487)	(90.73)
Cigaretes per week (0 for non-smokers, includes rollups)	13.04	238.746	241.117	114.49
N: 4156 ($h_1^* = 1.15$), *3287* ($h_2^* = 0.90$) , *3682* ($h = 1$)		(205.058)	(203.320)	(106.68)
Alcohol twice a week or more	55.39%	−261.691	−253.372	−208.94
N: 3050 ($h_1^* = 0.87$), *2358* ($h_2^* = 0.62$) , *3621* ($h = 1$)		(204.906)	(151.474)	(160.85)
Portions of fruits per day	3.20	18.171	22.422	13.77
N: 1646 ($h_1^* = 0.94$), *2146* ($h_2^* = 1.32$) , *1732* ($h = 1$)		(12.949)	(25.566)	(10.13)
Panel B. Lifestyle after two waves (apx. four years)				
Cigaretes per week (0 for non-smokers)	8.01	4.832	5.378	17.74
N: 4300 ($h_1^* = 1.52$), *3655* ($h_2^* = 1.20$) , *3092* ($h = 1$)		(2.193)	(2.205)	(87.84)
Cigaretes per week (0 for non-smokers, includes rollups)	10.39	6.202	6.483	73.83
N: 4765 ($h_1^* = 1.73$), *4144* ($h_2^* = 1.36$) , *3192* ($h = 1$)		(2.524)	(2.540)	(96.30)
Alcohol twice a week or more	42.82%	−9.001	−9.425	−263.22
N: 2546 ($h_1^* = 0.87$), *2125* ($h_2^* = 0.68$) , *2996* ($h = 1$)		(4.325)	(4.393)	(193.31)

Continued on next page

[25] The denominator is the impact on BP after one wave, even in the two waves lifestyle exercise.

Table 2.4: (Continued)

Dependent Variable at $t+1$	(1) Mean 1SD Below	(2) Loc Linear Triangular h_1^*	(3) Loc Linear Rectang. h_2^*	(4) Loc Quad Rectang. $h = 1SD$
Portions of fruits per day	3.27	0.648	0.702	9.17
N: 1033 ($h_1^* = 0.67$), *870* ($h_2^* = 0.52$) , *1568* ($h = 1$)		(0.410)	(0.405)	(9.14)
BMI: Body Mass Index (kg/m2)	27.86	0.587	0.743	20.47
N: 2092 ($h_1^* = 0.76$), *1679* ($h_2^* = 0.59$) , *2754* ($h = 1$)		(0.471)	(0.488)	(14.48)
Waist-to-height ratio (WHtR)	0.56	0.012	0.012	0.34
N: 1956 ($h_1^* = 0.69$), *1531* ($h_2^* = 0.54$) , *2750* ($h = 1$)		(0.007)	(0.008)	(0.22)
Obesity level 1 or above: BMI 30+	27.60%	8.144	7.668	227.71
N: 1958 ($h_1^* = 0.70$), *1532* ($h_2^* = 0.55$) , *2754* ($h = 1$)		(4.853)	(5.027)	(150.10)

Sample: Respondents aged 60 and younger at the moment of the nurse advice, who were not diagnosed with HBP or being taking medication for lowering BP.
Notes: Column (1) presents the mean of each dependent variable for those observations one standard deviation below the threshold. Columns (2) to (4) present different specifications for the trend (function $f(\cdot)$) between the outcome and systolic blood pressure. Robust SE are presented in parenthesis. Significance: * 10%, ** 5%, *** 1%.

2.4.3 Heterogeneity on the impact

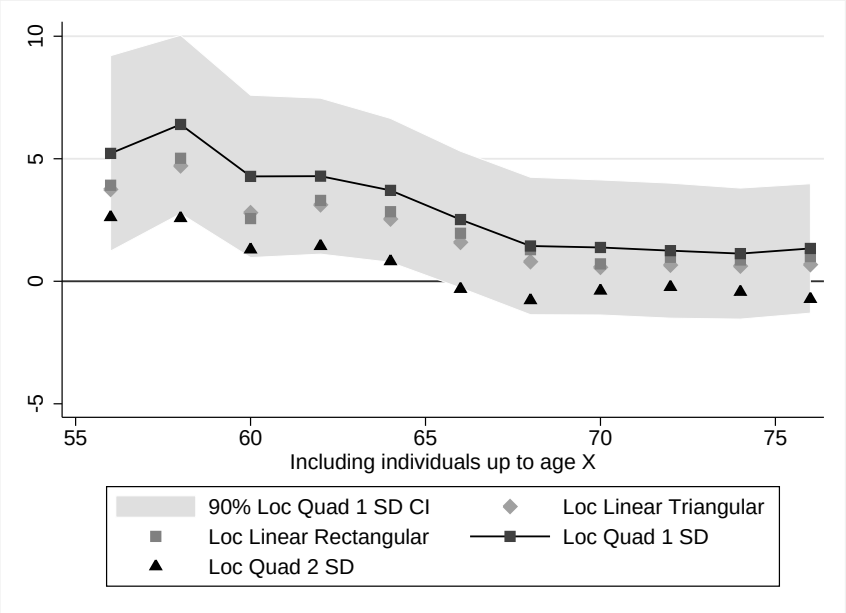
A relevant question is whether the impact is heterogeneous according to respondents' characteristics. First, the impact on BP prescription is declining with age. Figure 2.7 shows that if older individuals are included in the sample, the estimate of the discontinuity tends to 0. Second, the impact is concentrated on males with a 10-years CVD-risk above 8%.[26] Table 2.5 presents the discontinuity local linear triangular estimator for a selected group of variables. The difference with previous sections results is that the sample was stratified according to gender and CVD-risk. This reduces notoriously the sample size, resulting on larger standard errors. Differences on HBP medication, alcohol intake and fruits consumption are larger for men, all of them significant at least at 90% level. Finally, the estimates suggest that the impact is restricted to those individuals with a 10-year risk of developing a CVD of 8% or above.[27] With respect to smoking intensity, the estimated effect is

[26]CVD risk calculating using the Framingham equation D'Agostino et al. (2008). This is a standard risk calculator for individuals aged 30 to 74 without prior CVD. It involves age, gender, smoking status, total and HDL cholesterol levels, systolic BP, diabetes. For this study, while there are more accurate calculators for England population as QRISK (Hippisley-Cox et al., 2008), this method was selected for its simplicity given the available information.

[27]This category was defined on the basis of sample size, rather than clinical standards. However, a more standard 10% risk results on a similar point estimate but is not significant.

non-significant in all exercises.

Figure 2.7: Impact on BP medication estimator by age



These results suggest that the effect is zero on individuals with low overall risk of developing CVD. The fact that the effect is strong for men is likely to be related to the higher thresholds for advising respondents in the HSE (ELSA wave 0). In fact, NICE recommended drug therapy for those with a systolic BP of 160 mmHg or above (NICE, 2006, 2011). With respect to age differences, it is expected as older individuals have a higher demand for medical services, therefore the intervention should have no impact on them. Moreover, consequences of hypertension are higher between ages 40 and 70 (Chobanian et al., 2003).

Table 2.5: RDD by groups: general impact

$$Y_i = \delta(BP_i \geq 0) + f(BP_i|BP_i < 0) + f(BP_i|BP_i \geq 0) + u_i|X_i$$

RDD on systolic BP standarized around 140 mmHg. It is conditional on not been diagnosed before with HBP or being taking medication for blood pressure.

Restriction X_i	(1) HBP	(2) PILLS	(3) N CIGS	(4) ALCOHOL	(5) FRUITS
Base Result	2.98	4.28**	5.36	−7.93	0.77*
N PILLS: 3717 $(h = 1)$	(3.00)	(2.02)	(4.20)	(5.12)	(0.43)
Male					
Yes	6.09	7.92**	4.21	−15.02*	1.41**
N PILLS: 1478 $(h = 1)$	(4.70)	(3.18)	(7.00)	(8.06)	(0.64)
No	0.70	1.74	5.92	−3.53	0.20
N PILLS: 2239 $(h = 1)$	(3.86)	(2.59)	(5.21)	(6.58)	(0.57)
10 years CVD risk 8% and over					
Yes	3.34	9.01**	−12.13	−9.99	1.45**
N PILLS: 1054 $(h = 1)$	(6.14)	(4.13)	(8.18)	(9.30)	(0.65)
No	−0.75	−0.38	4.69	0.25	−0.26
N PILLS: 1540 $(h = 1)$	(5.05)	(3.19)	(4.27)	(7.99)	(0.55)

Notes: RDD on systolic BP standarized around 140 mmHg. Individuals aged 60 or younger who have not been diagnosed before with high blood pressure or any other cardiovascular related conditions. Column (1), *HPB*, presents estimates for the difference on the probability to be diagnosed with high blood pressure two years after the advice is given. In Column (2), *PILLs*, the dependent variable is the probability to be under medication for controlling blood pressure levels; in Column (3), *NCIGS*, it is the number of cigarettes consumed during the last week; in Column (4), *ALCOHOL*, the probability to have an alcoholic drink twice or more per week. Finally, Column (5) refers to the portions of fruit per day. Robust SE in parenthesis. Significance: * 10%, ** 5%, *** 1%.

2.4.4 Specification tests

Several tests were carried out in order to determine the quality of the results. The principal one was a balancing test, that is, if the 'treatment' can be considered as randomly allocated across a wide set of covariates. Table 2.6 presents the results from applying the same methodology but using as dependent variables basic demographic controls (panel A); and information on the main results' section outcomes but measured at the moment of the BP measurement (panels B, C and D). In the entire table, the only difference that is not statistically zero is non-white ethnicity and education level for some of the specifications. Nevertheless, there is neither difference in any health measurement, nor risky behaviour.

Further checks on the underlying assumptions of the regression discontinuity

are presented in Appendix 2.C.

Table 2.6: Balancing Test. RDD on covariates before receiving nurse advice

$$X_i = \delta(BP_i \geq 0) + f(BP_i|BP_i < 0) + f(BP_i|BP_i \geq 0) + u_i$$

RDD on systolic BP standarized around the cut-off.

Dependent Variable at t	(1) Mean 1SD Below	(2) Loc Linear Rectang. h_1^*	(3) Loc Linear Triangular h_2^*	(4) Loc Quad Rectang. $h = 1SD$
Panel A. Demographic Characteristics				
Age	53.80	−0.20	−0.21	0.07
N: 3744 ($h_1^* = 1.00$), *2990* ($h_2^* = 0.79$) , *3744* ($h = 1$)		(0.34)	(0.34)	(0.46)
Male	39.77%	−0.72	−2.43	0.10
N: 3343 ($h_1^* = 0.89$), *2620* ($h_2^* = 0.70$) , *3744* ($h = 1$)		(3.81)	(3.89)	(4.91)
Non white ethnicity	13.39%	4.84*	4.51*	4.66
N: 3304 ($h_1^* = 0.93$), *2767* ($h_2^* = 0.73$) , *3698* ($h = 1$)		(2.74)	(2.74)	(3.59)
Educ: Some medium qualif.	42.67%	6.23	8.17*	6.86
N: 2982 ($h_1^* = 0.78$), *2221* ($h_2^* = 0.61$) , *3733* ($h = 1$)		(4.12)	(4.29)	(4.99)
Educ: Some high level or above qualif.	34.61%	−1.46	0.72	−2.63
N: 3543 ($h_1^* = 0.95$), *2795* ($h_2^* = 0.74$) , *3733* ($h = 1$)		(3.53)	(3.62)	(4.70)
Married	76.74%	2.11	1.07	3.27
N: 4752 ($h_1^* = 1.35$), *3918* ($h_2^* = 1.06$) , *3744* ($h = 1$)		(2.73)	(2.76)	(4.22)
Panel B. Health-related Variables				
Diagnosed Major Cardiovascular Event ever (Stroke, Heart Failure, Infarction, An	1.29%	0.55	0.74	−0.14
N: 4064 ($h_1^* = 1.13$), *3333* ($h_2^* = 0.88$) , *3731* ($h = 1$)		(0.86)	(0.85)	(1.22)
Self-reported GOOD health	83.33%	1.10	1.50	1.82
N: 2990 ($h_1^* = 0.78$), *2224* ($h_2^* = 0.62$) , *3744* ($h = 1$)		(3.13)	(3.28)	(3.81)
Self-reported bad health	8.89%	−0.77	−0.21	−1.96
N: 4362 ($h_1^* = 1.21$), *3553* ($h_2^* = 0.95$) , *3744* ($h = 1$)		(1.96)	(1.97)	(2.93)
SSP: Chances to live to age 75	69.18	−0.50	−0.73	0.92
N: 2445 ($h_1^* = 1.40$), *2031* ($h_2^* = 1.10$) , *1868* ($h = 1$)		(1.83)	(1.84)	(2.78)
(D) Valid BMI - inc estimated>130kg	27.75	0.06	0.08	0.41
N: 3769 ($h_1^* = 1.06$), *3030* ($h_2^* = 0.83$) , *3603* ($h = 1$)		(0.35)	(0.35)	(0.49)
Waist-to-height ratio (WHtR)	0.57	0.00	0.00	0.00
N: 1191 ($h_1^* = 0.64$), *908* ($h_2^* = 0.50$) , *1835* ($h = 1$)		(0.01)	(0.01)	(0.01)
Overweight or above	70.00%	2.37	3.01	5.22
N: 2686 ($h_1^* = 0.78$), *2130* ($h_2^* = 0.61$) , *3603* ($h = 1$)		(3.79)	(3.95)	(4.59)
Obesity I or above	26.89%	4.46	4.12	6.72
N: 2686 ($h_1^* = 0.74$), *2130* ($h_2^* = 0.58$) , *3603* ($h = 1$)		(4.02)	(4.05)	(4.71)
Blood HDL level (mmol/l)	1.56	−0.01	−0.00	0.01
N: 2353 ($h_1^* = 0.93$), *1973* ($h_2^* = 0.73$) , *2623* ($h = 1$)		(0.04)	(0.04)	(0.05)

Continued on next page

Table 2.6: (Continued)

Dependent Variable at t	(1) Mean 1SD Below	(2) Loc Linear Rectang. h_1^*	(3) Loc Linear Triangular h_2^*	(4) Loc Quad Rectang. $h = 1SD$
Blood total cholesterol level (mmol/l)	5.95	0.01	−0.02	0.06
N: 2502 ($h_1^* = 0.98$), *1976* ($h_2^* = 0.77$) , *2627* ($h = 1$)		(0.10)	(0.10)	(0.13)
Panel C. Lifestyle				
Current smoker	18.57%	4.05	3.44	2.19
N: 4356 ($h_1^* = 1.21$), *3547* ($h_2^* = 0.95$) , *3738* ($h = 1$)		(2.70)	(2.72)	(3.94)
Cigaretes per week (0 for non-smokers, includes rollups)	6.93	−1.75	−1.26	−1.76
N: 3125 ($h_1^* = 0.95$), *2454* ($h_2^* = 0.75$) , *3288* ($h = 1$)		(3.21)	(3.18)	(4.24)
Alcohol twice a week or more	51.63%	0.03	−0.02	−2.01
N: 2064 ($h_1^* = 0.87$), *1696* ($h_2^* = 0.69$) , *2448* ($h = 1$)		(4.87)	(4.97)	(6.20)
Sedentary or low physical activity	15.60%	2.60	4.72	4.50
N: 1400 ($h_1^* = 0.75$), *1100* ($h_2^* = 0.59$) , *1878* ($h = 1$)		(4.37)	(4.49)	(5.15)
Portions of vegtables per day	6.04	0.83	0.99	1.24
N: 807 ($h_1^* = 0.83$), *613* ($h_2^* = 0.66$) , *959* ($h = 1$)		(1.07)	(1.07)	(1.30)
Portions of fruits per day	4.83	0.39	0.38	0.34
N: 1009 ($h_1^* = 1.07$), *809* ($h_2^* = 0.84$) , *963* ($h = 1$)		(0.55)	(0.60)	(0.71)
Panel D. Economic activity				
BU total weekly income (£ of May2005)	0.54	−0.00	−0.01	−0.03
N: 1836 ($h_1^* = 0.99$), *1470* ($h_2^* = 0.78$) , *1836* ($h = 1$)		(0.04)	(0.04)	(0.05)
BU total net (non-pension) wealth (1000£ of May2005)	363.46	105.21	107.27	116.08
N: 1372 ($h_1^* = 0.76$), *1080* ($h_2^* = 0.60$) , *1836* ($h = 1$)		(69.59)	(78.32)	(84.96)
Hours of work all jobs (employed or self employed)	36.65	−0.48	−0.92	−1.39
N: 970 ($h_1^* = 0.69$), *766* ($h_2^* = 0.54$) , *1414* ($h = 1$)		(2.02)	(2.07)	(2.38)

Sample: Respondents aged 60 or younger at the moment of the nurse advice, who were not diagnosed with HBP or being taking medication for lowering BP.
Notes: Column (1) presents the mean of each dependent variable for those observations one standard deviation below the threshold. Columns (2) to (4) present different specifications for the trend (function $f(\cdot)$) between the outcome and systolic blood pressure. Robust SE are presented in parenthesis. Significance: * 10%, ** 5%, *** 1%.

2.5 Conclusion

This chapter analysed the impact of a health check that advices to visit a family doctor for those individuals with blood pressure about certain threshold. Before continuing with the analysis, it is important to be clear on the limitations of this analysis. It is restricted for those individuals with mildly-raised blood pressure.

This is relevant in terms of policy analysis as they are the most likely to be affected by health checks for that specific condition. However, what would happen with other conditions in terms of behavioural response, with higher risks of further complications, might be different.

The first main question of this chapter is the impact of the advice on early detection of hypertension. Results show a large and significant impact of the advice on the probability to be under BP medication. However, it cannot be rejected that probability of being diagnosed with hypertension is the same for those people above and below the threshold. The present analysis cannot distinguish between two mechanisms. First, a temporal positive impact on both detection and medication prescription. Second, that the survey intervention might have increased the odds of choosing a medication-based HBP treatment by local family doctors. This is due to the lack on information about demand of health-care services in ELSA for waves 0 to 5.

The second element to discuss is the impact on lifestyle. Guidelines suggest a lifestyle intervention that curve smoking, bad dietary habits and heavy alcohol drinking. A clear impact in this direction is found for drinking frequency, and evidence for improved fruit portions' consumption is also found. On the other hand, there is evidence of risk compensation in smoking an caloric intake (reflected on obesity). In contrast, Steptoe and McMunn (2009) has previously shown that hypertense individuals in ELSA smoke less and drink more than non-hypertense individuals. This chapter, focused on those individuals who are in the borderline of the diagnosis, finds that the effect of the advice is precisely to reduce heavy drinking patterns.

Finally, whether the advice had a positive effect on respondents' health after nearly four years is an unresolved question. None of the effects on blood pressure, cholesterol levels, or sugar in the blood are statistically different from zero. However, this might be due to the limited sample size

These findings complement Kaestner et al. (2014) results on the use of statins, where an increase on obesity was found but at the same time physical activity increased for men. Also, such results can be contrasted with Fichera et al. (2016) who found that an increase in the quality of medical services in England im-

proved behaviour, including smoking and heavy drinking. Pooling together this evidence, risk compensation and complementary health-investment mechanisms are likely to be relevant elements to consider in preventive care policies. However, such responses are likely to be heterogeneous and hard to extrapolate to general circumstances.

2.A Sample selection

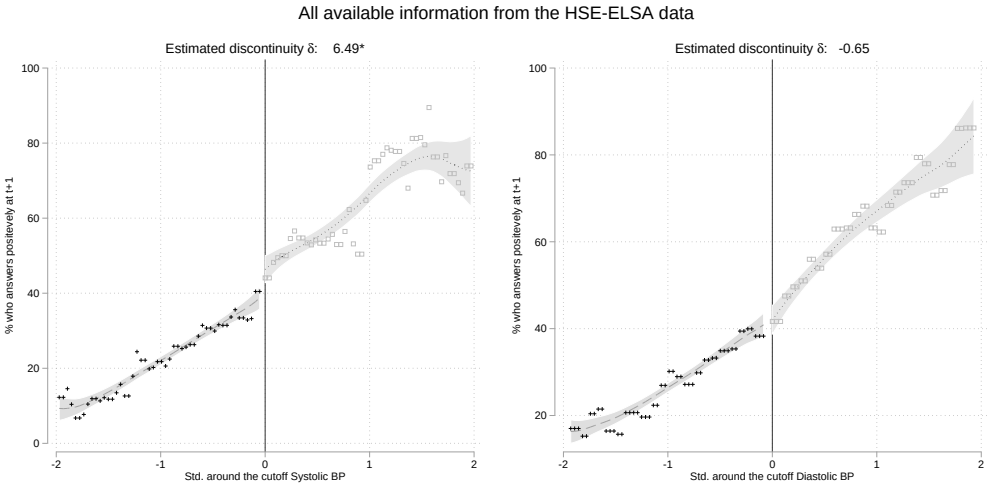
The analysis is carried out a subset of all the available data from the HSE-ELSA data. Individuals who were not diagnosed with HBP and were not taking BP lowering medication[28] are selected. Even though it drastically reduced the sample size, such restriction avoids potential biases as individuals who are above the threshold are more likely to report to be diagnosed with HBP before the nurse visit.

Using all the data, the jump estimator for the systolic BP is apx. 6.5 pp. while for the diastolic it is -0.6pp. Only the first one is statistically different from 0. However, Figure 2.9 shows that there might be a potential bias. Instead of reporting the proportion of those who are reported to have HBP at the following wave, the outcome is measured at the baseline. That is, what they reported before the nurses visited them. The same pattern is present: a jump of 5 pp. for systolic BP and of 1.1 pp. for diastolic BP. While it is not possible to reject the hypothesis of them equal to zero under the quadratic specification, the estimate for the systolic cutoff is different from zero under a triangular kernel. If we consider BP medication instead of the self-reported diagnosis, such issue is not present (Figure 2.10).

As a result, in order to avoid the potential bias provided by this discontinuity, I restrict the sample only to the new cases at the cost of larger standard errors.

[28] On the ELSA, everyone who is asked about BP medication reports to be diagnosed with HBP by design of the survey. That is not the case for the HSE, where the analysis of medication is much more detailed.

Figure 2.8: Nurse Advice and self-report of HBP at the following-wave

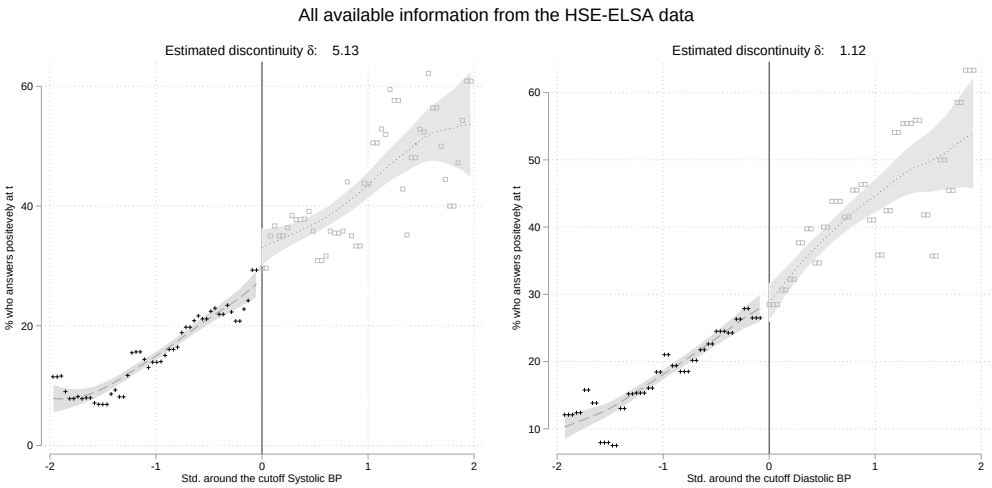


Sample: Individuals aged 60 or younger from the HSE-ELSA data.

Notes: Calculations using a quadratic function within 1 standard deviation of the cutoff. A 90% CI is presented.

Significance level: *90%, ** 95%, *** 99%

Figure 2.9: Nurse Advice and self-report of HBP at the same wave (balance test)

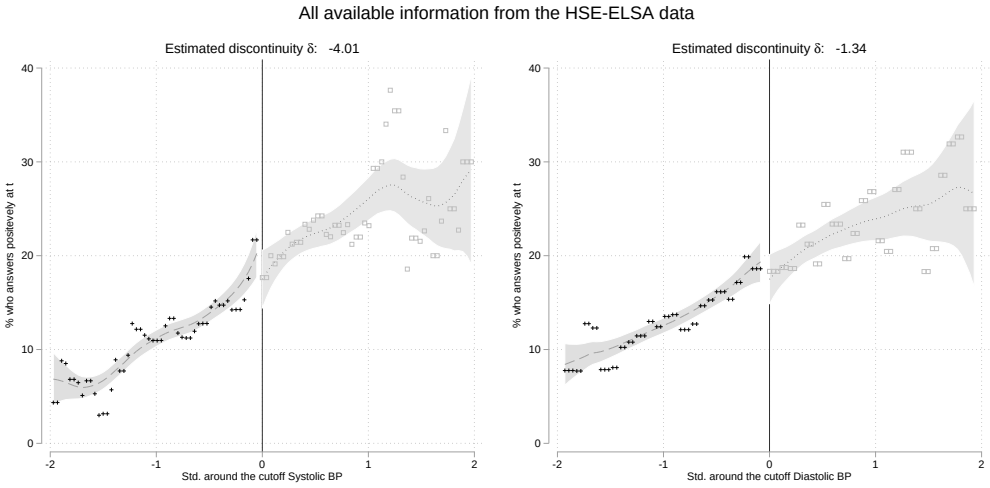


Sample: Individuals aged 60 or younger from HSE-ELSA data.

Notes: Calculations using a quadratic function within 1 standard deviation of the cutoff. A 90% CI is presented.

Significance level: *90%, ** 95%, *** 99%

Figure 2.10: Nurse Advice and BP medication in the same wave (balance test)



Sample: Individuals aged 60 or younger from HSE-ELSA data.

Notes: Calculations using a quadratic function within 1 standard deviation of the cutoff. A 90% CI is presented.

Significance level: *90%, ** 95%, *** 99%

2.B Alternative Specifications

There are many ways to implement the strategy. The general idea is to choose the parameters that minimize the distance between the observed outcome and the prediction from a model m, giving different weights K to each observation i as shown in Equation 2.3. Such a model, characterized by a set of parameters α, δ and restrictions, takes into account the relation between Y_{t+1} and the BP index BP_t^c measured at wave t and standardized according to the relevant advice cut-off c (it changes according to the year of the survey, gender and age as described before). The weights K are assigned using some arbitrary rule based on the forcing variable BP_t^c. The most simple specification gives equal importance to all observations between 0 and h standard deviations, and disregards the remaining data (rectangular kernel). A common alternative is the triangular kernel, where the relevance of observations decays linearly. For the main results, the value of h is determined following the rule of Imbens and Kalyanaraman (2011).

$$\min_{\{\delta,\alpha\}} \sum_{i=1}^{N} K(BP_{it}^c/h)(Y_{i,t+s} - m(\delta,\alpha,BP_{i,t}^c))^2 \quad ,s \in \{1,2\} \tag{2.3}$$

The model m specifies how BP and the outcomes are related, and in particular the parameter of interest, the difference δ between being above or below the cut-off. The relationship can be allowed to be different above and below the threshold as shown in Equation 2.4. In this expression, α is allowed to be specific above and below the $BP^c_{it} \geq 0$, a condition defined by the dichotomous variables W_{it}. The implementation was carried out following Nichols (2012), under different bandwidths and specifications for f.

$$m(\delta, \alpha, BP^c_{i,t}) = \delta W_{it} + \alpha_0 + f_l(\alpha_l, BP^c_{i,t}|W_{it} = 0) + f_r(\alpha_r, BP^c_{i,t}|W_{it} = 1) \tag{2.4}$$

Results are presented using three specifications for $f(\cdot)$:

- Local linear regressions: $f_l = \alpha^l_1 BP^c_{i,t}$ and $f_r = \alpha^r_1 BP^c_{i,t}$. Triangular and rectangular weights are considered. The difference between them is that triangular weights give more importance to the observations close to the threshold.

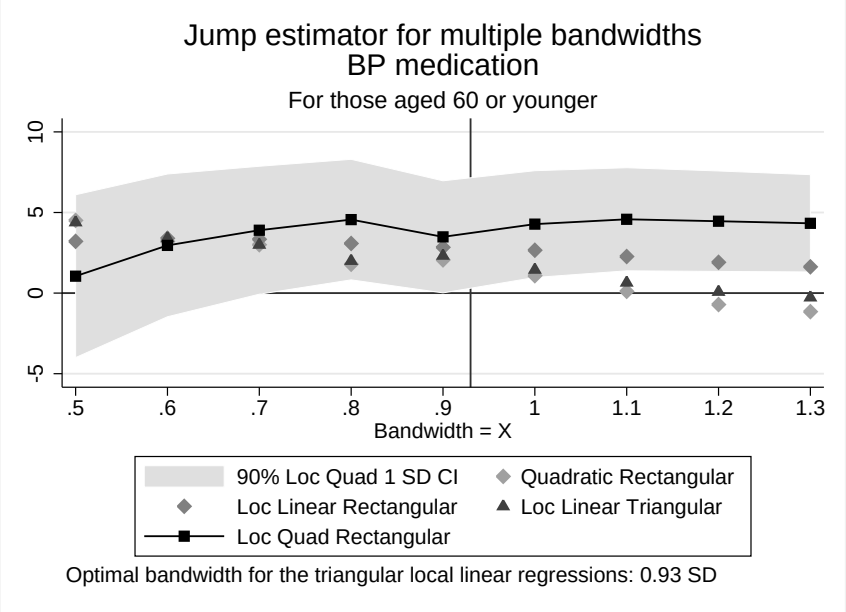
- Local quadratic function: using rectangular weights, $f_l = \alpha^l_1 BP^c_{i,t} + \alpha^l_2 {BP^c_{i,t}}^2$ and $f_r = \alpha^r_1 BP^c_{i,t} + \alpha^r_2 {BP^c_{i,t}}^2$.

2.C Further Robustness Checks

An usual concern with non-parametric estimators is their potential dependence on *ad hoc* parameters. In this particular case, the jump estimator might be very sensitive to the 'bandwidth' selection. Optimal selection procedures like the one presented by Imbens and Kalyanaraman (2011) help to determine a proper value for it. Nevertheless, the question on the sensitivity is still present. Figure 2.11 present the case of usage of BP-lowering medication for different values of the parameter, the horizontal axis, across multiple specifications. As a reference, the optimal bandwidth is highlighted by a vertical line and a confidence interval of 95%, which corresponds to the local linear triangular kernel. We can observe that the estimated values of the jump change notoriously according to the underlying assumption but results are relatively stable close to the optimal one.

One concern could be that the nurses registered a value above 140 for individuals who had BP levels slightly below it. This would have been translated into a

Figure 2.11: Jump estimator for multiple bandwidths (BP medication)



discontinuity in terms of the density at the threshold. Though the reasons for this potential manipulation are not clear, McCrary (2008) test can asses if that is the case. The test creates a histogram of the BP and defines the mid-point of each bin as the dependent variable of the RDD in a non-parametric way. Table 2.8 presents the estimator of the jump θ for different bin sizes around the optimal one. There is no evidence of a discontinuity in the density as such point.

Table 2.7: RDD sample restrictions

$$X_i = \delta(BP_i \geq 0) + f(BP_i|BP_i < 0) + f(BP_i|BP_i \geq 0) + u_i|BP_i \notin \Omega$$

RDD on systolic BP standarized around 140 mmHgDependent variable: whether diagnosed with HBP in the follow-up, conditional on not been diagnosed before with HBP or being taking medication for blood pressure.

Restriction Ω	(1) Quadratic 1 SD	(2) Loc Linear Rectang h^*	(3) Loc Linear Triangular h^*	(4) Local Quad 1 SD	(5) Local Quad 2 SD
Without restriction	1.76	4.19**	4.27**	5.59***	2.50
N: 2483 ($h^* = 0.59$), 4195 ($h = 1$) , 6562 ($h = 2$)	(1.48)	(1.97)	(1.73)	(2.02)	(1.60)
Taking out 139 mmHg†	2.11	4.77***	3.88***	7.10***	3.04*
N: 3657 ($h^* = 0.90$), 4088 ($h = 1$) , 6455 ($h = 2$)	(1.51)	(1.53)	(1.42)	(2.00)	(1.59)
Taking out 140 mmHg†	0.69	2.91	3.06*	5.23**	1.00

Continued on next page

Table 2.7: (Continued)

Restriction Ω	(1) Quadratic 1 SD	(2) Loc Linear Rectang h^*	(3) Loc Linear Triangular h^*	(4) Local Quad 1 SD	(5) Local Quad 2 SD
N: 2621 ($h^* = 0.65$), *4084* ($h = 1$) , *6451* ($h = 2$)	(1.61)	(2.04)	(1.86)	(2.49)	(1.88)
Taking out 141 mmHg†	2.02	5.12**	5.13**	6.63***	2.84
N: 2163 ($h^* = 0.56$), *4117* ($h = 1$) , *6484* ($h = 2$)	(1.62)	(2.30)	(2.05)	(2.29)	(1.80)
Taking out 139-141 mmHg†	1.19	6.15**	4.91**	9.10***	1.58
N: 2187 ($h^* = 0.59$), *3899* ($h = 1$) , *6266* ($h = 2$)	(1.85)	(2.81)	(2.46)	(3.35)	(2.27)

† For males aged 50 or over in wave 0, the values are 159, 160 and 161 mmHg. Robust SE in parenthesis. Significance: * 10%, ** 5%, *** 1%.

Table 2.8: McCrary Test for those aged 60 or younger

	Bin Size (mmHg)									
	0.279	0.335	0.391	0.447	0.503	**0.558** †	0.614	0.670	0.726	0.782
θ	0.06	0.06	0.08	0.08	0.09	0.12	0.13	0.14	0.16*	0.17*
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)

SE in parenthesis. Significance: * 10%, ** 5%, *** 1%. McCrary test on the continuity of the density at the threshold. Triangular kernels are fitted on the means of the bins of a particular bin size. The optimal bin size (†) and, the bandwidths are chosen following McCrary implementation of the test.

A further test consist of considering placebo discontinuities. In other words, given the index of standardized systolic BP, it is possible to perform the exercise but assuming that the jump is at values different from 0. Figure 2.12 shows that the only values in which a discontinuity is observed are those around 0.

A final test consist on including controls in our regressions. Normally they should not affect the results in any way. Table 2.9 presents such a regression for the case of medication. Essentially, there are no noticeable changes; the significance is only affected when the sample size is reduced due to the availability of information.

Figure 2.12: Placebo jumps over Std. Systolic BP index (BP medication)

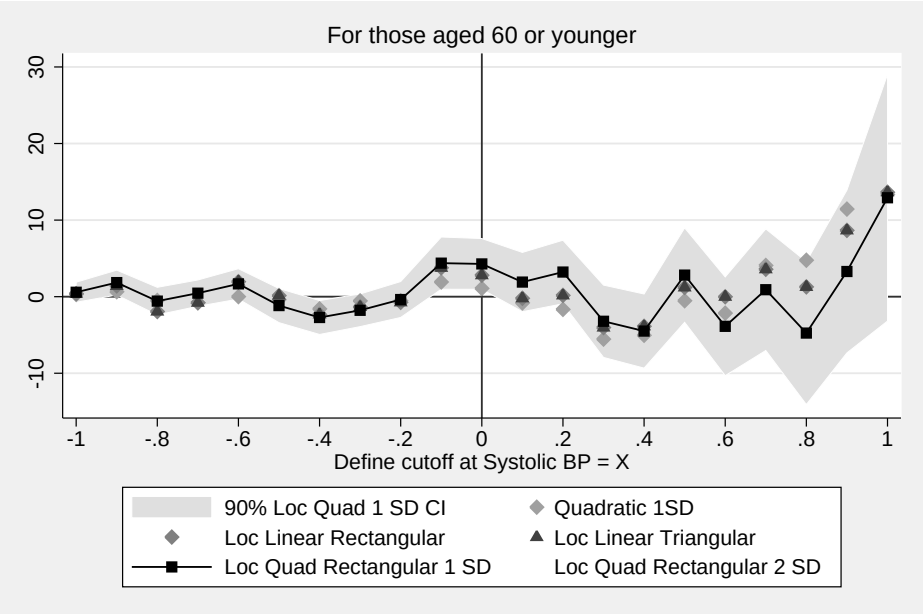


Table 2.9: RDD including covariates for those aged 60 or younger: Takes BP medication

$$Y_i = \delta(BP_i \geq 0) + f(BP_i|BP_i < 0) + f(BP_i|BP_i \geq 0) + X_i\beta + u_i$$

RDD on systolic BP standarized around the value. Dependent variable: Takes BP medication, conditional on not been diagnosed before with HBP or being taking medication for blood pressure.

Dependent Variable at $t+1$	(1) Mean	(2) Loc Linear Rectang. h_1^*	(3) Loc Linear Triangular h_2^*	(4) Loc Quad Rectang. $h = 1SD$
Without controls	3.36%	2.80*	2.56*	4.28**
N: 3322 ($h_1^* = 0.93$), *2784* ($h_2^* = 0.73$) , *3717* ($h = 1$)		(1.54)	(1.51)	(2.02)
Demographic	3.37%	2.67*	2.53*	4.21**
N: 3313 ($h_1^* = 0.93$), *2777* ($h_2^* = 0.73$) , *3706* ($h = 1$)		(1.52)	(1.51)	(2.01)
+ Health and Behaviour	3.32%	2.88*	2.70*	4.25**
N: 3176 ($h_1^* = 0.93$), *2656* ($h_2^* = 0.73$) , *3559* ($h = 1$)		(1.57)	(1.56)	(2.09)
+ Health and Behaviour (extended)	3.42%	2.27	1.94	2.94
N: 1263 ($h_1^* = 0.93$), *1043* ($h_2^* = 0.73$) , *1404* ($h = 1$)		(2.26)	(2.37)	(2.90)

Robust SE in parenthesis. Significance: * 10%, ** 5%, *** 1%. Demographic Characteristics: Age, gender, ethnicity, education level, marital status. Health and Behaviour: to report to be on bad health, BMI, any parental CVD-related death, smoking status. Extended variables: Cholesterol (Total and HDL), framingham CVD risk, alcohol and physical activity levels.

Chapter 3

On the economic value of preventive care: a life-cycle model perspective

3.1 Introduction

Health is an essential determinant of individuals' financial choices during their life cycle, as stated in the classic Grossman Health Capital model (Grossman, 1972). This is the case even if individuals' utility does not depend directly on health, as there are alternative mechanisms such as absence of from work due to disability, or out-of-pocket expenditures. For these considerations it is relevant to incorporate the economic consequences of health treatments when ranking the allocations of health resources. This is of particular interest when consequences are not immediate, such as for preventive care interventions. This paper introduces a framework that improves traditional economic evaluation techniques, such as cost-benefit and cost-effectiveness analysis, by modelling both health and economic consequences in a dynamic structural life-cycle model.

In general terms, the framework presented in this paper expands on standard Markov disease progression models, where health is categorised into states that account for the severity of illness, in two directions. First, it incorporates an index of risk of disease progression, which allows for heterogeneity in the transition probabilities between the states. Essentially, the probability of transiting into worse states of health is higher when the value of the index is higher. Thus, it is possible to model preventive care interventions as either exogenous transitory or permanent shocks to the value of the index. Second, the health model is complemented by a life-cycle model in which individuals, whose utility might be

affected by health status, are able to smooth consumption via savings and labour supply. This structure can be adapted to any chronic disease where early stage treatments are available, and can be estimated with datasets that already exist in many countries.

As a concrete example, I consider the role of a preventive care innovation, a class of medications called statins, which reduces the odds of serious health complications such as heart attacks or strokes, which might result in disabilities that inhibit working and increase the risk of death. This is relevant as improvements in medical care for cardiovascular disease (CVD) have been the main source of improvement in healthy life expectancy in the last 20 years (Chernew et al., 2016). First, the progression of CVDs is modelled with four health states: (i) having no signs of any related condition; (ii) having been diagnosed with mild conditions such as hypertension, which increase the odds of developing cardiovascular events; (iii) to have survive to such events; and (iv) death. In this setting, the empirical counterpart of the risk index is a summary variable that involves biomarkers that are known to be good predictors of future CVD complications such as blood pressure and cholesterol levels. This formulation is a contribution to the efforts directed towards modelling health investments (Kaestner et al., 2014; Hai and Heckman, 2015), when there is a clear distinction between the prevention and treatment of diseases (Ozkan, 2014). The model also includes a rule for the adoption of the treatment that incorporates a variation across local providers of health care services. This allows for a better understanding of the benefits of policies aimed at improving care in a population.

The model considers how health shocks affect individual labour supply and savings conditional on social security arrangements. This is built the on-growing literature that considers the role of health in shaping economic choices for individuals close to retirement age (Palumbo, 1999; French, 2005; Halliday et al., 2015; De Nardi et al., 2010; French and Jones, 2011; Yogo, 2016). Poor health derived from such complications has direct effects on wages, via a productivity effect, and on assets. It also introduces a penalty on the cost of working in terms of time resources of individuals. This, as in the original Grossman (1972) health capital model, has a direct effect on utility even if quality of health is not considered as a

separate commodity. By acknowledging these channels, it is possible to take into account selection into labour market participation derived by the onset of health complications (French, 2005). Also, it is possible to incorporate the economic gains of a medical treatment on top of conventional cost-benefit analysis of health interventions. In order to do this, a compensating variation is obtained by calculating the expected life-time utility from a counterfactual scenario in which the technology is not available, but a lump sum is given in compensation. Thus, the estimate of the value is the amount of the lump sum that minimises the difference in expected utilities from both the observed and counterfactual scenarios.

The framework presented below contributes to the literature devoted to assessing the value of health, medical innovations and policies while taking into account their economic implications. Murphy and Topel (2006) considers the social welfare value of improvements in longevity in the US over the 20th century using a life-cycle model for computing willingness-to-pay, but without considering reduced productivity or disabilities. The same consideration applies to Hall and Jones (2007), which is devoted to understanding rising medical expenditure. Papageorge (2015) introduced a major contribution by considering the labour market implications of a new treatment for HIV/AIDS. Moreover, he identifies willingness-to-pay for a treatment based on revealed preferences instead of relying on stated preferences or indirect approaches. While Papageorge's procedure is ideal, the availability of revealed preferences information is scarce as in most setups, consumer prices, if they exist, are not informative given that they suffer from a substantial impact of government interventions. For instance, work by Murphy and Topel (2006), Nordhaus (2003) or Hall and Jones (2007) matched the willingness-to-pay for extending life derived from their models with estimates from labour literature based on the wage premium from jobs according to risk or death. The procedure of this study introduces a novel alternative by deriving the *value of a statistical life* (VSL)[1] from the estimated bequest motive that accounts for assets' accumulation at the end of life (De Nardi, 2004; French, 2005). Thus the identification of the compensating variation in the framework derived in this

[1]This is the willingness-to-pay for a reduction of 1 unit on risk of death, normally extrapolated from calculations based on small variations on such a risk. See Murphy and Topel (2003) for an introduction of the concept in on the context of life-cycle models.

paper comes from the observed losses of people who suffered a CVD event, and from the value of assets of the elderly.

Another contribution of this project is that it improves the cost-effectiveness analysis of health interventions. McIntosh (2006) and Borghi (2008) mention that very few of the stated preference willingness-to-pay studies aggregate their results for conducting a cost-benefit analysis that will inform policy decisions. Instead, productivity gains and cost savings are used for this purpose, based on the idea that decision makers might have objectives other than 'making Pareto improvements' (Culyer, 1989; Brouwer and Koopmanschap, 2000). In fact, in the UK technology adoption decisions are explicitly made based on the cost of increasing 1 quality adjusted life year (QALY),[2] regardless of its impact on other elements that might affect individuals' utility. This model is able to improve this methodology by considering the effect of health on labour supply and then on labour derived income. This provides an additional variable to consider when taking decisions and allows for the calculation of potential resources that might revert back into the health care system via taxation. Standard cost-effectiveness analysis is produced by simulating the lives of a set of individuals under two scenarios, one with and one without treatment, and then adding up the total QALYs and the total costs for the reference population. Then, the additional QALYs are divided by the total amount of costs. With the framework presented above, on top of simulating health scenarios, it is possible to simulate consumption, savings, labour supply and income of individuals.

Another key difference from the literature is that the model is estimated using data from England, where out-of-pocket medical expenditures is low due to the presence of a publicly funded health care system. For this reason, medical expenses and health insurance are not considered. This allows for a lower bound on the potential value of preventive care innovations for those institutional setups, as such elements are crucial determinants of savings in old age as shown by Blau and Gilleskie (2008), De Nardi et al. (2010), and French and Jones (2011). The parameters are estimated using the method of simulated moments (MSM) with informa-

[2]It is a simple measure that combines quantity and quality of life. In it, for a given health state a *health utility* is assigned, which is a number that represents quality relative to a life without health problems (Phillips and Thompson, 2001).

tion from the *English Longitudinal Study of Ageing* (ELSA). This survey is similar to the *Health and Retirement Study* (HRS) in the US, and is devoted to improving the understanding of health, well-being and general economic circumstances of individuals aged 50 and older since 2002. It not only includes self-reported diagnoses of several chronic diseases or conditions, but also involves detailed information on blood pressure, cholesterol levels and other variables that provide an objective picture of the state of the current and future health of an individual. This dataset is linked with characteristics of the local medical services based on the post-code of residence, which provides exogenous variation to the availability of preventive care treatment.

The estimated model captures retirement and labour and non-labour supply patterns, and particularly, the impact of CVD onset, which is allowed to differ by gender and education level. The required compensation for removing statins therapy for primary prevention while keeping the same expected utility is on average £5300. This figure is 12% higher than the value that would be obtained if in the calculation of benefits we only consider a willingness-to-pay of £23.000 per QALY gained (Shiroiwa et al., 2010), which is between the thresholds for cost-effectiveness considered by the NHS (£20 to £30 thousand).

This result is based on productivity gains, extra leisure, and a reduced risk of death (the model predicts a VSL of £1.05 million). At the population level, aggregating such compensations provides the value of the drug: £79 billion by 2005. This drug, by NHS standards, is cost-effective as it costs £4641 to gain one QALY. As an additional benefit, labour-income is increased by £684 per year with such an investment.

The model also provides estimates for three crucial elements linked to the value of a drug. First, it allows for understanding how the value varies with respect to the efficiency of the drug. In the specific example, doubling the ability of statins to reduce the CVD-risk index implies an increase of 69% in its value. Second, the role of policies for increasing the availability of the treatment for whoever needs it can be understood, especially since the estimation relies on a variation in the probability of being on medication coming both from time and local family doctors' characteristics. An increase of 3 pp. in the odds of a prescription implies

an increase of 24.6% in the value of the drug (an elasticity of 0.59).

A final exercise shows how the value of the medication depends on the characteristics of institutional arrangements non-directly related to health. Given that there was an increase in the state pension age for women that covers some of the cohorts involved in ELSA, a natural exercise is to assess how much the value of the drug is affected by this policy change. For women with the characteristics observed in 2004, the model predicts an increase of 59.26% in participation in the labour market between the ages of 60 and 65 if they were exposed to such an increase in the state pension age. In terms of the drug, there is hardly any change in the overall value of the drug in terms of welfare, but there is an increase of 22 times in the labour gains per QALY gained.

After this introduction, a brief discussion of statins prescription in the UK is presented in Section 3.2 in order to provide some background for the case study. Next, the model is presented in Section 3.3, and it is followed by the compensating variation calculation in Section 3.4 and the cost-effectiveness discussion in Section 3.5. This is followed by the empirical component of the paper, which involves the procedure for the structural estimation of the model in Section 3.6, and the data used for this in Section 3.7. Finally, the results for the fit of the model and the value of statins are presented in Section 3.8, and Section 3.9 concludes.

3.2 Statins and prevention of cardiovascular diseases in the UK

Cardiovascular disease (CVD) is a heavy and growing burden for health care systems worldwide, making their prevention an important objective of public policy. Murphy and Topel (2003) calculated the value of a permanent 10% reduction in death rates due to major CVD events of around $5 trillion dollars of 1996. As a result, the pharmaceutical industry has been developing therapies aimed at reducing the risk of the onset of CVD such as statins. Before introducing the model, this section presents a brief introduction to the role of this medication in the prevention of CVDs, and how this therapy was adopted in the UK.

As described by Cutler et al. (2003), high cholesterol started to be pointed out as a risk factor for CVDs in the late fifties, and suggestive evidence and public

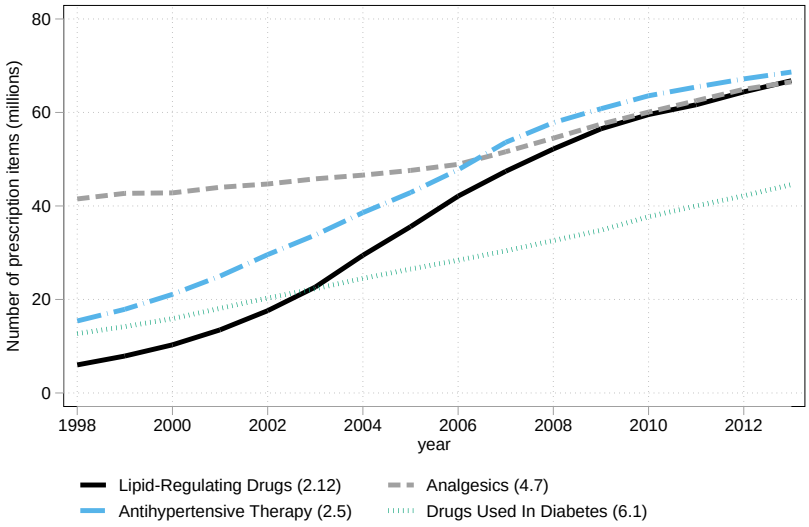
awareness built up until such a link was confirmed in the early eighties. This was translated into both a radical change in nutritional habits, but also into a race for developing safe and effective drugs for reducing cholesterol. Statins, which inhibit the production of LDL cholesterol by the liver, were introduced for the first time with the approval of Lovastatin (Tobert, 2003). Other molecules followed during the next decade. While they eventually became blockbusters for their developers, statins development and adoption involved several controversies about potential side effects, and about their effectiveness for reducing mortality.

In the 2000s in the UK, lipid-lowering drugs - mostly statins - prescriptions grew notoriously, specially if compared with the growth of well-established drugs as analgesics or anti-hypertensive therapies which have similar target population (see Figure 3.1). Between 2002 and 2012, there was an increase of 266% in the prescription of statins, which was the largest for any British National Formulary (BNF) section (HSCIC, 2013). This is similar in other countries like the US as discussed by Chernew et al. (2016). Since their initial approval, it was already clear that statins should be prescribed for secondary prevention of CVDs. In other words, use them in order to reduce the risk of repetition of CVD events on those patients who already suffered one. However, there was still a debate centred on whether they should be prescribed to general population who have no prior history of CVDs (primary prevention). As discussed by Tobert (2003), this last controversy was solved by the *Heart Protection Study* (Group et al., 2002), one of the largest medical trials ever.

Statins for primary prevention are prescribed by family doctors, also known as general practitioners (GPs). They are essential for access to this type of medication due to three main characteristics. First, given that GPs are the gatekeepers of the public health system, almost all non-urgent health care service starts with them. Second, individuals can choose a GP only if they live within a geographical zone defined by the practice where the physician is register (catchment area). They are appointed by regional administrative bodies called *Primary Care Trusts*[3] (PCT), which act as commissioners. By law, GPs are free to prescribe according

[3]PCTs have changed through the years. Prior to 2006/07, there where 303 of them but they were merged into 152 and most of them remained stable for the following years. A major reorganization took place in 2013, where PCTs were transformed into 211 Clinical Commissioning Groups (CCG).

Figure 3.1: Lipid-lowering medication prescriptions in England



Source: *Prescription Cost Analysis* (1998-2001) and *Prescriptions Dispensed in the Community* (2002-2012) for England by BNF Section, published by HSCIC.

to their criteria receiving suggestions by the guidelines provided by the *National Institute for Health and Clinical Excellence* (NICE). This institution is responsible in the UK for technical advice and care standards at primary care level. By 2006, NICE issued a technology appraisal suggesting Statins for primary prevention of CVDs (NICE, 2006).

Use of this medication is rewarded under the *Quality and Outcomes Framework* (QOF). This is a pay for performance scheme designed to improve quality of family doctors services in the UK.[4] Under this system, practices gain points for achieving certain targets, which are translated into cash for the practice every year. These goals range from administrative registries to precise clinical measures of people with a particular disease, and their achievement was monitored by their corresponding PCT. Since its introduction in 2004, QOF have rewarded GPs for keeping controlled the cholesterol level of their patients who have history of CVD or diabetes. In 2009, goals related to primary prevention were introduced for the first time. It incentivised doctors for assessing the risk of developing CVDs following a standard procedure for those individuals recently diagnosed with high blood

[4]How much and what is paid is negotiated between the *British Medical Association* (physicians union) and the National Health Service (NHS) every year, with suggestions made by NICE. Prior to QOF, GPs were paid only according to the population size under their care and their length of service within the NHS.

pressure. Such risk calculators are directly related to NICE guidelines for the use of statins. In 2011, usage of the drug became explicit for secondary prevention, and the same happened for primary prevention in 2013.

Assessing the impact of QOF on health is particularly challenging as there is no simple counter-factual (Gillam and Steel, 2013), but there is evidence of some modest gains when comparing incentivised with non-incentivised measures (Doran et al., 2011). One of the main goals of the QOF is to homogenize primary care services. In general, there was a gradient on performance with respect to deprivation which narrowed down after the first years (Dixon et al., 2010). In particular, there is evidence that the introduction of this policy in 2004 reduced heterogeneity in access to statins for secondary prevention of CVD (CQC, 2009), or even on general preventive measures as blood pressure monitoring (Ashworth et al., 2008). Apart from deprivation, other characteristics associated with performance on the scheme are related to the number of patients registered on the practice (Dixon et al., 2010), or the number of GPs working on it (Kelly and Stoye, 2014). Small practices typically underperformed during the first year of the QOF in several clinical indicators, but differences narrowed down in the following years (Doran et al., 2010; Ashworth et al., 2011).

3.3 Life-cycle model with health

This section develops the model which relates economic activity with health progression. Individuals life is modelled from age 52 until their death, which will come no-later than 100 years old. In the time being, health deteriorates affecting the probability to die as well as the trade-off of participating or not in the labour market. Such decision and the amount of resources to be saved or borrowed from the future are the choice variables available. With them, individuals maximize their expected utility conditional on their resources.

The model is organised on periods that cover 2 years of age ($t = 1, ..., T = 25$), following the data collection interval, and current health is discretised on four states ($S = 1, ..., 4$) including death. Health states cover progression of CVD diseases, allowing us to explore the role of statins. In this model calendar years (w) are not relevant for economic choices, however they do play a role determining the odds of getting a statin prescription. Apart from wealth, individuals are allowed to

be heterogenous on the risk of developing CVDs and the potential labour income that they would obtain if they work. Moreover, there is permanent heterogeneity on gender $g \in M, F$[5] and an index based on education level $EDUC \in 0, 0.5, 1$.[6]

In order to describe the model, I will start by explaining the components related to health progression. After this, I will continue with the choices and restrictions available by agents every period. This is followed by the optimization problem and how it is solved.

3.3.1 Health Progression

As described before, there are two different types of health in this model. First, a present *CVD-related health status*, discretised in the four states S described in the profiles section, which directly affects the utility function. Second, continuous CVD-risk index H, based on observed biomarkers, that affects the transition between these states. Essentially, by controlling the progression of H it is possible to diminish the odds of transition into states that reduce utility of individuals.

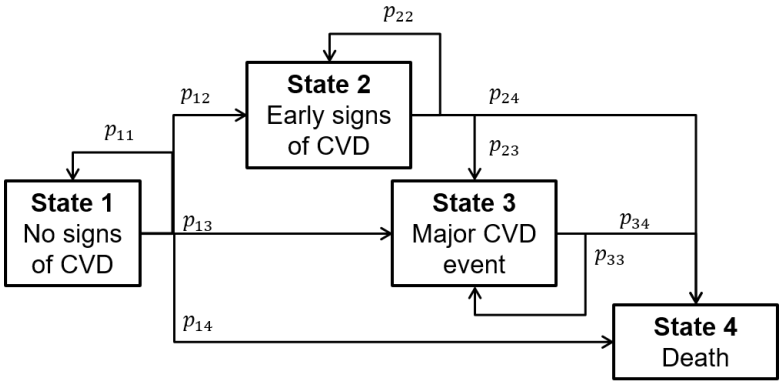
3.3.1.1 Markov Model

The are four health states. First, individuals do not present signs of any cardiovascular condition ($S = 1$). This might change, and they might present early signs denoted by persistent levels of high blood pressure, cholesterol or sugar in their blood ($S = 2$). Those are chronic conditions that can be managed, but not reversed. Additionally, they indicate that for them the risk of suffering a CVD event. This state ($S = 3$) makes it more difficult to work, as we will see in the following subsection, and drastically increase the odds of death.

[5] In the data,45.6% of the individuals ever observed from waves 1 to 6 are men.

[6] Level 0 represents no formal education, 31%(40.5%) of (fe)males. Level 0.5 covers from some formal education up to high school (up to NVQ3/GCE A level) or a foreign degree, 36.4%(38%) of (fe)males. Finally, level 1 is given to individuals with any tertiary education, 32.4% (21.5%) of (fe)males.

Figure 3.2: Health states and disease progression



A Markov process describes the transition between the four health states as described in Figure 3.2. First, it is assumed that transitions cannot be reverted, so individuals' health can only deteriorate[7]. Individuals can suffer a CVD event even without the diagnosis of milder conditions, or they can die without suffering this type of events. These restrictions simplify the transition between states at a given period t into the matrix $P_{t,t+1}$ presented in Equation 3.1.

$$P_{t,t+1} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ 0 & p_{22} & p_{23} & p_{24} \\ 0 & 0 & p_{33} & p_{34} \\ 0 & 0 & 0 & p_{44} \end{pmatrix} \tag{3.1}$$

Transition probabilities in $P_{t,t+1}$ are modelled using a multinomial logit structure. In this way it is possible to allow for different probabilities according to age t, gender g, education level index (EDUC), and CVD-risk index H but with a reduced number of parameters to estimate. In state 1, the process is governed by the three latent indexes (xb2,xb3,xb4) presented in Equation 3.2, and four extreme-value type I shocks (one per state) denoted by ξ_{it}^{s}. In state 2, only the last two of the latent indexes (xb3,xb4) and three of the shocks are valid. This is because the no-reversibility assumption stated above. Consequently, in state 3 only the last latent index (xb4) and two of the shocks are relevant.

[7]In epidemiological and health economics literature there might be several alternative CVD states according to the type of event, the number of years after it, or the secondary onset of another event. For simplicity, and for data restrictions, I will assume just one state, without any further differentiation.

$$\text{xb2}(t,H_t) = \tau_{21} * \text{age}_{it} + \tau_{22} * \text{age}_{it}^2 + \tau_{23} * \exp(H_{it}) + \tau_{24} * EDUC_i + \tau_{g;25} \tag{3.2}$$

$$\text{xb3}(t,H_t,S_t = s \in \{1,2\}) = \tau_{3,s,1} * \text{age}_{it} + \tau_{3,s,2} * \text{age}_{it}^2 + \tau_{3,s,3} * \exp(H_{it}) + \tau_{3,s,4} * EDUC_i + \tau_{g;3,s,5}$$

$$\text{xb4}(t,H_t,S_t = s \in \{1,2,3\}) = \tau_{4,s,1} * \text{age}_{it} + \tau_{4,s,2} * \text{age}_{it}^2 + \tau_{4,s,3} * \exp(H_{it}) + \tau_{4,s,4} * EDUC_i + \tau_{g;4,s,5}$$

For instance, for an individual who is in state 1, death (state 4) will come if three conditions hold at the same time: $\text{xb4} + \xi^4 > \text{xb3} + \xi^3$, $\text{xb4} + \xi^4 > \text{xb2} + \xi^2$ and $\text{xb4} + \xi^4 > \xi^1$. But if he is in state 3 already, it is only necessary that $\text{xb4} + \xi^4 > \xi^3$. Under this logic, in states 1 and 2, transition can be represented as a multinomial logistic process. For instance, the odds to transit from state 1 to 3 can be expressed as follows:

$$p_{13} = Pr(S_{t+1} = 3|t,H_t,S_t = 1) = \frac{\exp(\text{xb3}(t,H_t,S_t = 1))}{1 + \exp(\text{xb2}(t,H_t)) + \exp(\text{xb3}(t,H_t,S_t = 1)) + \exp(\text{xb4}(t,H_t,S_t = 1))}$$

In state 3, transition is simplified to a logistic process as their is only one state to transit into:

$$p_{34} = Pr(S_{t+1} = 4|t,H_t,S_t = 3) = \frac{\exp(\text{xb4}(t,H_t,S_t = 3))}{1 + \exp(\text{xb4}(t,H_t,S_t = 3))}$$

A limitation of the presented formulation is that it does not consider competing risk of death or disability. Chernew et al. (2016) presents evidence of a substitution between causes of death in the US, where mortality associated to CVDs have been declining in the last 20 years while those associated with respiratory and central nervous system are increasing. This means that the current model might overestate the value of extremely effective interventions if transition probabilities of observed survivors in the data might be higher not only for CVDs but also for those competing risks.[8] For practical purposes, ex-ante it might be hard to know if it is

3.3.1.2 CVD-Risk progression

Observed H is composed by permanent and transitory elements (Equation 3.3). This specification aims to capture the persistence of some conditions like choles-

[8] In other words, if it is the case that even if CVD become as treatable as hypertension (no difference on survival odds between states 2 and 3), the life-expectancy of those who transit into state 2 is lower than for those who never transit into such state. That is, if those at high risk of developing CVD are also at higher risk of developing dangerous complications non directly associated with CVD.

terol and blood pressure, jointly with the almost immediate effect (in weeks) of medication for controlling them.

The permanent component is the product of a long-run process that includes both an initial genetic endowment but also life-style choices as dietary habits, smoking and physical activity. These factors build up into chronic conditions, which are captured here with the persistence parameter ω_1 in Equation 3.4. Thus, heterogeneity on investments are captured by the observed initial variation on H. Such initial level will be transformed by *iid* normal innovations e (mean 0, standard deviation σ_e). In this AR(1) process, the long-run trend differs according to the education level and gender of the individual.

The transitory element captures the effect of health care into the index. While medication therapies reduce the levels of blood pressure and cholesterol, their effect is temporal. For instance, statins inhibits the production of an enzyme that is essential for the production of LDL cholesterol. Because the enzyme is only inhibited while the drug is in the body, statins are typically prescribed indefinitely. Hence, Equation 3.5 presents a specification where medication h has a linear effect on the temporary component of H. It also shows that each health state has a different constant which comes from other health-care interventions that might affect the index.

$$H_{i,t} = H^p_{i,t} + H^t_{i,t} \tag{3.3}$$

$$H^p_{i,t} = \omega_1 H^p_{i,t-1} + \omega_{g;2} + \omega_3 EDUC_i + e_{i,t} \tag{3.4}$$

$$H^t_{i,t} = \omega_4 h_{it} + \sum_{s=2}^{3} \omega_{5,s} \mathbb{1}(S_{i,t} = s) \quad , \quad s \in \{1,2,3\} \tag{3.5}$$

3.3.1.3 Medication choice

Medication choice is assumed to be exogenous to the individual. While it is true that in reality individuals might decide to accept or not a prescription by their doctor, it is not possible to distinguish such decision in the data. It is also assumed that medication's sole impact is through health transition probabilities. Hence, we are assuming that there are no noticeable secondary effects and that monetary costs are negligible.[9]

[9]The actual monetary cost was below 100£per year, or even 0 if individuals meet certain income conditions. However, other individual costs include to visit the GP practice for a repeated prescrip-

The probability to get a prescription is given by individuals' health (S, H), their education $(EDUC)$, local area characteristics that vary on time, and an unobserved component (ϵ). Equation 3.6 expresses such process for individual i at calendar year w and living in an area r. The model is a logit, as ϵ follows an Extreme Value Type I distribution. Local area is characterised by one characteristic, whether GP practices in the area are on average small or large $(SL \in \{0,1\})$, an effect that might change according to the calendar year associated to the survey collection time $(w \in \{4,5,6\})$. Such time variation according to practice size is allowed because of the increasing trend on prescription presented in Figure 3.1, and due to the introduction of pay-for-performance rewards related to primary prevention of CVDs in 2009 (see Appendix 3.F for further details).

$$
\begin{aligned}
h_{irw} &= \mathbb{1}(R > 0) \\
R &= \zeta_1 + \zeta_2 H^p_{irw} + \zeta_3 {H^p_{irw}}^2 + \zeta_4 EDUC_i \\
&\quad + \zeta_5 \text{Female} + \zeta_6 \mathbb{1}(w \geq 5) + \zeta_7 SL_r + \zeta_8 SL_r \cdot \mathbb{1}(w \geq 5) \\
&\quad + \epsilon_{irw}(h) \quad , \quad \epsilon_{iw}(h) \sim EV1(0,1)
\end{aligned}
\tag{3.6}
$$

The decision rule presented above can be understood as the reduce form of the optimal prescription behaviour of a practitioner that is partly altruistic and gets utility from the health of their patients. However, such decision depends on the constraints and incentives faced by her local practice which are captured by the size of it, and the calendar year.

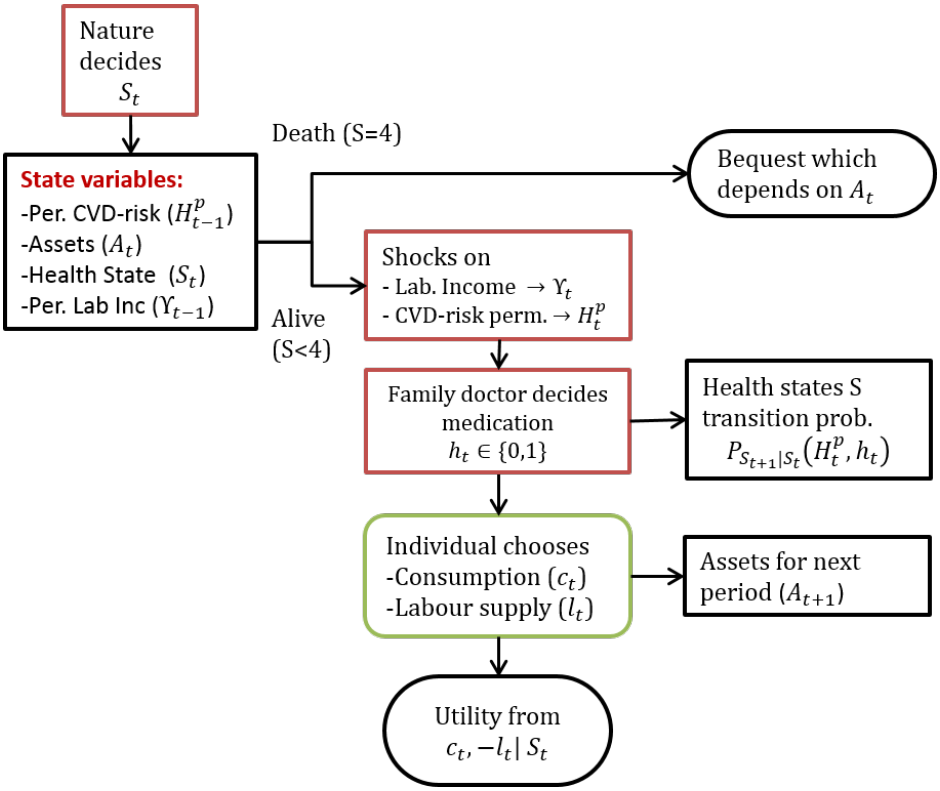
3.3.2 Life-cycle model

This is a Finite-Horizon Life-cycle model, for which each period timing, states and choices are sketched in Figure 3.3. Each period, individual i chooses hours of leisure $1 - l$ and consumption c conditional on being prescribed with lipid-lowering medication $h \in \{0,1\}$. Individuals are assumed to consider that the prescription rule is fixed and will not change in the future. It is important to notice that individuals are constrained by their assets A_{it} and exogenous shocks that might affect their potential income, which is a function of their education level

tion. On adverse effects, it is not expected to be harmed due to statins (Ebrahim et al., 2014).

and has a persistent component Υ_{it}. When an agent dies, he gets utility from leaving a bequest b. Apart from death ($S = 4$), the only health-state that has a direct impact on utility is to have suffered a CVD-disease ($S = 3$). However, notice that individuals will carefully consider their current CVD-risk index H and health state S in order to make decisions, as they rationally forecast the future of their health.

Figure 3.3: Life-cycle model diagram



3.3.2.1 Utility

Individual i, in period t chooses consumption c and hours of leisure ℓ. Utility follows a Cobb-Douglas function between consumption and leisure and it is nested within a constant relative risk aversion function (CRRA), with parameters that are gender and education specific. Consumption has a weight of $\eta_{g,ED}$ and risk aversion $\gamma_{g,ED}$.

Individuals younger than 90 can work either full-time (FT,$l = 40$ hours) or part-time (PT, $l = 20$ hours[10]) out of an available endowment of $\bar{L} = 112$ hours, a

[10]On practice, I followed Institute for Fiscal Studies derived variables which sets as full-time any work with more than 35 hours per week.

number that comes from the average awake time of people in the UK (Lader et al., 2006). Part-time and full-time working costs in term of hours are also shifted by parameter $\theta_{P;g,ED}$. Notice that if people have suffered from a CVD event ($S_{it} = 3$), there is a penalty of $\phi_{1;g,ED}$ of their time if they want to work ($l_{it} > 0$). This penalty might be higher or lower if they decide to work PT (ϕ_2).

If in a given period the individual is alive $(S = 1, S = 2, S = 3)$, they get utility according to Equation 3.7, but if it is death by this period, he will get it from their bequest as shown in Equation 3.8. Notice that this is equivalent to a non-depreciated durable good, generating a fixed flow of utility from the time of dead until the last potential life period T. This ensures that for any given period, the comparison between death and life has the same sign. As a result, it is not necessary to add a constant to the utility function to ensure that individuals prefer to be alive to be death.

$$
\begin{aligned}
U(c_t, l_t; S_t, A_t, \mathrm{Y}_t, X) =& u(c_{it}, l_{it}; S_{it}) = \frac{(c_{it}^{\eta_{g,ED}} \ell_{it}^{1-\eta_{g,ED}})^{1-\gamma_{g,ED}}}{1-\gamma_{g,ED}} \quad \text{if } S_{it} < 4 \qquad (3.7)\\
& \ell_{it} = \bar{L} - l_{it} - \theta_{P;g,ED} * (l_{it} = 40)\\
& \qquad - (\phi_{1;g,ED} + \phi_{2;g,ED} * (l_{it} = 20)) * (S_{it} = 3) * (l_{it} > 0)\\
& l_{it} \in \{0, 20, 40\}\\
U(c_t, l_t; S_t, A_t, \mathrm{Y}_t, X) =& b(A_{it}) = \theta_B \frac{(A_{it} + \theta_K)^{(1-\gamma_{g,ED})\eta_{g,ED}}}{1-\gamma_{g,ED}} \quad \text{if } S_{it} = 4 \qquad (3.8)
\end{aligned}
$$

The central parameters for our analysis are $\phi_{1;g,ED}, \phi_{2;g,ED}$. These parameters tell us about the burden of CVDs. Such effect is unlikely to be homogeneous across occupations, and for that reason it is allowed to differ by education level. Another crucial element is the utility of 'death', which is typically captured by the bequest model θ_B. This motive is typically identified by the amount of asset holding of the oldest individuals. Bequest motive is particularly important in the UK context: Blundell et al. (2016) report that in ELSA nearly 70% of households with one member aged 70 and older consider at least fairly important to leave property or money as an inheritance, in comparison with just 50% of similar households in the US. Notice that apart from these characteristics, health does enter directly into the

utility function.[11]

3.3.2.2 Assets progression and income processes

Individuals get resources for consumption from their household assets, which evolve according to Equation 3.9. Assets accumulate with the financial returns from the difference between income and consumption, given a risk-free and constant interest rate. Income comes either from labour ($y^L_{i,t}$) or other alternative sources ($y^{NL}_{i,t}$).[12] Borrowing is limited by the maximum amount of resources that they could potentially earn given a minimum consumption (Equation 3.10).

The onset of a CVD event has a one-off impact on assets accumulation, captured by parameter ϕ_3. This parameter considers out-of-pocket expenses, but in the UK this are more related to temporal disability, informal care provided by other economically active members of the household, and in general other costs related to the recovery time Liu et al. (2002). Expenses associated to demand of health care services, which increased notoriously during and after these events,[13] are normally covered by the publicly funded National Health Service (NHS).

$$\frac{A_{i,t+1}}{1+r} = A_{i,t} + y^L_{i,t} * \kappa * l_{i,t} + y^{NL}_{i,t}(\text{age}_{i,t}, l_{i,t}, S_{i,t}, \text{EDUC}_i, g, b) - c_{i,t} - \phi_3 \cdot (S_{i,t} = 3) \cdot (S_{i,t-1} \neq 3) \quad (3.9)$$

$$A_{i,t} \geq BC_t \text{ where } BC_i = \sum_{i,t=1}^{T} \frac{BC_{i,t+1}}{1+r} - y_{min} + c_{min} \quad (3.10)$$

As shown in Equation 3.11, income from labour $y^L_{i,t}$ (κ transforms weekly income into yearly one,[14]) either from wages or self-employment earnings, is a function of a stochastic idiosyncratic productivity process Y, current CVD-status, gender and of their education level. The productivity process follows an AR(1) given an initial draw $Y_{i,0}$ and is shaped by *iid* normal innovations v.

Non-labour income is the deterministic process presented in Equation 3.12. As it involves income from non-labour sources, earnings from other family mem-

[11] An alternative would be to weight Equation 3.7 by a function of the current health status as in De Nardi et al. (2010); Palumbo (1999).

[12] Even tough periods cover a two-years span, parameters in the income equations will be calculated for a year for interpretation purposes. In the data, income is elicited at week level, so reported figures are multiplied by 52 weeks and then doubled.

[13] See Appendix 3.E for more details.

[14] $\kappa = 52/1000$, for 52 weeks per thousand GBP.

bers, and the benefits and tax system, the two main drives behind this process are age and education level. In other words, instead of formally modelling such elements of the household, I will estimate a reduce-form version of it. First, the base level is determined by the education level and gender, but with a permanent jump at state pension age ($\text{SPA}_{g,b}$) which is gender (g) and cohort-of-birth specific (b). In the model, SPA is 65 for all men and for women born on and after 1954, while it is 60 for women born before such date.[15] Second, it has a positive trend with age which changes at SPA. The trend is quadratic after SPA in order to avoid huge non-labour incomes as age approach to 100. The third element is a response to labour supply: features of the benefit social welfare system like job-seeking allowance increase non-labour income as working hours decrease, which changes according to education level and age. The fourth element are pensions on top of the state pension: individuals who do not work will receive part of their potential labour income according to their participation status.

$$ln(y^L_{i,t}) = \iota^L_{g;1} + \iota^L_2 * \text{EDUC}_i + \iota^L_3 * (S_{it} = 3) + \text{Y}_{i,t} \tag{3.11}$$

$$\text{Y}_{i,t} = \rho \text{Y}_{i,t-1} + v_{i,t} \quad , \quad v_{i,t} \sim N(0, \sigma_v)$$

$$\begin{aligned} ln(y^{NL}_{it}) =& \iota_{g;1} + \iota_{g;2} * \text{EDUC}_i + (\iota_{g;3} + \iota_{g;4} * \text{EDUC}_i) * [\text{age} \geq \text{SPA}_{g,b}] \\ &+ \iota_{g;5} * (\text{age} - 50) * (\text{age} < \text{SPA}_{g,b}) \\ &+ [\iota_{g;6} * (\text{age} - \text{SPA}_{g,b}) * (\text{age} \geq \text{SPA}_{g,b}) \\ &+ [\iota_{g;7} * (\text{age} - \text{SPA}_{g,b})^2 * (\text{age} \geq \text{SPA}_{g,b}) \\ &+ [\iota_{g;8} + \iota_{g;9} * \text{EDUC}_i] * (\text{age} \geq \text{SPA}_{g,b})] * l_{it} \\ &+ [(\iota_{g;10} * (l_{it} = 1) + \iota_{g;11} * (l_{it} = 2)] * \ln(y^L_{it}) * [\text{age} \geq \text{SPA}_{g,b}] \end{aligned} \tag{3.12}$$

3.3.3 Solution

Individuals solve:

[15] Traditionally, the SPA was 65 for men and 60 for women. However, the 1995 Pensions Act equalised it to 65, with a transition regime from April 2010 to 2020. A new Pensions Act in 2011 speeded up the transition and increased the age for both genders to 66. For simplicity, the model assumes that SPA jumps immediately from 60 to 65 for women born in 1954 and afterwards, and that there is no change from 65 to 66. It is assumed that women knew well in advance the first change, so it is effectively incorporated in their behaviour.

$$\max_{(c_t,l_t)_{t=1}^{T}} u(c_1,l_1;S_1,X) + E\left[\sum_{t=2}^{T}\beta^{t-1}U(c_t,l_t;S_t,A_t,\mathrm{Y}_t,X)\right] \tag{3.13}$$

given a vector of parameters $\Theta = \{\beta,\eta,\gamma,\phi,\theta_B,\theta_K,\mu,\sigma_v,\rho,\iota,\omega,\tau,\zeta\}$, conditional on fixed characteristics X (education level, gender and birth-cohort,[16] calendar year and type of nearby GP practice). They have to decide the full stream of choices considering the potential utility every period (Eqs 3.7 and 3.8) subject to the progression of income (Eqs 3.11, 3.12), assets (Eqs 3.9, 3.10), health states (Eq 3.2), CVD-risk index (Eqs 3.3, 3.4 and 3.5) and odds of prescription (Eq 3.6).

An essential assumption for guaranteeing a solution is that in $T+1$ the individual will be death with certainty. This problem can be reformulated using bellman equations, which allows for the problem to be solved backwards from period $t = T$. As a result, all the estimated parameters will be relative to such arbitrary value set by the bequest function. This is a fundamental assumption for calculating the compensating variation for the drug, and will be further discussed in the welfare section.

Conditional on education level and prescription status, the value function for an individual who is still alive is shown in Equation 3.14. This means that rational individuals consider the potential paths of both their health (CVD-risk index, prescription, health status) and income (labour and non-labour). Notice that calendar year is not considered, this means that individuals take current prescription process as granted and do not expect it to change in the future. For notation benefit, individual subscript was removed, and time notation was simplified: $V = V_t$ and $V' = V_{t+1}$. More details on how the problem is solved are in the Appendix 3.D.1.

There are two main output of the solution. First, the expected maximised utility at a given point of the state space, $V(A,H,S,\mathrm{Y};X)$, presented in Equation 3.14. It is crucial as it allows us to calculate the measure of value as will be described in the next section. Second, the arguments of the problem at a given point of the state space, $A'^{*}(t,S,A,H,\mathrm{Y},h;X)$[17] and $l^{*}(t,S,A,H,\mathrm{Y},h;X)$. These policy functions are

[16] There are three options: males, females born before 1954, and females born on and after such date.

[17] A' (future assets) is stored instead of c as it is in the same units and magnitude of the state variable A (assets). Nevertheless, results should be similar using either of them.

used to predict choices while doing simulations. More details on the simulation procedure can be found in Appendix 3.D.4.

$$\begin{aligned} V(A,H,S,\mathrm{Y};\mathrm{X}) =& V_{h=1} * \Pr(h=1) + V_{h=0} * (1-\Pr(h=1)) \quad if \quad S<4 \\ V(A) =& b(A) \quad if \quad S=4 \end{aligned} \tag{3.14}$$

$$V_h = \underset{(c,l)}{\operatorname{argmax}}\, u(c,l;S) + \beta E\left[V'(A',H',S',\mathrm{Y}';\mathrm{X})|A,H,S,\mathrm{Y},h,c,l\right]$$

3.4 The measure of value: compensating variation

The measure is obtained by comparing the expected utility of both scenarios. The question, as show in Equation 3.15, is to determine the amount of money π_i that would be required in order to compensate an individual i for living in the no-medication world if we want him to attain the same expected utility as in the current scenario. It can also be interpreted as the maximum amount of assets that the individual will give up in order to live in a world where the medication system is available.

$$\pi_i \in \operatorname{argmin} \left| E_0 \sum_{t=\tau_i}^{T} \beta^t U(l_{it},c_{it}|h^*_{it},A_{it}) - E_0 \sum_{t=\tau_i}^{T} \beta^t U(l_{it},c_{it}|h_{it}(S_{it}=2)=0,A_{it}+\pi_i) \right| \tag{3.15}$$

This compensation variation can also be interpreted as the willingness-to-pay for the drug. How much will a person give up in order to enjoy the benefits of a treatment? A substantial difference from literature is that the value of this technology is not based solely on the trade-off life vs. death but also considers observed differences on labour hours and consumption. This means that it considers the value of quality of life improvements of the technology.

Papageorge (2015) exploits rich data on medication prices and choices, allowing him to identify willingness-to-pay based on revealed preferences. This is not feasible for most applications, therefore most of them are based on the implicit values of life present in the trade-off between wages and mortality risk of cer-

tain occupations (Viscusi, 1993). While this approach does not take into quality of health, it allows to incorporate estimates of the *Value of a Statistical Life* (VSL)[18] already established on the literature. For instance, Hall and Jones (2007) calibrate their utility function, where death is normalised to 0, in order to match VSL estimates.[19] In the model presented above, the sole driver of the difference between life and death in the model is the bequest motive presented in Equation 3.8, which is identified by the shape of the assets rather than from an observed trade-off that involves mortality risk.

In order to contrast the implications of this assumption I will calculate the VSL implied by the model. This is done by considering a counterfactual reduction in the mortality rate (Murphy and Topel, 2010).

3.5 Improving Cost-Effectiveness Analysis

The most common approach for evaluating whether or not to adopt a technology is the cost-effectiveness analysis. In the UK, NICE considers that an intervention provides value for money if it is able to gain one quality-adjusted life year[20] (QALY) if it costs £20.000 or less to do so. By using this criteria, this valuation strategy avoids economic considerations and centers the decision on the pure 'health' benefits. However, such economic benefits might produce extra resources to the health system.

QALYs calculation comes from calculating the incremental cost-effectiveness ratio (ICER), in the scenarios with and without the drug. For it, it is required to simulate two datasets under the estimated policy rules and the counter-factual scenarios. In total k datasets of size N are simulated in both scenarios. It is normally interpreted as the amount of money that would be required in order to

[18] How much of their assets an individual is willing to give up in order to reduced probability of death by 1 unit. However, calculations are based on extrapolating small reductions, for instance, of 1/1.000 or of 1/10.000. See (Viscusi and Aldy, 2003) for a good review.

[19] Hall and Jones (2007) also considers quality of life by introducing extra elements on the utility function that are calibrated with quality adjusted life years weights.

[20] In the health economics literature, health states are given a factor called *utility* that represents the physical and mental capacity of each state. Typically, an utility of 1 is considered one quality-adjusted life year. Such factors are normally derived from surveys that consider perceived quality of life and health conditions. For example, according to the EuroQol 5 dimensions (EQ-5D) measure of health status (Devlin and Krabbe, 2013) which is typically used in NICE guidelines, if the number 1 represents the best health state, a value of 0.079 is given for a state in which individuals require assistance for daily-life activities. For a simple introduction see Phillips and Thompson (2001) and Malek (2001).

gain 1 QALY using this treatment. Even though this interpretation has assumes that the production function is linear on resources and that there are no fixed costs, it is a simple and widely used reference for cost-effectiveness evaluation:

$$\text{ICER} = \frac{\text{Cost}_{\text{statin}} - \text{Cost}_{\text{no statin}}}{\text{QALY}_{\text{statin}} - \text{QALY}_{\text{no statin}}} \tag{3.16}$$

where $\text{Cost}_{\Xi} = \sum_{i=1}^{kN}\sum_{t=1}^{T} cost_{it}(\Xi)$ and $\text{QALY}_{\Xi} = \sum_{i=1}^{kN}\sum_{t=1}^{T} QALY_{it}(\Xi)$, with $\Xi =$ statin when h_{it}^{*} operates, and $\Xi =$ no statin when $h_{it}(S_{it} = 2) = 0$.

The same equation can be used for determine not the cost but any other simulated variable. For instance, we can obtain how much extra labour income is associated to 1 extra QALY in the treatment-available world as follows:

$$\frac{Y^{L}_{\text{statin}} - Y^{L}_{\text{no statin}}}{\text{QALY}_{\text{statin}} - \text{QALY}_{\text{no statin}}} \tag{3.17}$$

where $Y^{L}_{\Xi} = \sum_{i=1}^{kN}\sum_{t=1}^{T} Y^{L}_{it}(\Xi)$.

3.6 Structural Estimation of the Model

The model was estimated in several steps, separately for men and women. First, adjusted profiles are obtained in order to derive moments from the data. Most of the moments are unconditional means per age, but they also include transition probabilities, variances and serial covariance of labour income and the health-index. Next, initial conditions for simulating predictions by the model are derived. This involves an imputation procedure for the potential labour-income of those who are not working. This is carried out using a Heckman selection model where participation is instrumented with a dummy for having a partner who reports to be in bad health (see appendix 3.D.3 for more details).

The next step is to determine the value of the parameters that will reproduce the profiles observed in the data. First, the health-model parameters are estimated independently, and given their results, life-cycle model parameters are estimated for men and women separately. The discount factor is calibrated from literature and set to $\beta = 0.9604$[21] and the interest rate set to $r = 0.030225$.[22], all other parameters θ, a $q \times 1$ vector, are estimated via the Method of Simulated Moments

[21] This is the two-years equivalent to the more standard $\beta = 0.98$.
[22] This is equivalent to $r = 0.015$ in a one-year term.

(MSM). This requires to simulate a dataset similar to ELSA, which involves also to consider sampling and attrition. Details on how the simulated dataset is prepared are presented in appendix 3.D.4.

Let $\mathbf{x} = \{x_i\}_{i=1}^{N}$ be the data where we have N individuals. Each individual is observed at least two times or at most 6 between age 53 and T. Such data is transformed by a vector functions $\mathbf{m}$ of size $p \times 1$, such that $p > q$. The individual functions, m_j select a variable, or a combination of variables, at a specific age. For instance, wage; conditional on being observed at wave 2, being 53 to 54 years old, and being working at such year. Appendix 3.D.2 presents in detail each of the moments considered. These conditional moments are transformed into a unconditional ones using an indicator function. In other words, in order to produced sample averages, the sum of individual contributions is divided by the total number of individuals N, rather than the number of individuals who meet the condition set by the moment.[23] Given such information is possible to construct a sample average, the observed moment, $\psi_j(\mathbf{x})$, such that as $N \to \infty$

$$\psi_j(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} m_j(x_i) \to E(m_j(\mathbf{x}))$$

Given the model and the vector of parameters, it is possible to construct a simulated data analogue, $\mathbf{x}_{(k)} = \{x_i\}_{i=1}^{Nk}$, which size is k times the original sample. With it, we can apply the same vector $\mathbf{m}$ and construct simulated moments $\psi_j(\mathbf{x}_{(k)})$, such that as $kN_j \to \infty$

$$\psi_j(\mathbf{x}_{(k)}) = \frac{1}{kN}\sum_{s=1}^{kN} m_j(x_s) \to E(m_j(\mathbf{x}_{(k)}, \theta))$$

If the model is correctly specified $\forall j$, $E(m_j(\mathbf{x}_{(k)}, \theta)) = E(m_j(\mathbf{x}))$. This motivates the MSM estimator (McFadden, 1989; Pakes and Pollard, 1989; Duffie and Singleton, 1993):

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \left(\psi_j(\mathbf{x}) - \psi_j(\mathbf{x}_{(k)})\right)' \mathbf{W} \left(\psi_j(\mathbf{x}) - \psi_j(\mathbf{x}_{(k)})\right) \tag{3.18}$$

Where $\mathbf{W}$ is a diagonal matrix constructed with the diagonal of the inverse of

[23] See for instance Chamberlain (1992) and French and Jones (2011).

the covariance matrix of the data:[24] $\mathbf{W}^{-1} = \text{diag}(\mathbf{S})$ and $\mathbf{S} = \text{Var}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{m}(x_i)\right)$; and $\mathbf{J}$ is the jacobian matrix of the simulated statistics with respect to θ, $\mathbf{J} = \frac{\partial \frac{1}{kN}\sum_{i=1}^{kN}\mathbf{m}(x_i)}{\partial\theta}$. In practice, it is derived numerically using steps of 25% of the original parameter value, or of 0.01 if it is 0.

Then $\hat{\theta}$ is a consistent estimator of θ, and its asymptotic distribution is given by $\sqrt{N}(\hat{\theta} - \theta_0) \to N(0, Q)$. Matrix $Q = \left(1 + \frac{1}{k}\right)\left(\Sigma_1^{-1}\Sigma_2\Sigma_1^{-1}\right)$, with $\Sigma_1 = \mathbf{J}'\mathbf{W}\mathbf{J}$ and $\Sigma_2 = (\mathbf{J}'\mathbf{W}) * \mathbf{S} * (\mathbf{W}\mathbf{J})$.

The health model was estimated using an amoeba algorithm starting from a defined set of starting values. Parameters of the transition equations were initialised at 0, while starting values for CVD-risk dynamics and prescription equations come from estimating them outside the model. In terms of identification, the Markov health transitions and prescription equation involve a collection of logistic regressions which are identified by observing transitions between states across ages, education levels, GP-sizes and different values of the CVD-risk index. For the CVD-risk dynamics equation it also required to include moments that account for variance and autocovariance (4 and 8 years). The list of the moments constructed with such information is in Appendix 3.D.2. As any logistic models, parameters are relative to the variance of the innovations that have been standardised to 1.

On terms of the economic choices model, parameters are relative to β and r which were fixed. The are two motives for savings in this model: risk aversion, captured by the parameter γ_g; a the bequest motive represented by θ_B. While the shape of the assets' profiles is dominated by both parameters (inducing labour participation), the bequest motive allow us to understand why there is not a sharp drop in assets for the elderly (above age 80). These utility parameters were inititalised to standards in the literature, but several combinations were manually calibrated. Transitory and persistent components of labour income per hour are identified by including up to 4 autocovariances of the variable (from 2 to up to 8 years), on top of its cross-sectional variance.

[24] The optimal weighting matrix might be biased in small samples, so the diagonal is used as recommended by Pischke (1995).

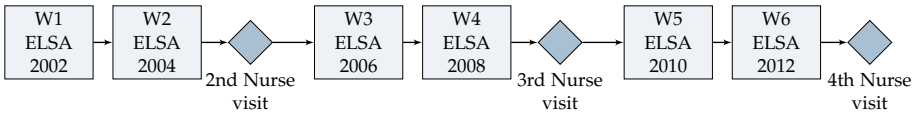
3.7 Data

This section presents the main sources of data for estimating a model that jointly considers life-cycle choices, health progression and statins prescriptions. First, a description of the individual-level dataset will cover the first two elements. Second, GP practice and PCT levels data will allow us to understand how health-care services are related to individual choices.

3.7.1 ELSA

The *English Longitudinal Study of Ageing* (Marmot et al., 2013) is a panel of individuals aged 50 and over, representative for those living in private households in England. I consider its first 6 waves spanning the period 2002/03 to 2012/13. Together with socio-economic information, it includes biomarkers collected by nurses in waves 2,4 and 6 (Figure 3.4).[25] Financial variables as wealth, income and non-labour income[26] of the household based on the *Institute of Fiscal Studies* income calculations. Financial figures were adjusted to May/June 2005 constant prices using monthly regional consumer prince index from the Office for National Statistics. A final element which makes the dataset attractive is that mortality information was obtained by ELSA directly during field work but also by it with data from the *Department for Work and Pensions* and the *National Health Service Central Register*.

Figure 3.4: Survey timing



While ELSA includes information from partners and other household members, the present study includes individuals who are part of the core dataset, that is, those selected in order to make the sample representative of England's population. Panel A of Table 3.1 shows the number of individuals observed in each wave according to their original inclusion. From the 11391 observations included

[25]For further information, please refer to:`http://www.ifs.org.uk/ELSA/about`

[26]This is calculated as non-housing and non-pension wealth of the benefit unit minus the labour income of the individual. All income measures in ELSA are net of tax. See Section 3.A.2 for more details about wealth calculations.

in wave 1, 8780 were reinterviewed in wave 2. The other 2611 are either death (506 of them) or not considered because they asked not to be re-contacted, it was not possible to do it, or moved out of Britain. Hence, it is possible to follow-up 81.5% of the original sample.[27] According to ELSA reports (Bridges et al., 2015), missing information is correlated with age, ethnicity, region and subjective health measures.

Panel B of Table 3.1 presents the number of individuals included in the main estimation. In wave 2, nurses visited 88% of the interviewed core members. The remaining refused to be contacted for such purpose (Scholes et al., 2008). From 81% of those who were visited by a nurse, a blood sample was collected (6231 individuals). The figure in Table 3.1 is smaller as the measurements had to be valid for blood pressure and cholesterol levels, as it will be discussed in Section 3.7.1.2. This is a common problem for this kind of instruments (Heidi Guyer, 2010), and as a general rule, the resulting analysis will be over a population who is in better health relative to the general population.

ELSA provides population weights in order to recover representative estimates. Cross-sectional weights are going to be used in this analysis when discussing the willingness-to-pay calculations from wave 1 perspective, however, longitudinal weights are not considered as the model will incorporate attrition.

[27] Given that ELSA wave 1 is based on the *Health Survey of England*, ELSA wave 2 report estimated the longitudinal response to be around 46.6% of the original sampling framework (Scholes et al., 2008).

Table 3.1: ELSA Structural Estimation Sample

	Number of individuals					
	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
Source Sample	2002/03	2004/05	2006/07	2008/09	2010/11	2012/13
Panel A: Original sample						
Since wave 1	11391	8780	7535	6623	6242	5659
Since wave 3	0	0	1275	972	936	888
Since wave 4	0	0	0	2291	1912	1796
Since wave 6	0	0	0	0	0	826
Total	11391	8780	8810	9886	9090	9169
Confirmed Deaths	0	506	1033	1663	2229	2701
Total Observed	11391	9286	9837	11499	11286	11867
Panel B: Included sample						
Used for the moments	11358	9260	9807	11450	11214	11737
CVD-risk index	0	5058	0	5631	0	5439

Source: Own calculations based on ELSA waves 1-6, core members.
Notes: Panel B consider observations that are used for calculating the moments for the structural estimation of the model and includes registered deaths. In this context, observations without a valid age, work or health status were discarded. Also, notice that CVD-risk index is based on those individuals for which there is valid measurements of blood pressure and cholesterol.

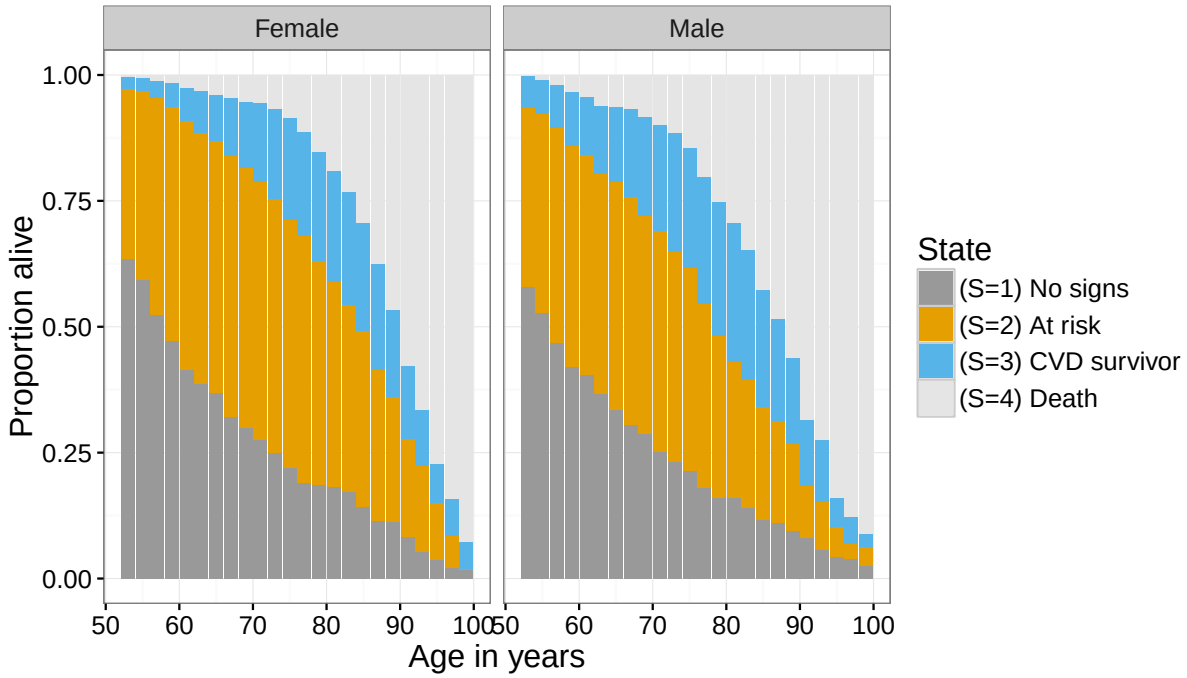
3.7.1.1 Data profiles

Following the model, health is categorise in four states which are based on reported conditions diagnosed by a physician. First, individuals do not present signs of any cardiovascular condition ($S = 1$). This might change, and they might present early signs denoted by persistent levels of high blood pressure, cholesterol or sugar in their blood ($S = 2$). Those are chronic conditions that can be managed, but not reversed. A more serious scenario is to suffer an CVD event, which might generate permanent limitations for work ($S = 3$). Last state, $S = 4$, is death.[28]

Figure 3.5 shows the proportion of individuals in each health state, at a given two-years age group (52-53, 54-55 and so on).[29] For both men and women the proportion of individuals in high risk of suffering from a CVD event is increasing until age 70. For those who survived such event, the peak age is around 80. The main difference is that more male are affected by CVD than women, but more women move into the 'at risk' status.

[28]Death is observed via an administrative linkage from ELSA with ONS records.
[29]Age is grouped in order to boost the number of observations per group.

Figure 3.5: Observed progression of health states

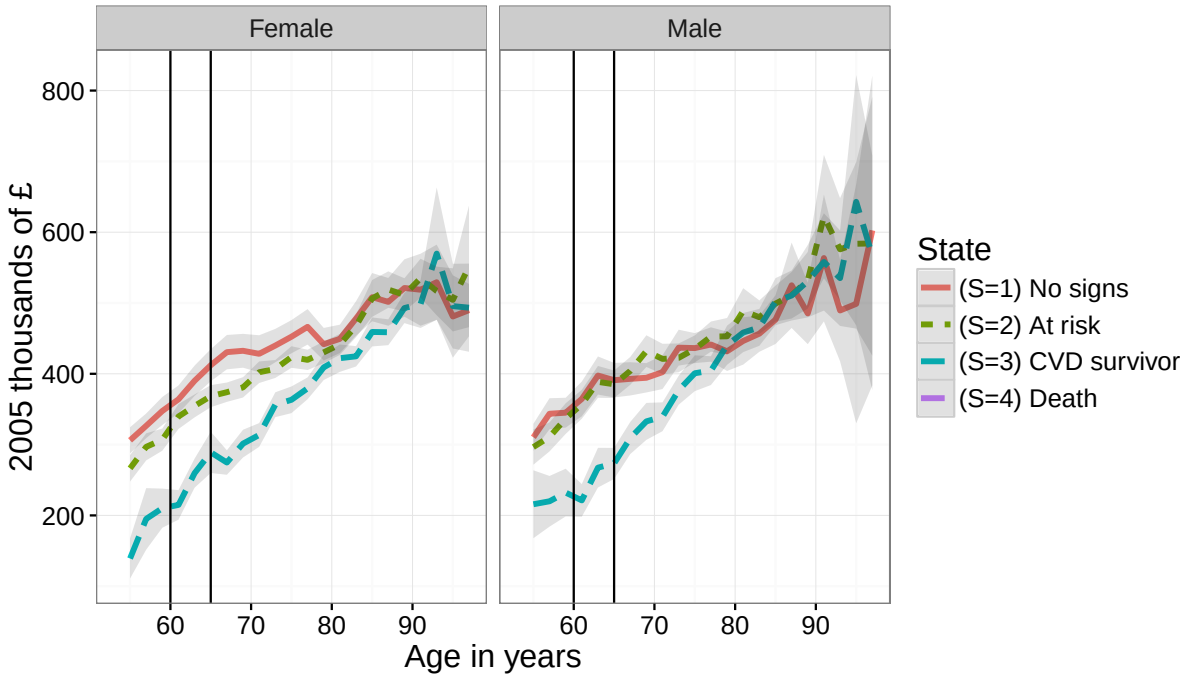


Figures 3.6, 3.8 and 3.7 show the assets, income and labour supply profiles after netting out the effects of cohorts, family size, and seasonal and business cycle.[30] I follow a procedure similar to French (2005) which using linear regressions estimates the contribution of all this confounding elements in each one of the outcomes, while controlling for the contribution of health states and age. Given this estimates, variables are adjusted by removing the specific effects in favour of common normalisation. Specifically, the profiles adjustment aims to set all individuals in a world with a fix unemployment rate of 5, a family size of 2, and to be born in 1946. Appendix 3.A explains in detail this procedure.

First, Figure 3.6 shows that assets are considerable smaller for CVD survivors. Part of this might be triggered by lower participation in the labour market as shown in Figure 3.7. As a second observation, it is clear that there are differences on this aspect, specially before the SPA. The effect is larger in full-time participation for men, and in part-time for women. Third, Figure 3.8 shows that earnings per hour are slightly lower for those participating in the labour market after surviving a CVD event. This suggest a productivity effect if those working are those who are more likely to earn more.

[30] 90% confidence intervals are based on the standard error of the mean

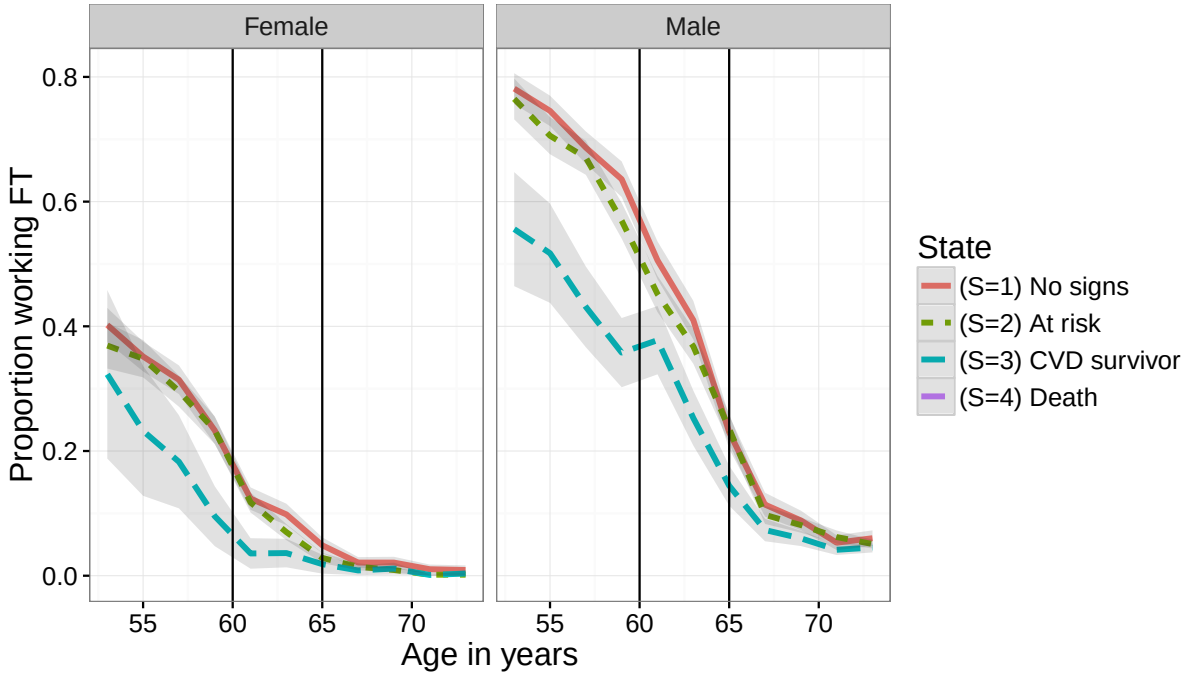
Figure 3.6: Observed Assets Profile



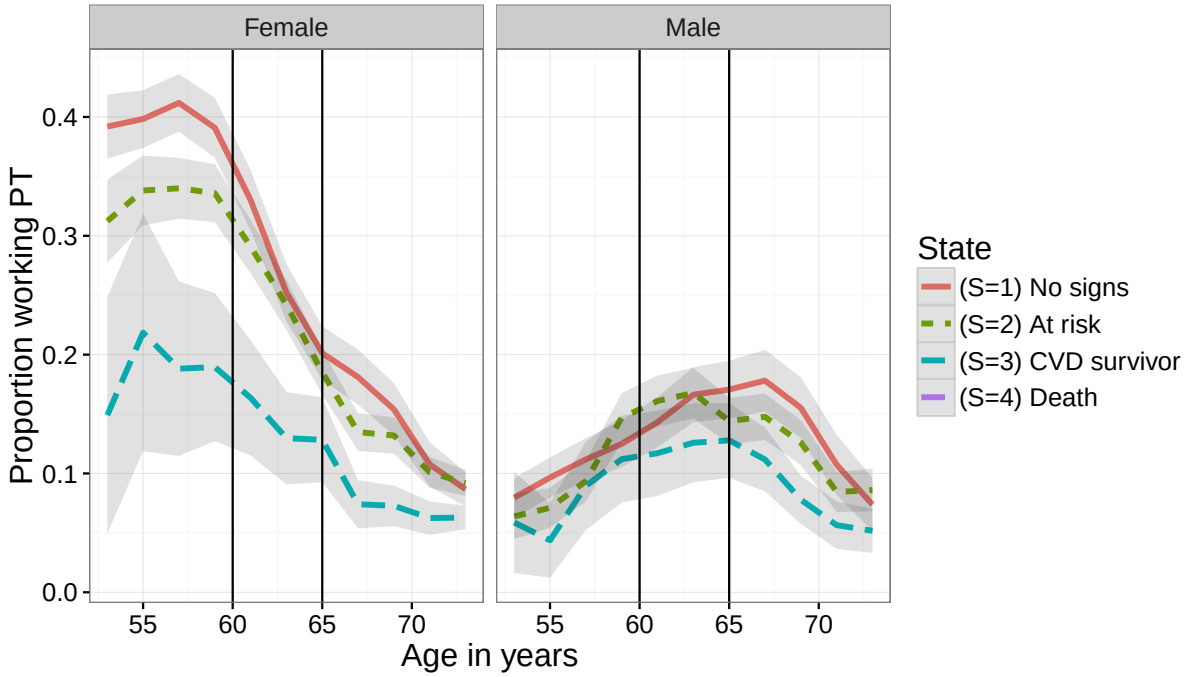
Notes: Own calculations based on IFS derived variables on ELSA waves 1 to 6. Adjusted for Birth cohort, family size and regional unemployment. Includes 90% CI.

Figure 3.7: Observed Labour Supply Profiles by Health State

(a) Full-time



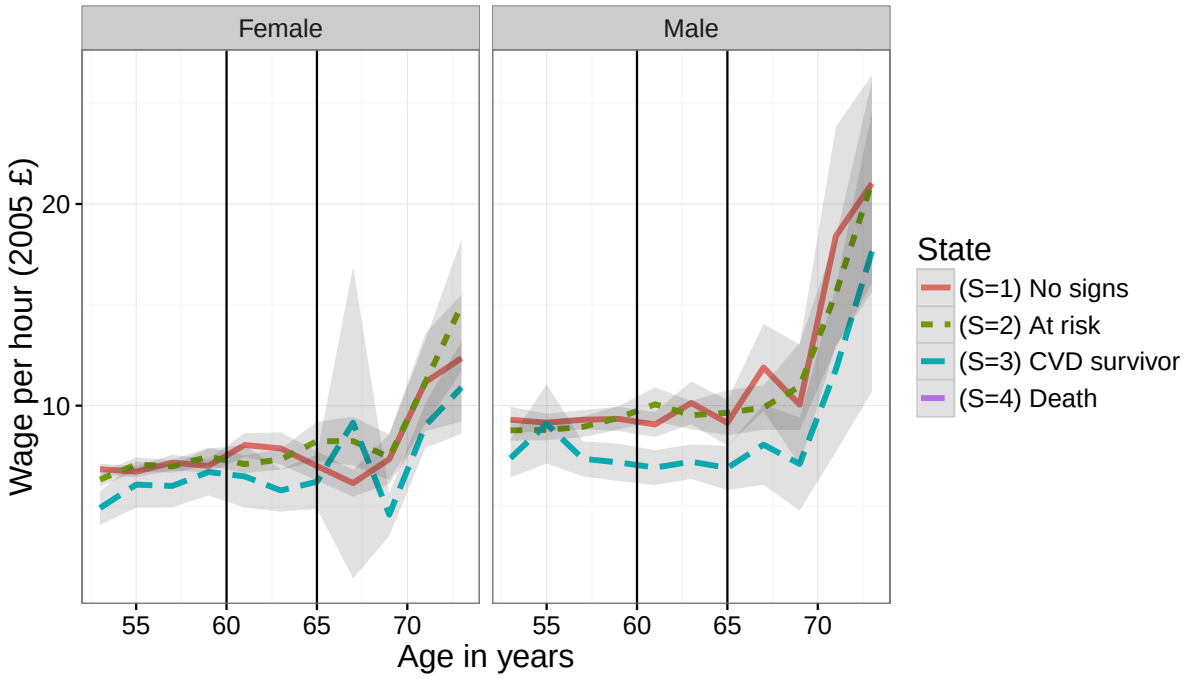
(b) Part-time



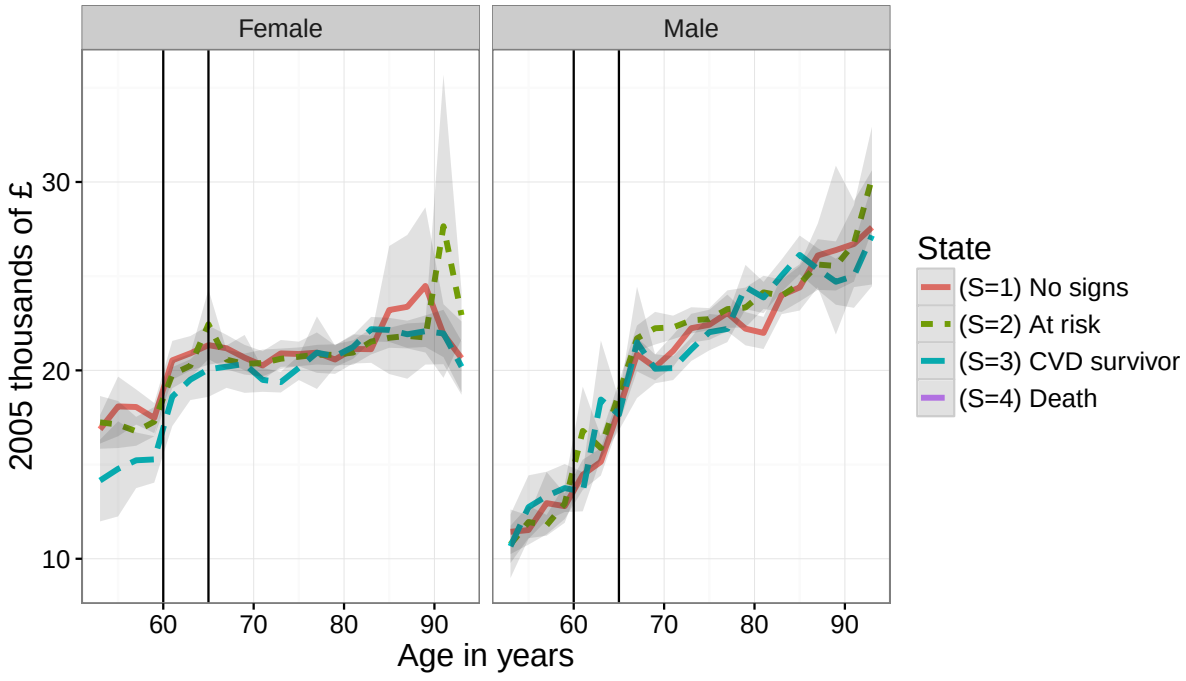
Notes: Own calculations based on IFS derived variables on ELSA waves 1 to 6. Adjusted for Birth cohort, family size and regional unemployment. Includes 90% CI.

Figure 3.8: Observed Income Profiles by Health State

(a) Wage per hour



(b) Yearly non-labour, non-financial income



Notes: Own calculations based on IFS derived variables on ELSA waves 1 to 6. Adjusted for Birth cohort, family size and regional unemployment. Includes 90% CI.

3.7.1.2 Cardiovascular Risk Index

A CVD risk index was built using some of the biomarkers collected by ELSA: systolic and diastolic blood pressure, non-HDL[31] and total cholesterol. These variables are summarized into a single index using the Anderson (2008) summary index. Anderson's summary index weight standardised version of the variables in order to maximise de variance of the index. Panel A of Table 3.2 presents the weights given to each variable given such criteria.[32]

Table 3.2: Anderson Summary Index (2008) for CVD-Risk Biomarkers and Statins

A. Anderson Summary Index (2008) for CVD-Risk Biomarkers

Variable	**Mean**	**SD**	**Weight**
Systolic BP (mmHg)	132.655	16.908	0.346
Diastolic BP (mmHg)	74.253	10.978	0.209
Total Cholesterol (T-C, mmol/L)	5.684	1.203	0.253
Non-HDL Cholesterol (Non-HDL-C, mmol/L)	4.103	1.125	0.191

B. Cochrane's (2013) estimates of Statin effect on Cholesterol

Variable	**Reported**	**SDvs**	
Total Cholesterol (T-C, mmol/L)	-1.05	-0.873	
Non-HDL Cholesterol (Non-HDL-C, mmol/L)†	-1.01	-0.898	
Total effect (ω_4)			-0.392

Notes: † Cochrane (2013) reports effects on LDL Cholesterol (-1.00 mmol/L). This number was translated into Non-HDL-C given the observed relationship in the ELSA sample: 1 mmol/L of LDL-C is equivalent to 1.005 to 1.0155 mmol/L of Non-HDL-C. Moreover, the correlation coefficient between LDL-C and Non-HDL-C is 0.9476. The advantage of Non-HDL-C is that fasting is not required for the validity of the measure, which is essential for LDL-C calculation.

[31] While the dataset includes LDL cholesterol, sample size is smaller than for HDL cholesterol due to the nature of the test: LDL requires fasting prior to the blood sample, while HDL does not.

[32] The index is a weighted average of the standardised variables, with weights that maximize the variance of the index. For more details, see Appendix A from Anderson (2008). Alternatively, a factor model was calculated. While both of them are highly correlated, the Anderson's version is a better predictor of the Framingham score. Fundamentally, the weight for each variables is calculated as the sum of the elements of the respective row of the inverse of the covariance matrix, with respect to the total sum of the elements of the same matrix. Therefore, more weight is given to those variables for which their variation is less captured by other variables variation. In comparison, the factor analysis captures the common variance of those variables. In this particular application, it is important to take into account both individuals with either high cholesterol or high blood pressure, rather than those with high levels of both biomarkers. The reason is that each measure might include essential information for predicting CVD risk.

The resulting index, a number between -2 and 2, is a measure of the underlying risk of developing a CVD event. Figure 3.9 presents its distribution according to their medication status, where despite selection,[33] there is a clear difference between those who are and not taking statins, possibly due to their beneficial effect on CVD risk.

The role of statins in the model is defined in Equation 3.5. This parameter is calibrated based on evidence from literature. In particular, Taylor et al. (2013) review on the effect of statins on primary prevention of CVDs, estimates the impact of the drug usage on the reduction of cholesterol levels. Given that the CVD-risk index mixes standardised versions of cholesterol and blood pressure, Panel B of Table 3.2 translates those point estimates into the index.

Figure 3.9: Distribution of the log CVD-risk index $\ln(H)$

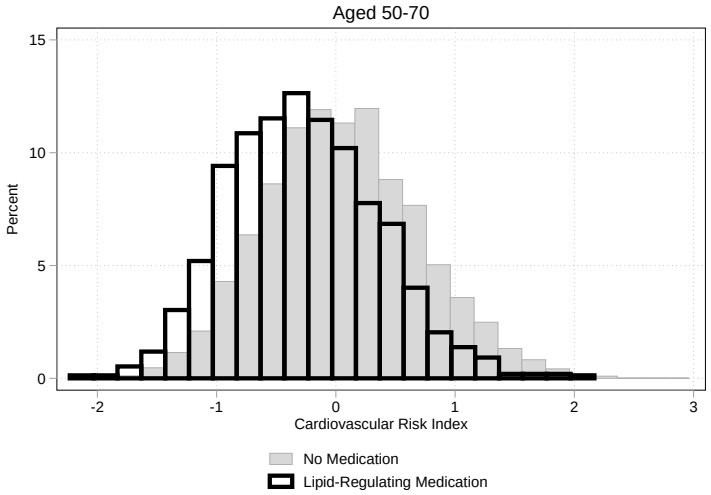


Figure 3.10 presents the age profiles for the resulting index and medication intake, showing a similar picture for both men and women. In general, there is a large decrease on the CVD-risk index for those who survive a CVD event. Appendix 3.E discusses the potential channels behind these trends in detail.

One concern is the potential effect of the drug-therapy on other factors that are related with the risk of developing a CVD or death. Strazzullo et al. (2007) review of clinical literature shows that Statins cause a small (1.1 mmHg) but significant reduction on systolic blood pressure. This implies an underestimation of the value

[33] Those under medication should have, on average, a higher CVD-risk than those who are not.

of the drug. Nevertheless, a risk compensation mechanism might be triggered, Kaestner et al. (2014) have found that Statins' adoption is related to an increase on BMI and alcohol use. If such compensation is not of the same magnitude in the clinical trials, this might generate an upward bias on the current estimates of the drug's benefits.

3.7.2 Practices data and prescription rule

Equation 3.6 presented the probability to be prescribed with statins. Information for constructing moments required for its identification comes from administrative data on primary care services from the QOF registers[34] which apart from performance for each GP practice, includes data on raw prevalences of some diseases derived from clinical records, and the number of registered patients.

Panel A presents the size of GP practices measured according to the number of patients registered in them. As a general trend, the number of practices has reduced while the number of patients per practice has increased. On average, each practice has around 60 new patients per year. At the same time, there was an increase in the adjusted number of physicians per capita. However, detected prevalence of risk conditions as hypertension and diabetes have increased, part of it due to ageing of English population and to better quality of primary care services.

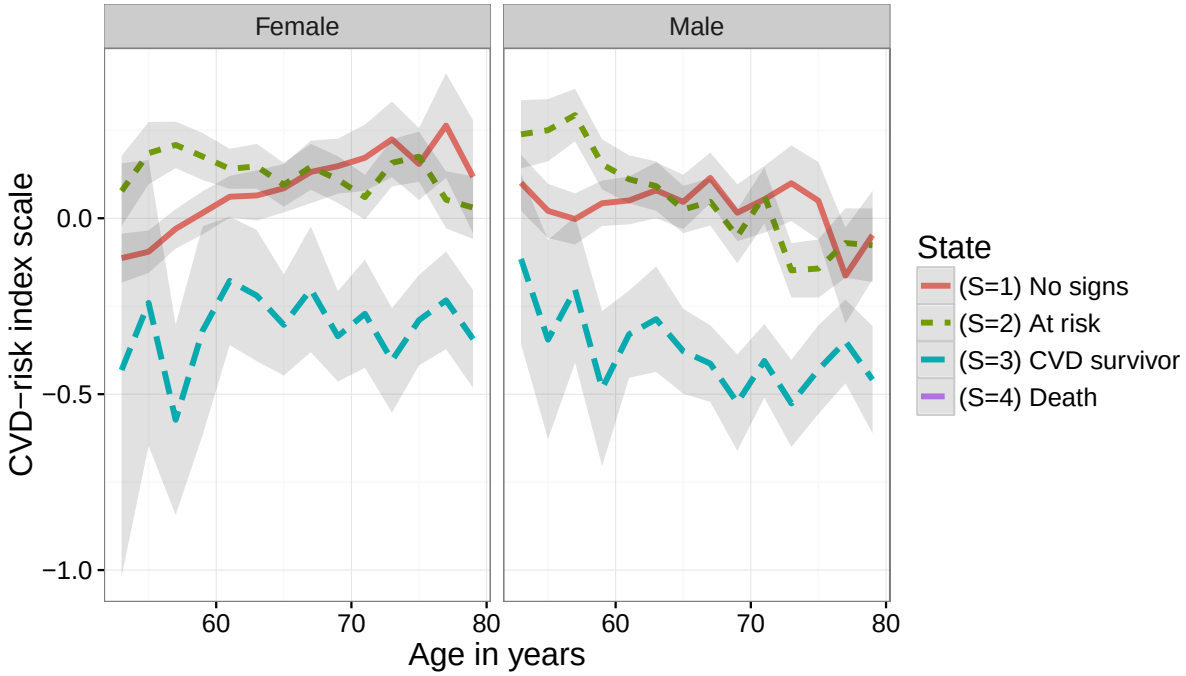
ELSA data is linked to GP practices' information based on respondents' post code of residence. While PCTs cover a well defined geographical area, practices catchment areas definition is not standard and overlap in urban areas. As a result, individuals are able to choose among nearby practices. In order to stablish a measure of the type of available health care services, I have considered information for the 10 closest GP practices within 15 Kms, according to Euclidean distance. For each respondent, list size information of the selected practices was averaged with equal weights as long as they are within 1 Km of the postcode, based on 2009/10 information. For observations outside that range, their relative weight decays with the inverse of distance.[35]

[34]This data archived and published by the *Health & Social Care Information Centre* (HSCIC). It is recorded in the *Quality Management and Analysis System* (QMAS), which is the source for deriving the payments calculations.

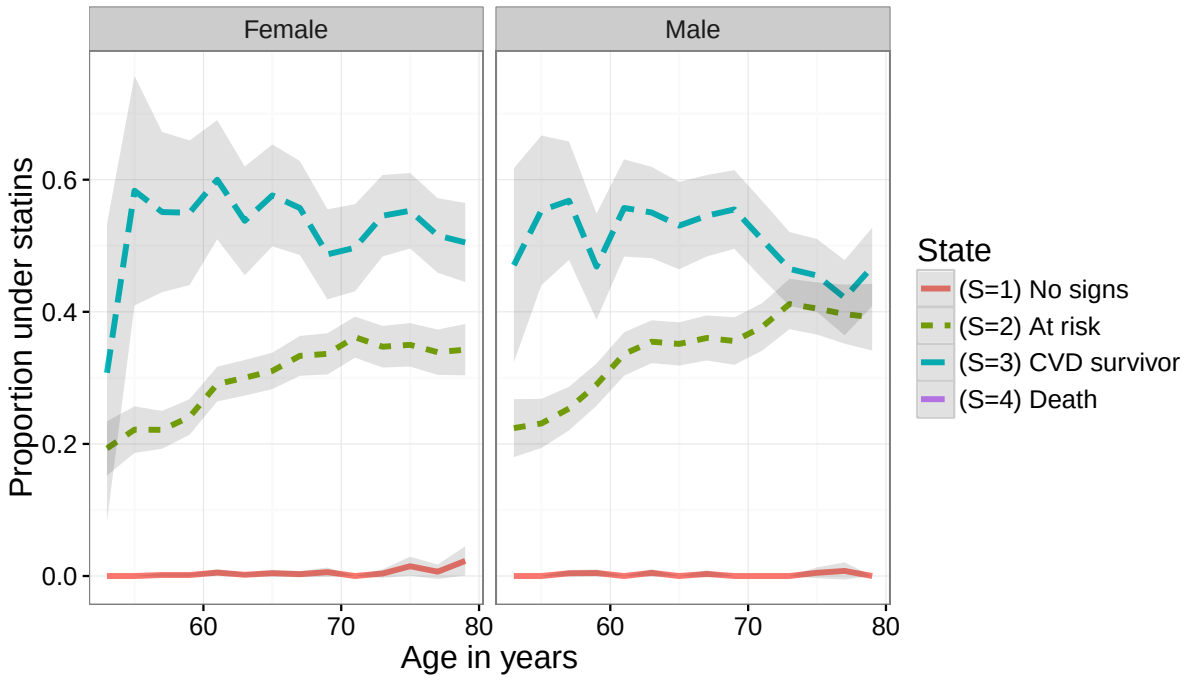
[35]See Appendix 3.B for further details.

Figure 3.10: Observed progression of CVD-risk Index and Medication

(a) CVD-risk index



(a) Medication intake



Panel B of Table 3.3 shows the result of the process. Nearly 80% of the observations included in the model (Panel B of Table 3.1) analysis were matched to at least one GP practice. However, if we do not consider those who are already

death by each wave, almost 90% of the observations have a valid match. Around a third of them live in areas where the closest GP practices are *small*, that is, with an average of less than 6000 patients. Such practices are almost half of the more than 8000 GP practices in England.

Table 3.3: GP Practices List Size Summary

Source Sample	Wave 3 2006/07	Wave 4 2008/09	Wave 5 2010/11	Wave 6 2012/13
Panel A: GP Practices		Average by practice		
Number of patients in thousands (list size)	6.41	6.60	6.69	6.98
Less than 6000 patients	0.54	0.52	0.51	0.49
Between 6000 and 8300 patients	0.18	0.18	0.18	0.19
8300 patients and above	0.28	0.30	0.31	0.33
List size year variation	.	60.56	58.64	91.62
Number of practices	8372	8229	8245	8020
Panel B: ELSA matched with nearby GP practices 2009/10 data		Average by individual		
Number of patients (1000s)	7.48	7.51	7.51	7.50
Less than 6000 patients	0.28	0.28	0.28	0.28
Between 6000 and 8300 patients	0.39	0.38	0.39	0.39
8300 patients and above	0.33	0.34	0.33	0.33
QOF: PP01 achievement	.	.	80.67	82.61
Number of individuals included in the sample	7683	9213	9005	9377

Source: Own calculations based on the Attribution Data Set GP-Registered Population.
Notes: Prevalences and list size (number of patients) information is derived from QOF data. PP01 is a clinical indicator of the QOF on primary prevention of CVDs.

3.8 Results

3.8.1 Parameters and Fit

The main goal of the model is to be able to forecast lifetime choices and states for English population aged 53 and over at a given moment. As a result, a first impression of such ability is how it fits the actual progression of observed data. The next set of figures present the data profiles both from ELSA and from a dataset simulated using the model. Such simulations are based on the estimated coefficients presented in Tables 3.4 and 3.5.

Table 3.4: Parameters from the Structural Model I: Utility and Income Process

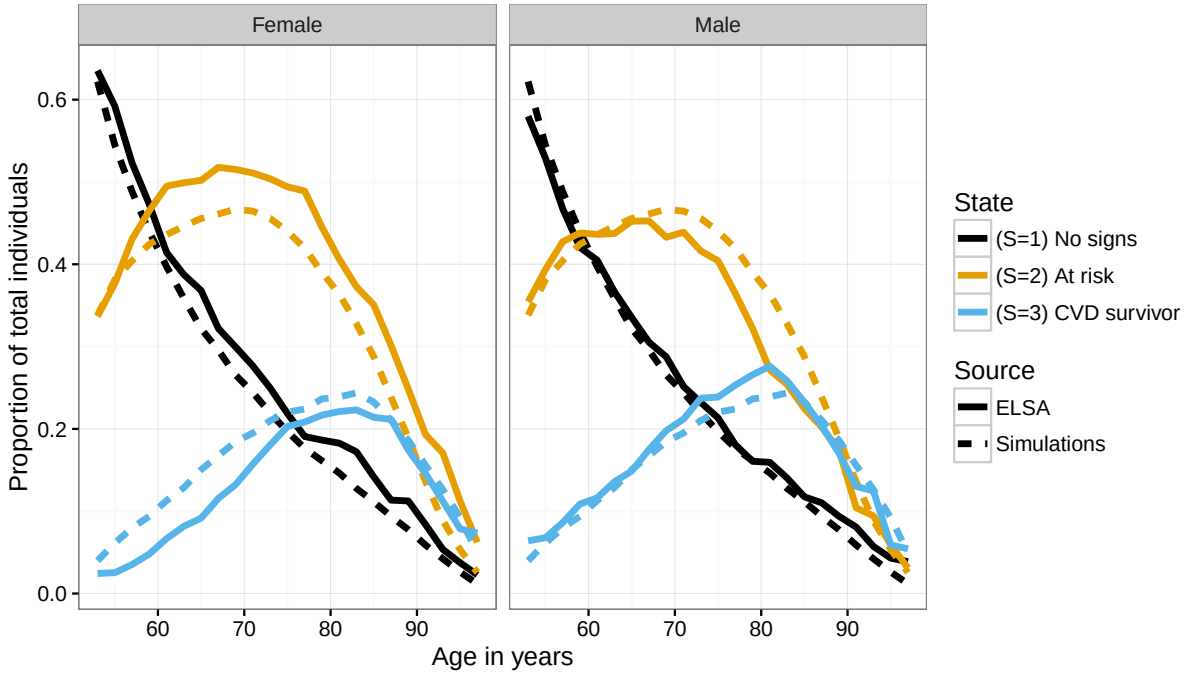
	Parameter	Male		Female	
		Value	SE	Value	SE
Utility					
β	Discount factor **(Cal)**	0.9604		0.9604	
No Education (ED=1)					
η	Importance of consumption in utility	0.5018	0.0049	0.4307	0.0012
γ	Relative risk aversion	3.9561	0.1656	3.2435	0.0148
θ_p	Unflexibility penalty for FT work	-3.1488	0.2074	1.3808	0.2554
ϕ_1	Disutility of bad health while working	7.7135	1.2082	6.2889	1.1380
ϕ_2	Additional disutility of bad health while PT	-1.5579	0.5250	-3.1569	0.9432
At least high school (ED=2,3)					
η	Importance of consumption in utility	0.5491	0.0033	0.5951	0.0003
γ	Relative risk aversion	3.4819	0.1765	3.9129	0.1569
θ_p	Unflexibility penalty for FT work (ED=2)	-3.6253	0.1298	-0.3484	0.0944
θ_p	Unflexibility penalty for FT work (ED=3)	-3.9688	0.1110	-2.0189	0.1593
ϕ_1	Disutility of bad health while working	7.5768	0.5876	7.5214	1.0271
ϕ_2	Additional disutility of bad health while PT	-2.4389	0.1949	-4.6596	0.9254
All levels of education					
ϕ_3	Asset one-off cost of CVD event	-54.0091	6.6823	-35.0947	3.4689
θ_B	Bequest importance	0.6926	0.0388	0.4646	0.0237
θ_K	Base for not leaving bequest penalty **(Cal)**	30.0000		30.0000	
Earnings per hour					
σ_u	Variance innovations	0.5804	0.0158	0.5349	0.0102
ρ	Persistence innovations	0.5507	0.0216	0.3897	0.0222
ι_1^L	EDUC	0.8778	0.0476	0.9161	0.0361
ι_2^L	To be in State 3	-0.1125	0.0269	-0.5587	0.0804
ι_3^L	Mean	1.1904	0.0345	1.0199	0.0272
Non-labour Income					
ι_1	Constant	2.1653	0.0189	2.5744	0.0033
ι_2	EDUC	0.3302	0.0236	0.2536	0.0210
ι_3	Age $\geq SPA$	0.5931	0.0245	0.1323	0.0161
ι_4	Age $\geq SPA$ * EDUC	-0.4526	0.1549	-0.8998	0.1940
ι_5	(Age<SPA) * (Age-50)	0.2136	0.0161	0.0890	0.0153
ι_6	Hours of work	-0.4307	0.0242	-0.2076	0.0181
ι_7	Hours of work * EDUC	0.4988	0.0393	0.2034	0.0284
ι_8	Hours of work * Age $\geq SPA$	0.0437	0.0125	0.1690	0.0163
ι_9	Hours of work * Age $\geq SPA$ * EDUC	0.0563	0.0188	0.0228	0.0122
ι_{10}	Age $\geq SPA$ * Y^L * Not work	0.1011	0.0078	0.1961	0.0036
ι_{11}	Age $\geq SPA$ * Y^L * Part-time	0.0653	0.0041	0.1017	0.0022

Table 3.5: Parameters from the Structural Model II: Health Process

	Parameter	Value	SE
CVD-risk index			
σ_ε	Variance innovations	0.4242	0.0008
ω_1	Persistence	0.7234	0.0005
ω_2	Constant	0.0232	0.0011
ω_3	EDUC	-0.0296	0.0015
ω_4	Health investment effect **(Cal)**	-0.3900	
$\omega_{5,2}$	To be in State 2	0.1315	0.0029
$\omega_{5,3}$	To be in State 3	-0.2107	0.0034
ω_6	Female	0.0371	0.0005
Prescription Model			
ζ_1	Constant	-2.8955	0.0476
ζ_2	H	0.0943	0.0143
ζ_3	H^2	2.0189	0.0741
ζ_4	EDUC	1.1630	0.0350
ζ_5	To be in State 3	1.6119	0.0339
ζ_6	Female	-0.0558	0.0100
ζ_7	Wave 6	0.4128	0.0489
ζ_8	Small list area	0.5984	0.0310
ζ_9	Small list area after 2009	0.5825	0.0456
XB2: latent evolution of transition into State 2 (CVD risk)			
τ_{21}	Age	0.0000	0.0019
τ_{22}	Age Sqd/100	0.0000	0.0016
τ_{23}	CVD-risk index H	0.7994	0.0109
τ_{24}	Constant	-3.4827	0.0748
τ_{25}	EDUC	-0.0054	0.0031
τ_{26}	Female	0.0059	0.0058
XB3: latent evolution of transition State 1 into 3 (CVD event)			
τ_{311}	State 1: Age	-0.3103	0.0004
τ_{312}	State 1: Age Sqd/100	0.2747	0.0017
τ_{313}	State 1: CVD-risk index H	0.2807	0.0151
τ_{314}	State 1: Constant	3.5065	0.0866
τ_{315}	State 1: EDUC	0.9040	0.0516
τ_{316}	State 1: Female	0.1802	0.0248
XB3: latent evolution of transition State 2 into 3 (CVD event)			
τ_{321}	State 2: Age	0.0000	0.0022
τ_{322}	State 2: Age Sqd/100	0.0000	0.0018
τ_{323}	State 2: CVD-risk index H	0.3239	0.0093
τ_{324}	State 2: Constant	-3.5000	0.0867
τ_{325}	State 2: EDUC	-0.1784	0.0198
τ_{326}	State 2: Female	-0.6525	0.0164
XB4: latent evolution of transition State 1 into 4 (death)			
τ_{411}	State 1: Age	-0.3103	0.0004
τ_{412}	State 1: Age Sqd/100	0.2747	0.0015
τ_{413}	State 1: CVD-risk index H	0.0009	0.0022
τ_{414}	State 1: Constant	4.6067	0.0632
τ_{415}	State 1: EDUC	-0.0621	0.0096
τ_{416}	State 1: Female	-0.7480	0.0437
XB4: latent evolution of transition State 2 into 4 (death)			
τ_{421}	State 2: Age	-0.2929	0.0004
τ_{422}	State 2: Age Sqd/100	0.2683	0.0010
τ_{423}	State 2: CVD-risk index H	0.7348	0.0094
τ_{424}	State 2: Constant	4.1552	0.0391
τ_{425}	State 2: EDUC	-1.0362	0.0234
τ_{426}	State 2: Female	-0.8065	0.0129
XB4: latent evolution of transition State 3 into 4 (death)			
τ_{431}	State 3: Age	-0.2974	0.0020
τ_{432}	State 3: Age Sqd/100	0.2657	0.0027
τ_{433}	State 3: CVD-risk index H	0.7410	0.0258
τ_{434}	State 3: Constant	4.3937	0.0462
τ_{435}	State 3: EDUC	-0.0079	0.0029
τ_{436}	State 3: Female	-0.6562	0.0215

Figure 3.11 presents the proportion of individuals per health state by gender. The continuous line presents the proportion per age derived from ELSA while the dashed one comes from the simulations. The other main components of the health model are presented in Figure 3.12, where Panel A presents the average CVD-risk index progression while Panel B the proportion of individuals under statins. In general terms, the model fits adequately such profiles.

Figure 3.11: Observed and Simulated Health States progression



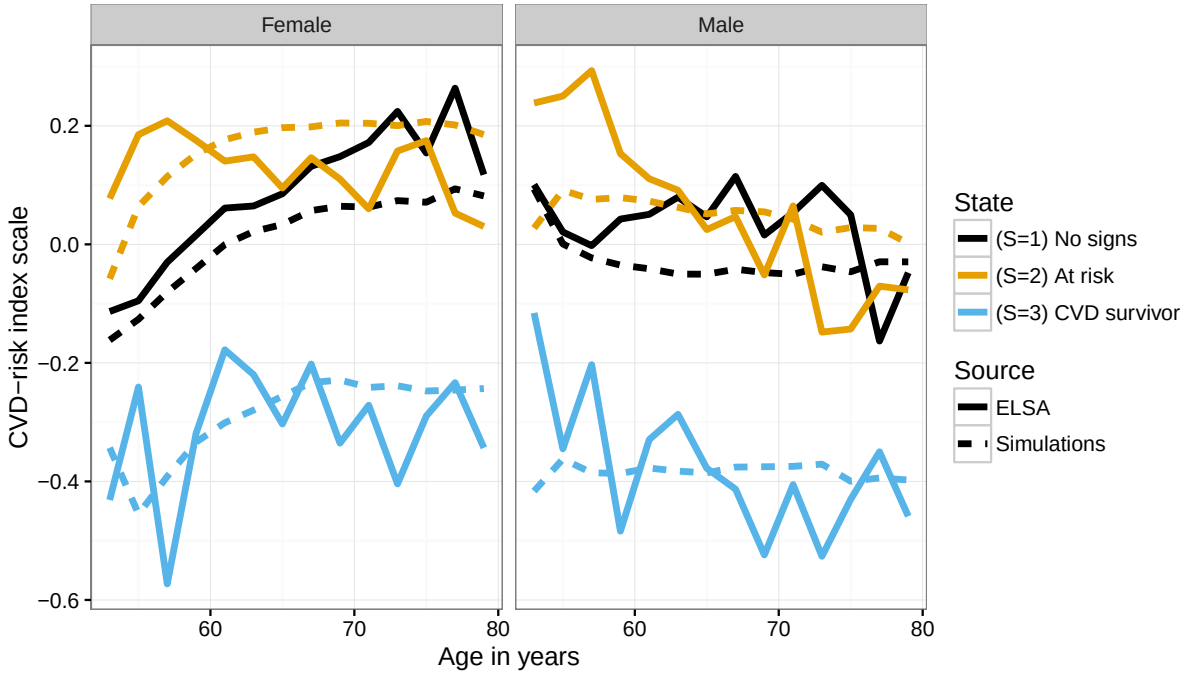
Notes: Own calculations based on IFS derived variables on ELSA waves 1 to 6. Adjusted for Birth cohort, family size and regional unemployment.

Assets are presented in Figure 3.13, labour supply in Figure 3.14 and income in 3.15. As a whole, the model captures the general trends of the data. Labour participation is reduced after the onset of a CVD principally due to the extra cost of labour (parameters ϕ_1, ϕ_2), the reduced payoff per hour (parameter ι_3). Such behaviour is reflected in a reduction on assets, which is exacerbated by an onset shock on expenditures (parameter ϕ_3). A clear limitation is the inability of the model to accurately forecast selection into work according to potential labour income. In this case, despite of the wage penalty, labour income of those who

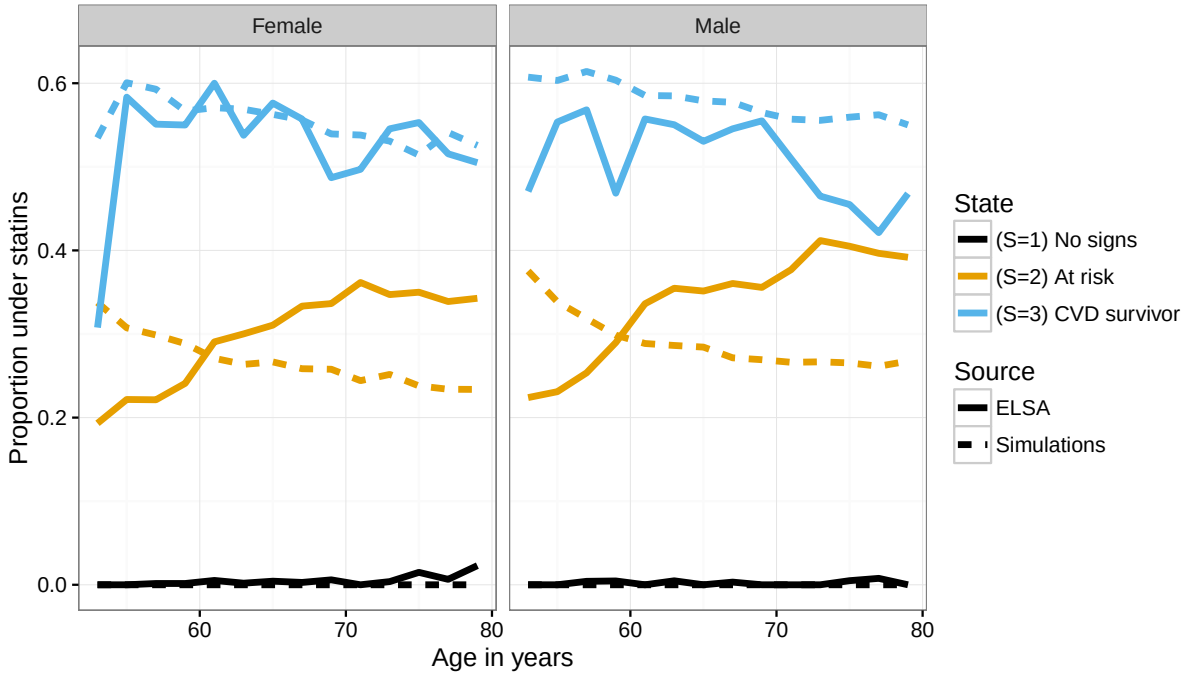
decide to participate is higher than those who have not suffered such events. This is evident in Panel A of Figure 3.15 for females.

Figure 3.12: Observed and Simulated Health Model Profiles by Health State

(a) CVD-risk index

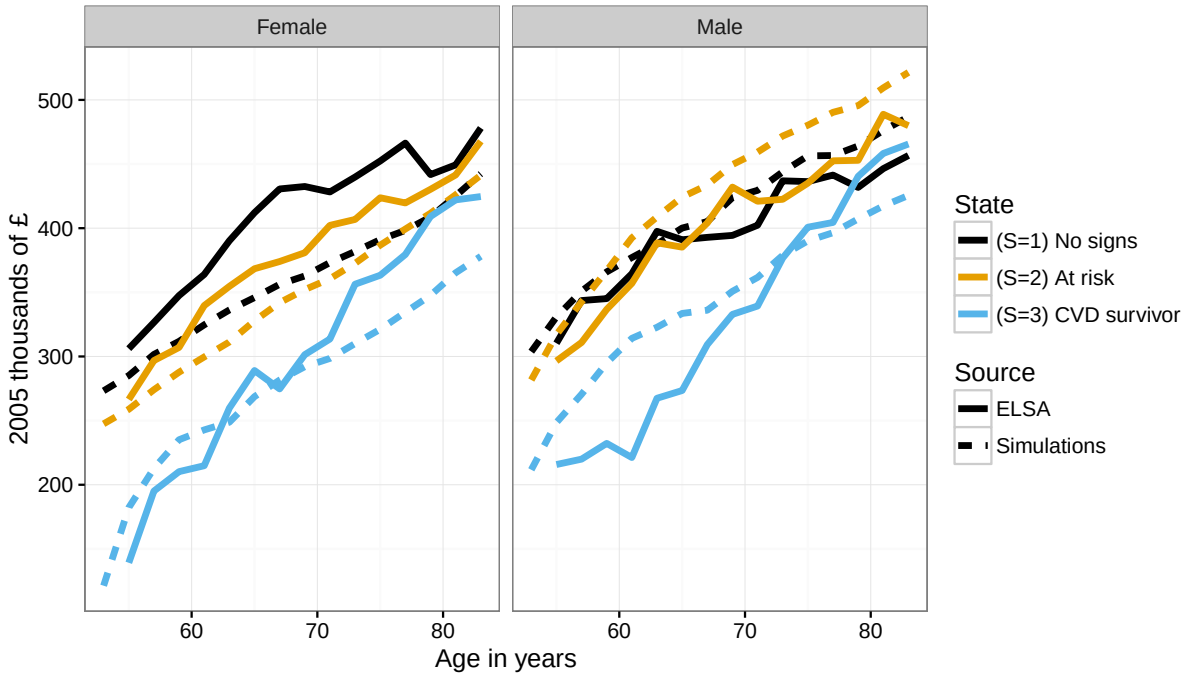


(b) Under Statins Treatment



Notes: Own calculations based on IFS derived variables on ELSA waves 1 to 6. Adjusted for Birth cohort, family size and regional unemployment.

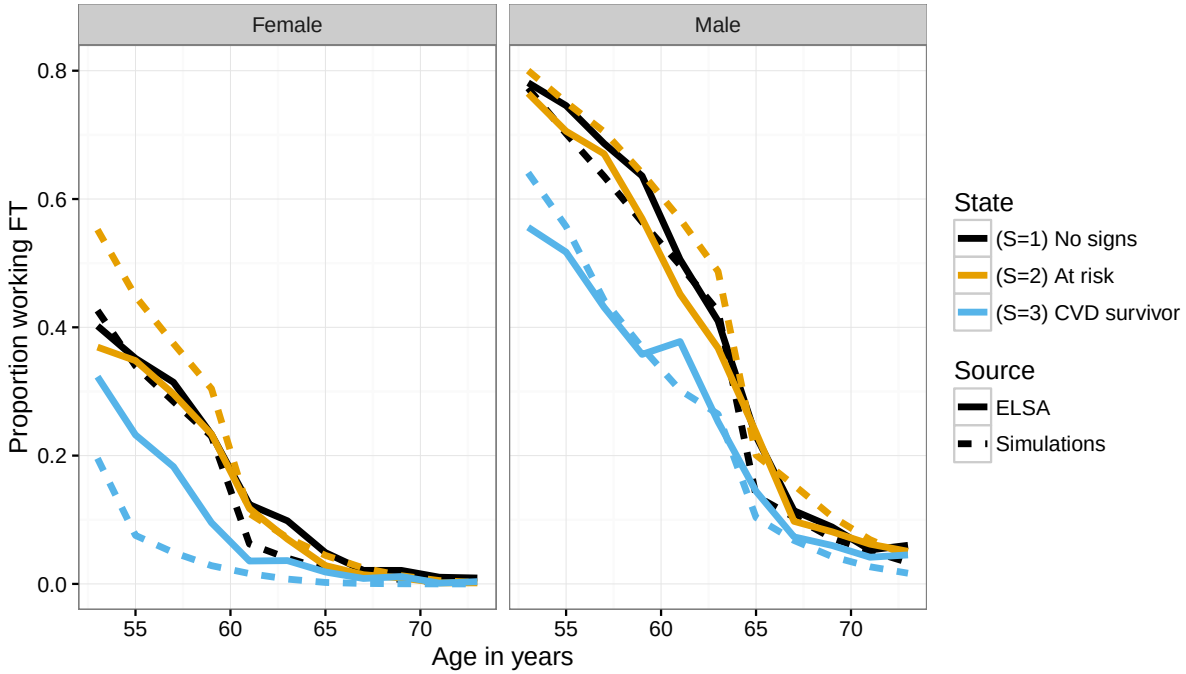
Figure 3.13: Observed and Simulated Assets Profiles by Health State



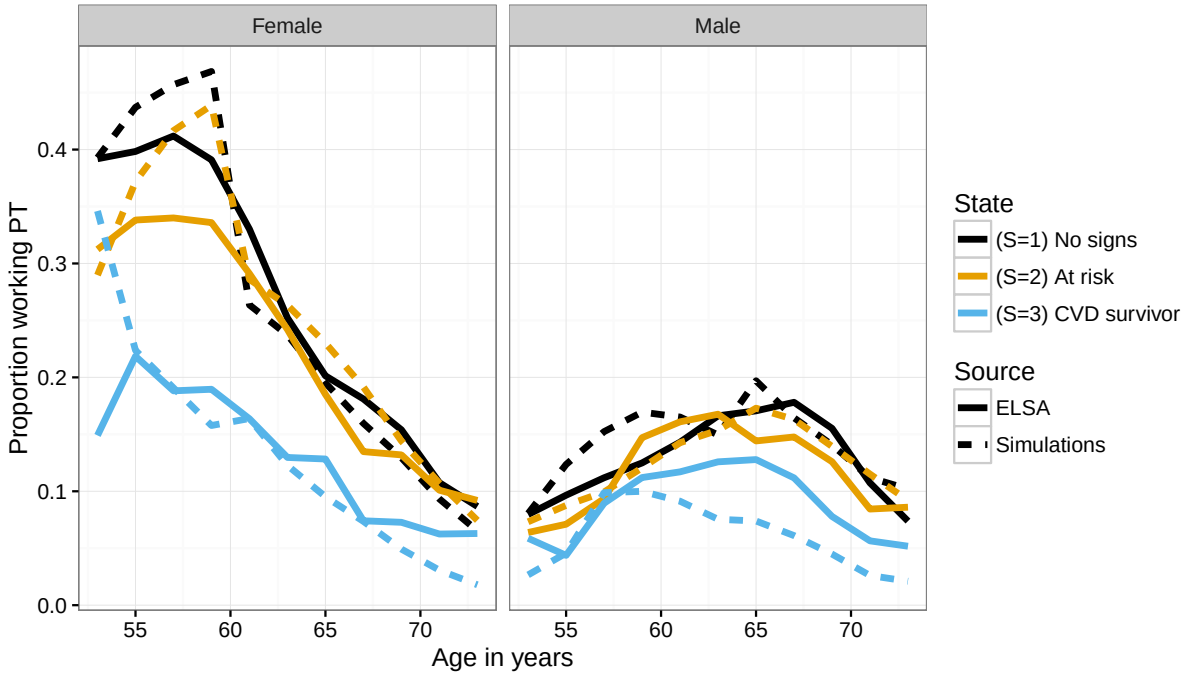
Notes: Own calculations based on IFS derived variables on ELSA waves 1 to 6. Adjusted for Birth cohort, family size and regional unemployment.

Figure 3.14: Observed and Simulated Labour Supply Profiles by Health State

(a) Full-time

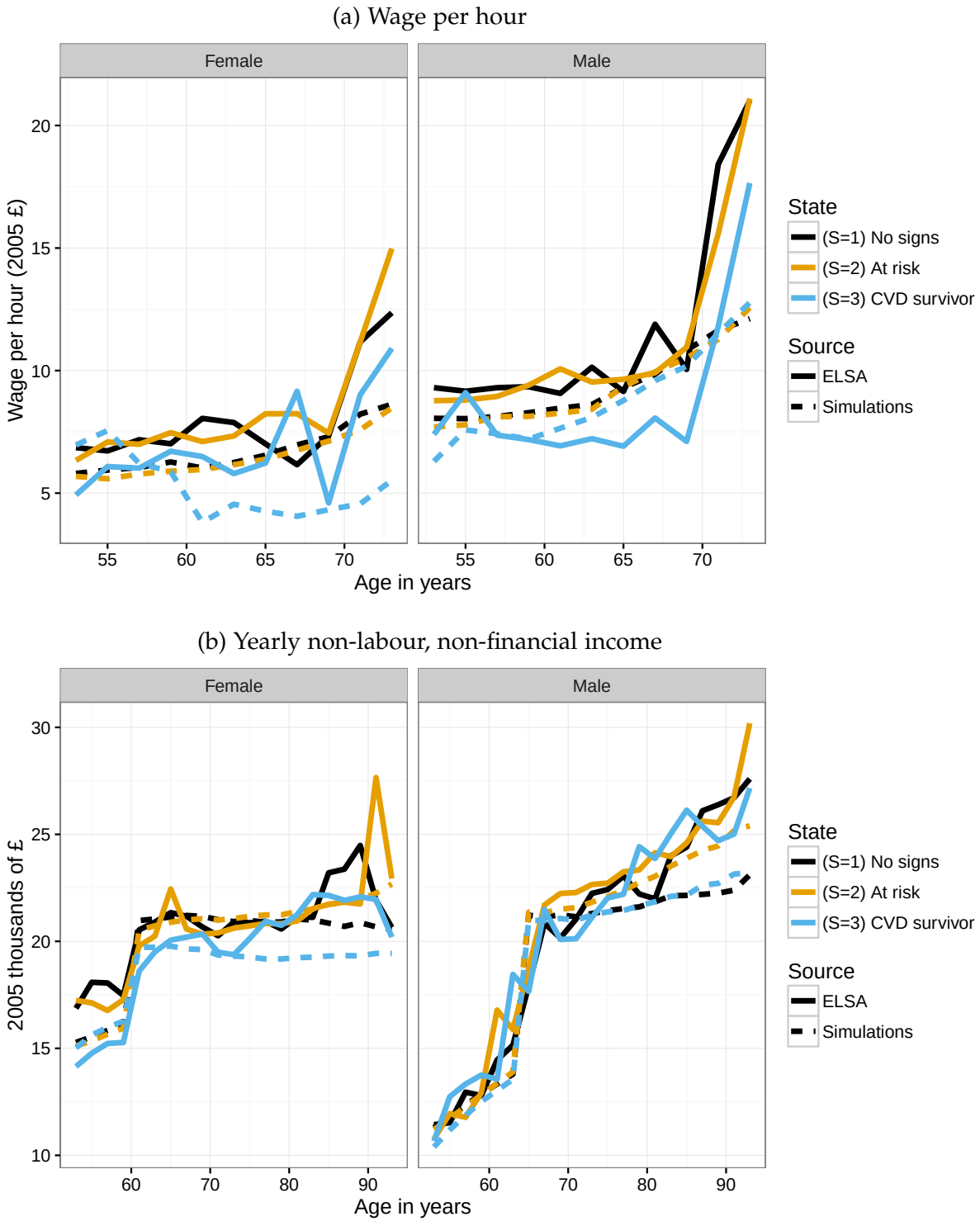


(b) Part-time



Notes: Own calculations based on IFS derived variables on ELSA waves 1 to 6. Adjusted for Birth cohort, family size and regional unemployment.

Figure 3.15: Observed and Simulated Income Profiles by Health State



Notes: Own calculations based on IFS derived variables on ELSA waves 1 to 6. Adjusted for Birth cohort, family size and regional unemployment.

3.8.2 The Value of Statins

Table 3.6 presents the main elements for an economic evaluation of a lifetime treatment with statins for primary prevention of CVD for age 53 onwards. In other words, here we are comparing the observed state of the world, where someone at risk of developing CVDs ($S = 2$) obtains a statins prescription with a probability given by Equation 3.6, against the counterfactual where statins are only available for those who already have been diagnosed with a CVD ($S = 3$).

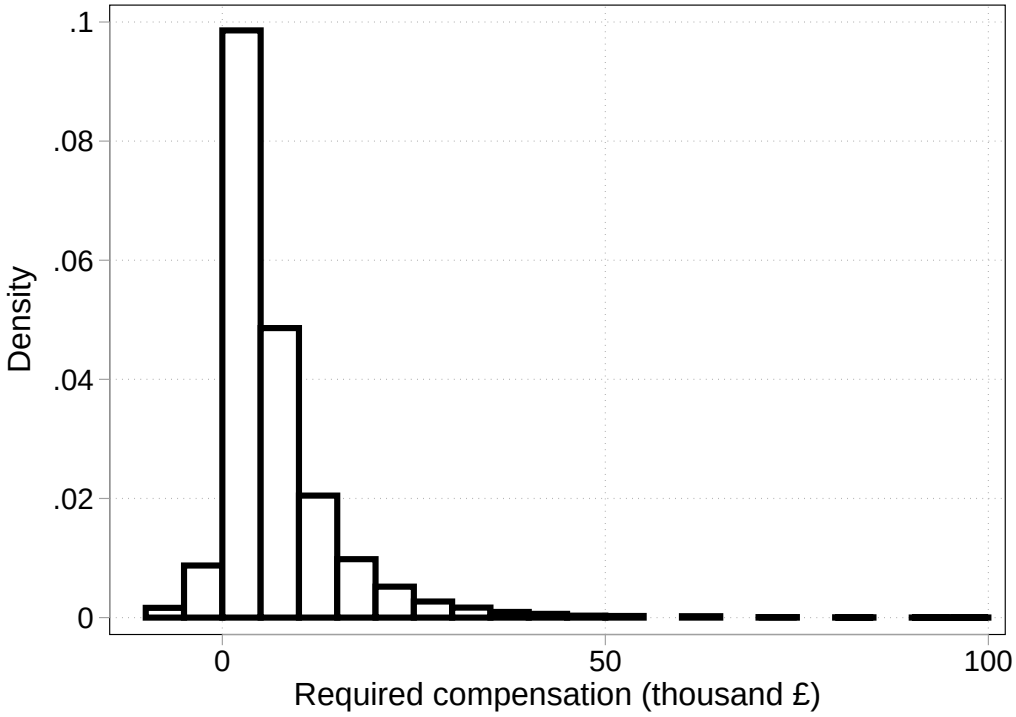
In first place, panel A presents the value of the drug taking into account individuals' utility. In this case, they compare their expected utilities under the current state of the world, against an scenario where preventive care medication is not available but that includes a monetary compensation. The amount of such compensation require to make both expected utilities the same is the willingness-to-pay. It is on average £5300. By aggregating such amounts[36] we are able to obtain the value of the drug if give an equal weight to all individuals: £79 billion 2005 pounds.

As expected, there is substantial heterogeneity on the required compensation. Figure 3.16.A presents its distribution according to the current health state of individuals. Figure 3.16.B shows that valuation is heterogeneous along many dimensions apart from gender. It shows the value of the intervention across age for a selected set of characteristics, leaving all the other state variables constant. It does depend on CVD-risk and health status, but also on the main drives of financial gains: income (education level in the graph) and assets. A remarkable result to discuss is that while the gains are monotone on CVD-risk index, that is not the case for assets. The reason for this is the existence of a bequest motive that depends only on such variable.

[36]ELSA sample is weighted according to age and gender in order to obtain a figure that is representative of England's population aged 53 and older by 2004. Notice that this procedure implies an utilitarian welfare function.

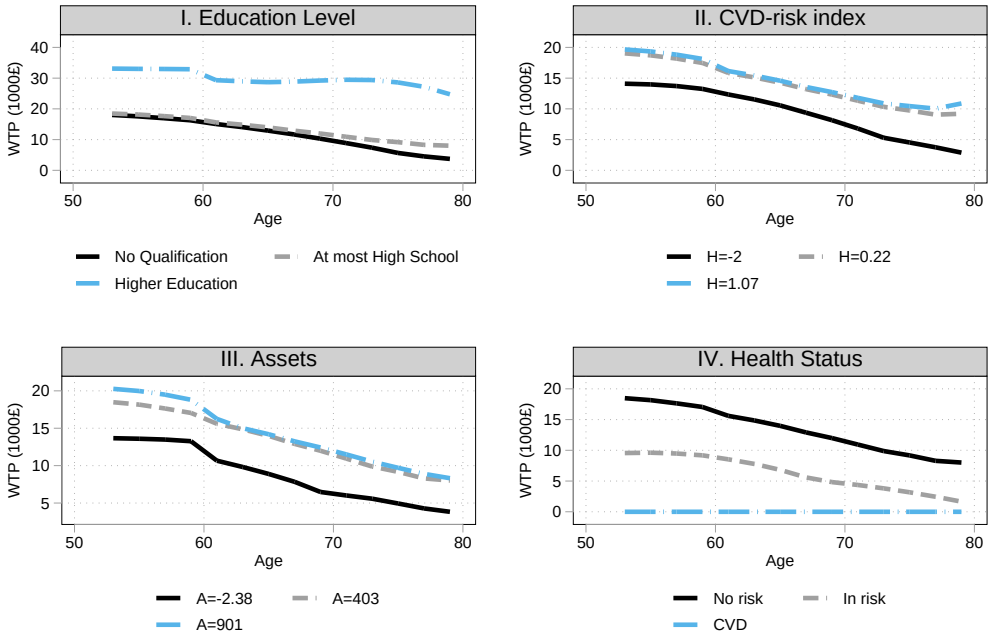
Figure 3.16: Heterogeneity on the compensating variation for primary prevention of CVDs

A. Distribution



Note: Based on 42547 simulations which replicate ELSA wave 2 age-gender structure

B. Calculated compensation with respect to age



Marginal WTP calculated at: Females, high school education, no detected risk of CVD, A=402.72, Y=1.11, H=-0.21, in 2002

The estimated value can be compared with a QALY based equivalent. Panel B of the table present the total amount of QALYs gained if we simulate the lives of all individuals considered in Panel A until their death, weighting their contribution in order to be representative on England's population in 2004. In total, nearly 3.1 million QALYs would be obtained through all the remaining years of life if we compare the current progression of health with a counterfactual without access to primary prevention statins. Shiroiwa et al. (2010) estimate the willingness-to-pay per QALY in the UK on £23.000, which is between the cost-efficiency threshold used by the NHS as a reference for the value of each QALY (£20 to £30 thousand). The total value of the drug would be between Â£61 to Â£92 billion. Hence, the compensating variation estimate would be nearly 12% higher with respect to the £23 thousand value per QALY.

Nevertheless, the most common economic evaluation approach for health is the cost-effectiveness analysis, presented in panel C. Details on how QALYs and costs are computed is presented in Appendix 3.C, following Ward et al. (2007) who did a systematic review of the cost-effectiveness of statins for primary prevention of CVDs using UK data. In terms of life years gained (LYG) and QALYs gained, estimates are smaller than to those presented by Ward et al. (2007). Undiscounted estimate for ICER with 50% non-compliance[37] is of £4641 while it is around £10.000 in Ward et al. (2007) for people of such age and with a compliance rate of 65% after 2 years and of 50% after 4 years. The ICER is normally interpreted as the amount of money required to obtain 1 QALY. If compared with NICE threshold of £20.000, this is a cost-effective treatment.

[37]Non-compliance refers to those individuals who are prescribed but do not take the drug. Here I assume that for each person prescribed and who takes the drug, there is another who does not take it. Effectively, this doubles de costs of prescription per patient. Discounting under usual rates from medical literature has little impact on the estimates, for this reason is not presented.

Table 3.6: The Value of Statins for Primary Prevention of CVD at age 53

	Male	Female	Both
Panel A: Compensating variation by 2004/05 (Wave 2)			
Required compensation (thousand £) (mean)	5.7	5	5.3
Required compensation (million £) (sum)	38965	40407	79372
Panel B: Total gains:			
Total gained QALYs (millions)	1.1	2	3.1
Value if each QALY is worth £20.000 (million £)	21729	40019	61748
Shiroiwa et al. (2009): WTP of £23.000 QALY (million £)	24988	46022	71010
Value if each QALY is worth £30.000 (million £)	32593	60029	92622
Panel C: Cost-Effectiveness:			
Undiscounted ICERs:			
1 Life Year Gained	1890	1312	1508
1 Quality Adjusted Life Year (QALY) Gained	2699	1900	2172
Undiscounted ICERs with 50% non-compliance:			
1 Life Year Gained	3865	2893	3221
1 Quality Adjusted Life Year (QALY) Gained	5518	4187	4641

Notes: Own calculations. Panel B is based on 41028 simulated individuals. The initial sample replicates ELSA wave 2 age-gender distribution.

In order to calculate the number of QALYs gained, it was only required to use the health component of the model. In this respect, the only difference with traditional analysis is the inclusion of the prescription equation. However, we can take advantage of the economic side of the model to add value to this exercise. The ICER definition in Equation 3.16 allows to approximate the cost-per-QALY gained, hence the same exercise can be used for any other variable, in particular labour supply variables. Table 3.7 presents such exercise. According to it, with every 1000 QALYs gained, approximately 1.8 individuals who would participate in the labour market one year in the observed state of the world, would not do so in a world without statins for primary prevention. The mean reason for this is that risk averse individuals might reduce their labour intensity if there is a less risky future ahead. An additional effect is that some individuals shift their labour supply in the intensive market as well: the treatment implies that some individuals decide to change their decision of participate part or full time. The last consideration is the productivity effect of statins. As part of the model estimation, CVDs were

found to impose a penalty of 30% on wage per hour. The table shows that with the treatment it is possible to avoid the burden of such penalty and increase wage-per-hour in £.3 per QALY gained. This result is enough for offsetting reduced part-time participation for men.

The result is that £1972 per year will be produced for men and of £17 for women. Population-wise, the net effect is an increase of £684 . Thus, if it costs £4641 to produce 1 QALY, it will come with £684 which is a potential source of income for the health system.

Table 3.7: Labour market implications of primary prevention

	Male	Female	Both
Undiscounted effect on labour supply per 1 QALY gained			
Persons working per year	.089	-.049	-.0023
Persons working FT per year	-.022	-.065	-.05
Persons working PT per year	.11	.016	.048
Working Hours per week	1.3	-2.3	-1.1
£ from yearly labour-income	1972	17	684

Notes: Own calculations. Panel B is based on 41028 simulated individuals. The initial sample replicates ELSA wave 2 age-gender distribution.

3.8.2.1 Bequest motive and the Value of a Statistical Life (VSL)

A final element for discussion is the interpretation of previous results as a willingness-to-pay estimate. As discussed before, in this model difference between life and death is based on the bequest motive. This is a central difference with respect to value-of-life literature that traditionally calibrates this difference using occupation-based estimates of the value of a statistical life. Table 3.8 presents an equivalent exercise for the simulated dataset in 2004/05: which is the maximum amount of money that an individual would give up in order to attain a reduction of 1/1000 on the odds of death for the rest of his life? The average is around £1050, or £1582 for individuals younger than 60. This implies a VSL of £1.05 million, or £1.582 million for those aged 60 or younger.[38] Such figure is similar to the average value of preventing a casualty in 2004 estimated by the Department for Transport in the UK, £1.4 million (DfT, 2005). However, is below VSL estimates for the UK

[38]Willingness-to-pay for risk reductions is not linear, this is just a standard normalization of the estimates in the VSL literature.

derived from risk-compensating premium in wages. Following Viscusi and Aldy (2003), the implicit VSL from Sandy et al. (2001) is of £7.65 to £99.6 in 2005 prices.[39] Thus, this paper presents a lower bound of the willingness-to-pay measure that would be obtained under such literature.

Table 3.8: The Value of a Statistical Life (VSL)

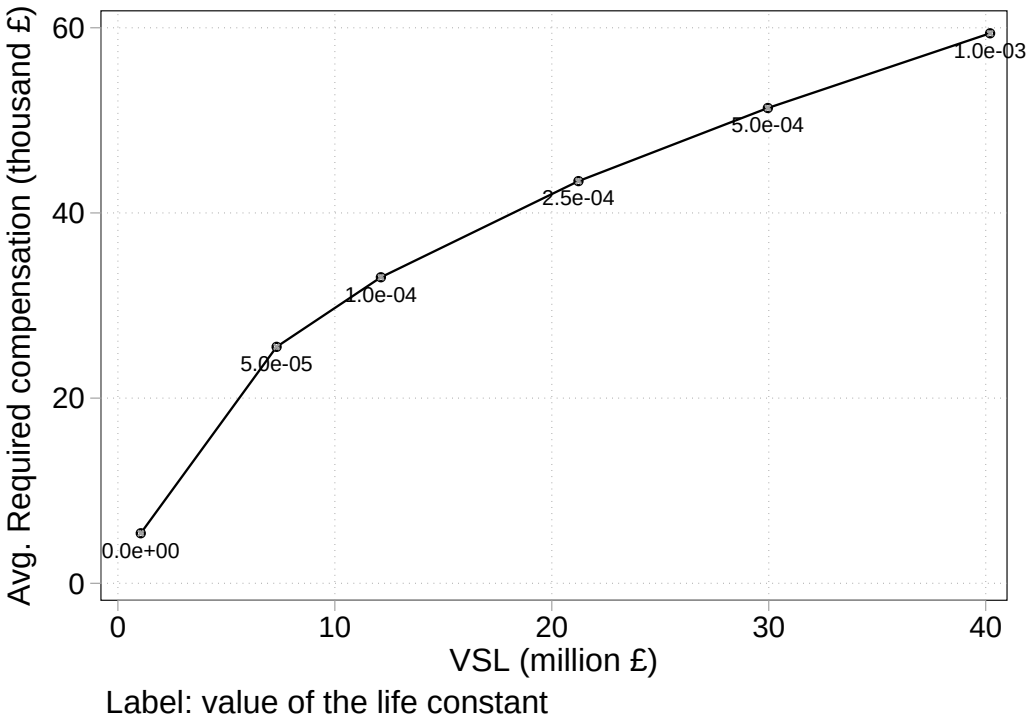
	Male	Female	Both
Required compensation (million £)	1.193	.9339	1.05
Required compensation (million £), $age \leq 60$	1.875	1.337	1.582

Notes: Own calculations. Based on a counterfactual reduction of 1/1000 on the odds of death for the remaining life for a simulated sample of individuals that replicates ELSA Wave 2 age-gender composition.

While the model design implies a VSL, it is still possible to use (Hall and Jones, 2007) strategy and introduce a constant to Equation 3.7, which plainly would represent the difference on utility between of life and death in any given period. This one can be calibrated it in order to match a particular VSL value. Figure 3.17 presents the resulting average value of statins under different implied VSL. The point on the lower left is the present estimate. It shows that the value of the treatment greatly depends on such figure, and after a £10 million figure, the relationship is almost linear.

[39] $5.7 - $74.1 millions, 2000 USD. Translated into 2000 £using an exchange rate of 1.51 £/USD.

Figure 3.17: Simulated Labour Supply under SPA=60 and SPA=65



3.8.3 Retirement age and the value of preventive care

One of the central implications of the model presented in this paper is that the value of a medical innovation depends on the institutional arrangement of a society. Specifically, the model considers the role of social protection when considering non-labour income as a function of retirement age in Equation 3.12. As discussed when the equation was introduced, there are two retirement regimes for women in the model. This is because the 1995 Pensions Act implied a radical increase on retirement age for women during our study period, from age 60 to age 65 depending on the date of birth. A natural question is that given everything else equal, women from these two generations value differently the introduction of statins for primary prevention.

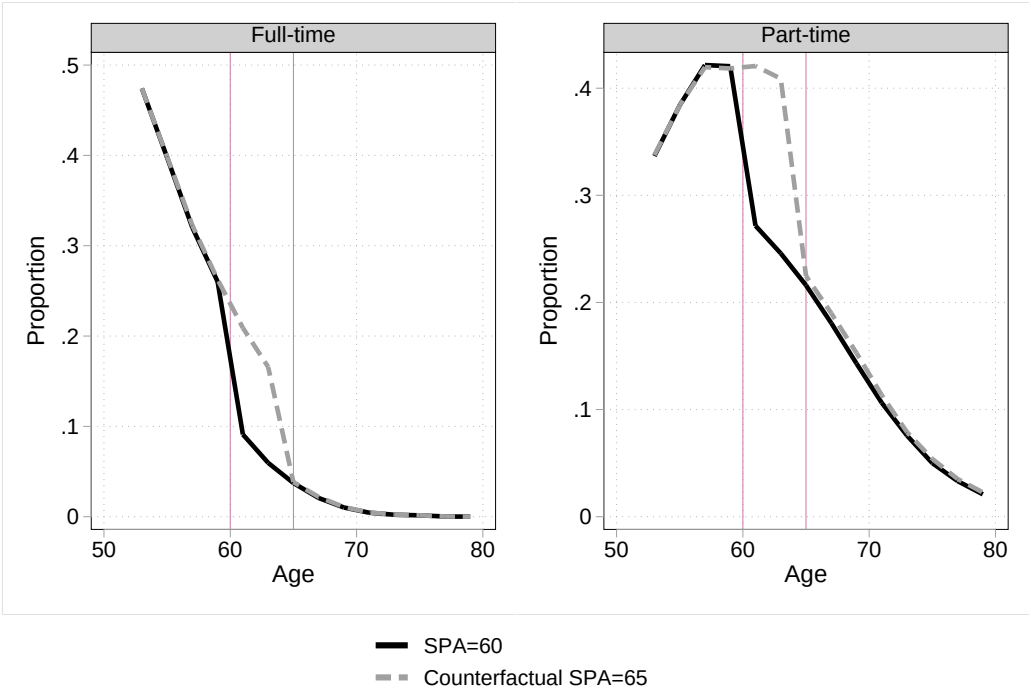
In other to answer this question, the value of statins is computed for women considering both retirement regimes separately. For this, a counterfactual scenario is constructed: all simulated women in wave 2 are switched to the new 65-years-old standard pension age.[40] Figure 3.18 presents the simulated age profiles for

[40]All women aged 52 and over at wave 2 would be eligible for the former retirement regime

labour market participation under both scenarios. It shows that the model predicts an increase on full-time labour supply participation after the age of 60 due to this retirement-age reform. In total, there is an increase of 59.26% in participation in the labour market between ages 60 and 65 .

Table 3.9 presents the calculated value of statins in both scenarios. The first column shows results under the regime applicable to women in wave 2, and the second the counter-factual scenario. First, willingness-to-pay is slightly higher in the 60-years-old SPA regime than in the new alternative, but there are extra gains of 22 times of the income produced when obtaining 1 extra QALY with the medication.

Figure 3.18: Simulated Labour Supply under SPA=60 and SPA=65



Predicted increase of 59.26% on labour market participation between ages 60 and 65

(SPA=60) as they were born before 1954.

Table 3.9: The Value of Statins and Retirement Age

	Female SPA=60	Female SPA=65
Panel A: Compensating variation by 2004/06 (Wave 2)		
Required compensation (million £)	40407	39918
Panel B: Undiscounted effect on labour supply per 1 QALY gained		
1 person working per year	-.049	-.0055
1 person working FT per year	-.065	-.06
1 person working PT per year	.016	.054
1 Working Hour per week	-2.3	-1.3
£ of yearly labour-income	17	399

Notes: Own calculations based on 22720 simulated individuals that replicates ELSA Wave 2 age-gender composition.

3.8.4 Counterfactual technology scenarios

Our analysis above depends crucially on the evolution of CVD-risk index and medication prescription when individuals are at-risk of developing CVDs (S=2). With the model we can analyse how the compensating variation would change if we modify the two basic pieces of this process: how effective is the drug and how likely is that one gets a prescription. These two channels allow for an improvement on health of the population from different perspectives. The first one is a pure medical innovation result, while the second can be attained under government policies.

With the model we can consider alternative values for the effect of statins in the CVD-risk index. Figure 3.19.A presents the average required compensation for not having access to the drug (vertical axis) according to its effect on reducing cholesterol (horizontal axis). The vertical line corresponds to the value ω_4 (Equation 3.5 calibrated from RCTs discussed in section 3.7.1.2, and the corresponding average compensation is the value presented in Table 3.6 for both men and women. It shows that the value is increasing on the effect of the drug (a more negative ω_4), as expected. If there is a drug that doubles current statins efficiency, its value will be 69% larger than the estimated one for statins. This calculation involves both the drug effect on people at-risk of CVD and those who already survived at least one of such events. The diminishing returns are produced by the functional forms of the latent indexes presented in equation 3.2. As the risk-index H enters exponen-

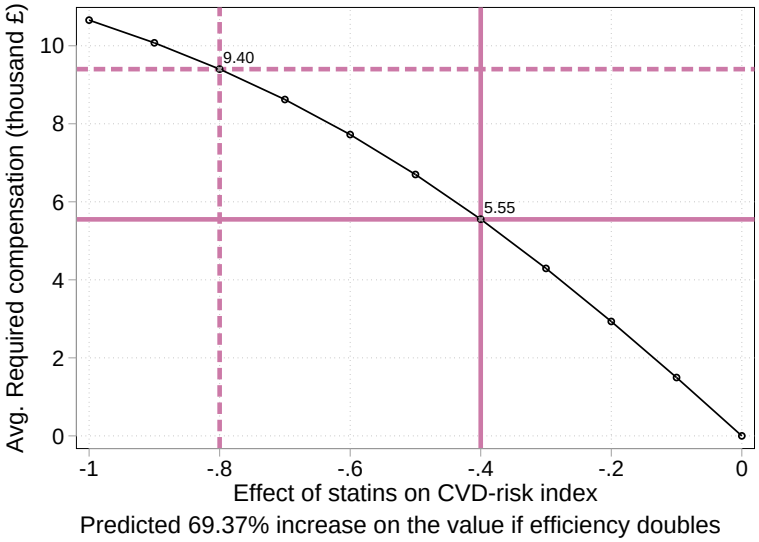
tially on the transition probabilities, it implies that a reduction when the index is high is much more important for cutting the odds of bad health events that when the index is low.

Figure 3.19.B considers the value under different values of the parameter ζ_1 from Equation 3.6. This parameter governs the probability to get a prescription for statins when regardless of CVD-risk and gender, thus the horizontal axis considers the derived average probability rather than the value of the coefficient. In order to consider only the primary-prevention extent, parameter ζ_3 is adjusted in order to keep the probability of prescription constant for those who already suffered a CVD event. A striking result is that increasing the probability to access a prescription might increase the value of the drug notoriously. For instance, the model predict that the value of a coverage of nearly 80% for those at-risk of CVD events would be higher than a potential drug that is more than twice as effective in reducing the CVD-risk index. This is a suggestive result, because the model does not considers differently those individuals who are at-risk but should not take statins.[41] More informative is to consider an impact on prescription probability similar to the increase obtained in the quality and outcomes framework for individuals living in small relative to large GP practices. A 3 pp. increase on the odds of prescription is obtained (see Appendix 3.F). And additional advantage is that the model is estimated taking into account his exogenous variation. Then, an increase of 3 pp. on the probability to be treated results on an increase on the value of the treatment of 24.6%. This implies that at such point there is an elasticity of the value with respect to the prescription probability of 0.59.

[41] Because they interact with other medication increasing the risk of adverse effects. See `https://www.gov.uk/drug-safety-update/statins-interactions-and-updated-advice-for-atorvastatin`.

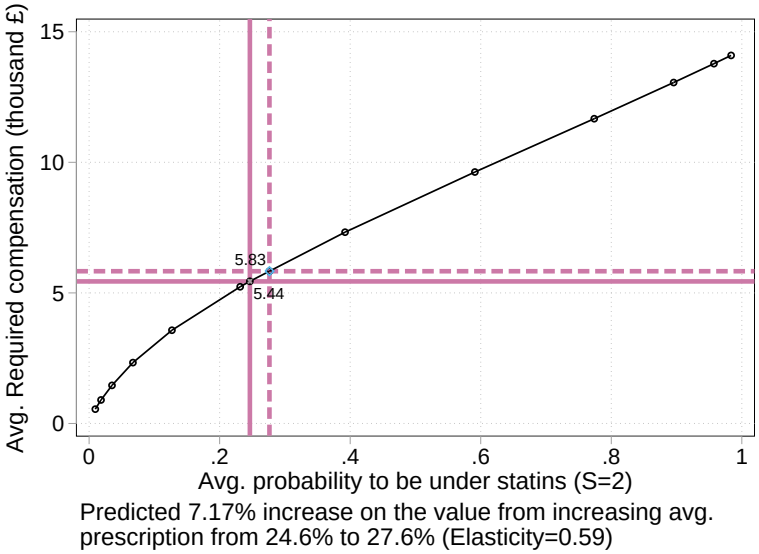
Figure 3.19: Counterfactual technology scenarios

A. Value as a function of the effect of statins on CVD-risk



Note: Based on 42610 simulations which replicate ELSA wave 2 age-gender structure. Counterfactual value of the drug was computed by modifying the value of parameter ω_4 (Equation 3.5). The vertical line corresponds to the estimated value of the parameter calibrated as discussed in section 3.7.1.2.

B. Value as a function of the average probability of prescription in (S=2)



Note: Based on 42610 simulations which replicate ELSA wave 2 age-gender structure. Counterfactual value of the drug was computed by modifying the value of parameter ζ_1, which is the constant of the prescription equation (3.6). The vertical line corresponds to the value of the estimated parameter.

3.9 Conclusion

This paper introduced a framework for assessing the economic benefit derived from the adoption of a health-care technology. It introduces a life-cycle model for savings and labour supply decisions into a Markov health progression model.

It is found that statins therapy for primary prevention, a drug for reducing the odds of CVD onset, had a pure economic value of £79 billion by 2005. This figure is 12% higher than the value that would be obtained if in the calculation of benefits we only to consider a rate of £23.000 per QALY gained, which represents the willingness-to-pay per QALY in the literature.

As is widely accepted in the clinical literature, primary prevention is cost-effective in comparison with NICE standards. In terms of the drug, it costs £4641 to gain one QALY (undiscounted, 50% non-compliance), but by doing so, it increases labour-income by £684 per year. Such additional gains depend on the trade-off between the direct avoidance of CVD events and a reduction in the precautionary motive for asset accumulation. This is reflected in the sharp differences by gender; the labour income gains of using statins for primary prevention are £1972 per year for men and £17 for women. The role of a recent retirement age increase for women was also considered, and almost no difference was found in the value of the drug.

A relevant question is the relevance of the economic benefit of a health intervention when considering its adoption. In particular, labour income gains will also be reflected in extra taxable income, which should be deduced from the cost figures. Let us consider a back-of-the-envelope calculation assuming that the total net of tax £684 comes from a pure increase in taxable income. For an individual subject to deductions of 25%, assuming than half of those resources are directed to the health service, the total gains for the health system would be of £327.

3.A Data profiles and cohort effects

3.A.1 Adjustment procedure

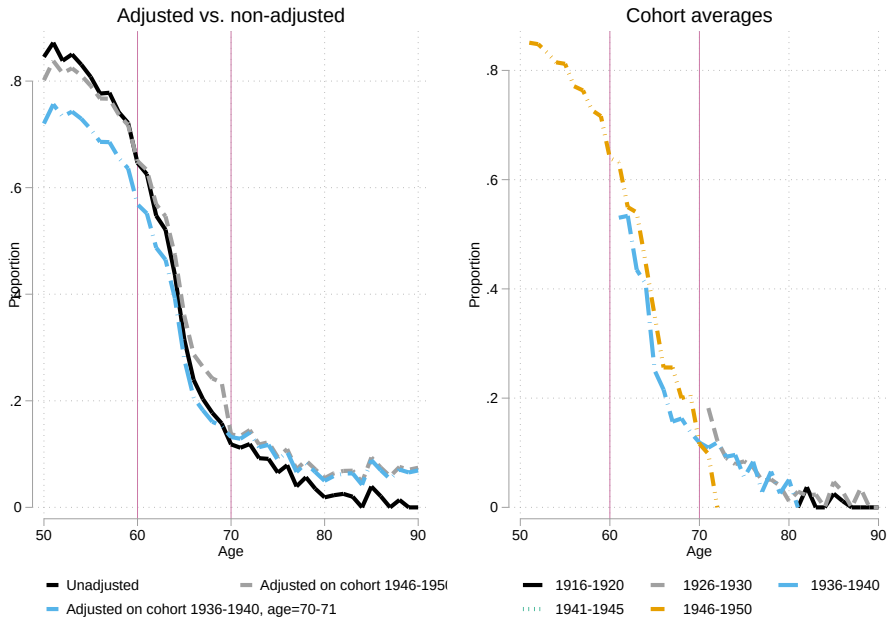
First, Equation 3.19 is estimated from the data, where Y corresponds to the relevant variable to be adjusted, for individual i at period t living in region r. Second, the predicted effects of unemployment, family size and cohort are removed from the data, while keeping the predicted values from the age-health status interactions. Third, the predicted effect of an unemployment rate of 5, a family size of 2 and of cohort 1946 are added to all observations. Notice that cohort effects are allowed to be different from age 70 onwards. The reason for this is to avoid high levels of participation on the labour market after SPA.[42]

$$\begin{aligned} Y_{itr} = & \sum_{s=1}^{3} \sum_{k=50}^{80} d_{itr}^{ks} \times (age_{it} = k) \times (S_{it} = s) \\ & + \sum_{f=1}^{F} d_{itr}^{f} \times (famsize = f)_{itr} + \Pi_U \times U_{tr} \\ & + \sum_{c=1}^{C} d_{itr}^{C1} \times (cohort_i = c) \times (age < 70) + \sum_{c=1}^{C} d_{itr}^{C2} \times (cohort_i = c) \times (age \geq 70) + u_{itr} \end{aligned} \tag{3.19}$$

Figure 3.20 illustrates how profiles are affected by this procedure. The graphs on the left show that it induces a notorious divergence on labour participation after SPA. This is driven by the difference between cohorts presented on the right graph: younger generations are more likely to be working. Also, notice that the adjusted profile is a normalization with respect to a given cohort at a specific time. Thus, if we consider a young cohort as the base, the procedure implies a big difference at old ages. I have consider this cohort as the benefits of a reduction of CVD-risk is aimed towards individuals below the SPA.

[42] Younger cohorts are observed to work more as their are observed before SPA, which implies an upward correction on those cohorts which having retired for most of the observed time. In practice, without the differential cohort effect by age, nearly 20% of individuals aged 80 and over are suggested to be working while the observed figure is almost 0%. This is attenuated with this strategy, as shown in Figure 3.20.

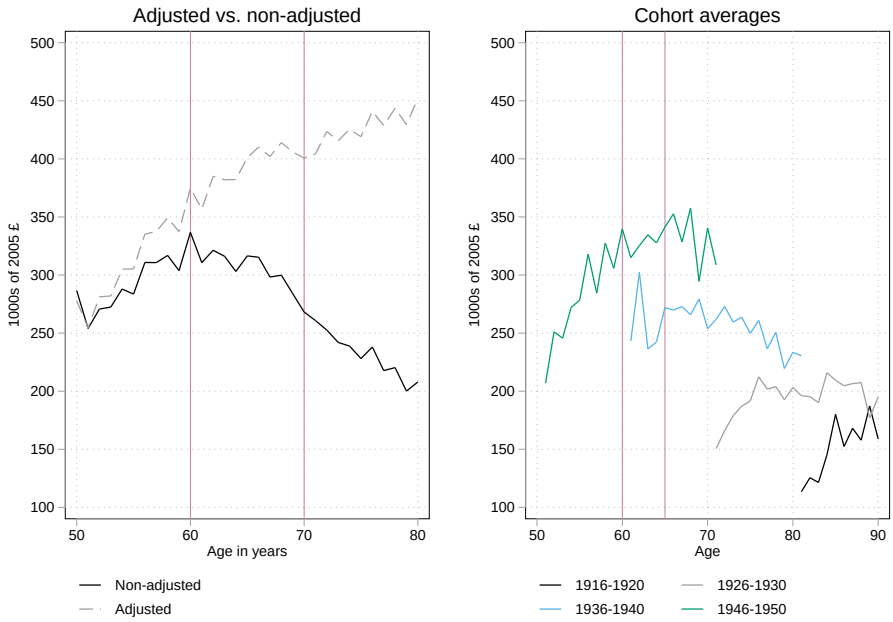
Figure 3.20: Observed Income Profiles by Health State



Notes: ELSA data profiles for individuals aged 50 to 80. Cohort, unemployment and family size effects were removed in the 'adjusted' series.

The most affected variable is assets. Figure 3.21 shows how a hump-shape average is generated if no correction is applied to the data. However, if we group data by birth cohort, an increasing pattern is observed. Hence, the correction has a strong implication on the profile. One the reasons behind is the strong cohort effects on housing ownership in the UK, described by Banks et al. (2012): older cohorts are notoriously less likely to own their dwelling with respect to younger ones due to a Government policy that allowed individuals to buy the council house that they were renting. However, as discussed in Appendix 3.A.2, the pattern is also present in non-housing assets. This is central as the value of assets late on life identify the bequest motive with is the main driver of the statistical value of life in the drug value calculations. If individuals rapidly de-accumulate at the end of life, it means that the bequest motive should be small.

Figure 3.21: Assets by Birth Cohort



Notes: ELSA data profiles for individuals aged 50 to 80. Cohort, unemployment and family size effects were removed in the 'adjusted' series.

3.A.2 Assets

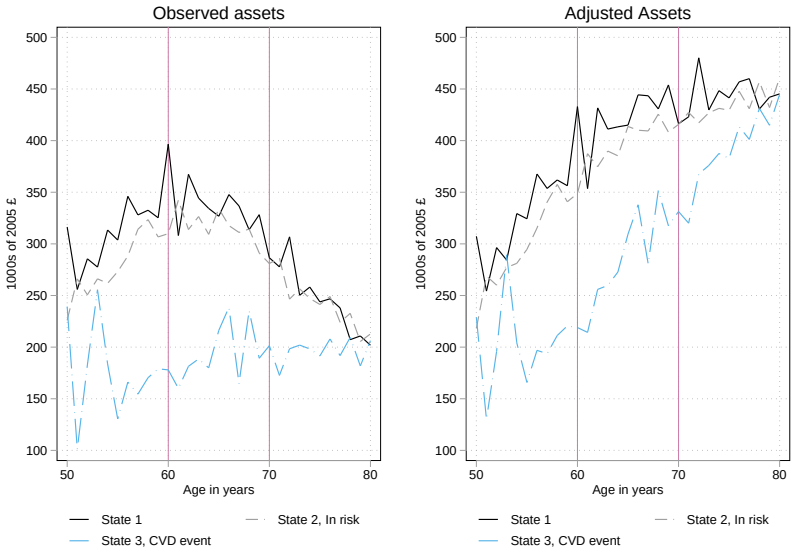
Assets include all liquid and non-liquid financial positions reported by the household excluding pension wealth. The main reason for this is that its returns are already included in the non-labour income equation (Eq. 3.12). Also, before 2015 it was not possible to borrow against pension-wealth before age 55 without paying a tax of 55%. After it, individuals can get up to 25% tax-free. In this model, as pension release is not the main focus of the analysis, it is assumed that individuals cannot access to their pension pot at all.

An important element of interest in housing. One of the potential reasons for the increasing pattern of assets with respect to age could be house prices. If housing wealth increased notoriously during all the decade for all cohorts, I might be confounding the savings pattern with the commercial hike on prices. Figure 3.22 presents assets measures both with and without housing, before and after adjusting them from cohort and other variables effects. In both cases the

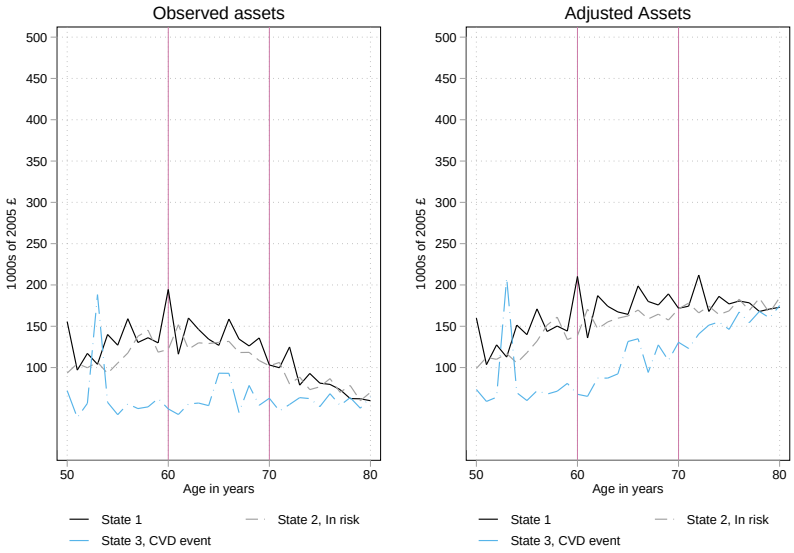
hump-shape of assets is transformed into an age-growing trend, meaning that both housing and non-housing assets are corrected.

Figure 3.22: Assets Correction

(a) Including housing



(b) Excluding housing



Notes: ELSA data profiles for individuals aged 50 to 80. Cohort, unemployment and family size effects were removed in the 'adjusted' series.

3.A.3 Non-Labour Income

Non-labour income is the total non-financial household income minus employment and self-employment earnings of the respondent. Total household income in ELSA includes employment, self-employment, benefits, state pension, annuities or private pension, assets income and others. Asset income is not included as asset's law of motion equation (Eq. 3.9) captures capital gains directly, and not as a part of the non-labour income equation (Eq . 3.12). In ELSA, asset income includes: interest from savings, TESSA, ISA or National Savings; income from Bonds, PEPs, Shares, Trusts, Bonds and Gilts. Also, it considers gains from renting property, or returns from businesses and farms.

The non-labour income process of this article makes several assumptions on its functional form. This reduce-form equation roughly summarizes the benefits and retirement system in place.[43] This appendix explains some of them in detail.

1. *To be working has different returns before and after SPA:* ELSA data shows that in aggregate terms, retirement age drastically change the profile of labour and non-labour income. Part of this can be backed-up with the design of benefits system: state pension replace other type of benefits as job-seeking allowance or income support. Also, the fact that couples retire together also explains why after SPA there is no difference on non-labour income.

2. *EDUC variable only affects the mean of both working and non-working profiles:* While potentially many aspects of non-labour income might differ according to past economic background, evidence shows that profiles are almost parallel for different levels of EDUC.

3. *Health states do not influence the profiles:* This is probably the most surprising element of this model, however it is a result of the data. As can be seen in the figure, there is almost no difference on the profiles across health states. Although not presented, similar results are found for separate working and not-working profiles.

[43] See Bozio et al. (2010) for a detailed exposition of the main characteristics and changes of the system.

3.B Matching GP practices and ELSA respondents

Figure 3.23 presents an example for a postcode in London (NatCen office). The first 8 practices are within 1 Km of the postcode while 9th and 10th are around 1180 meters from it. As a result, while the weight for the first eight is $\frac{1}{8+2/1.18}$, for the last two are $\frac{1/1.18}{8+2/1.18}$.

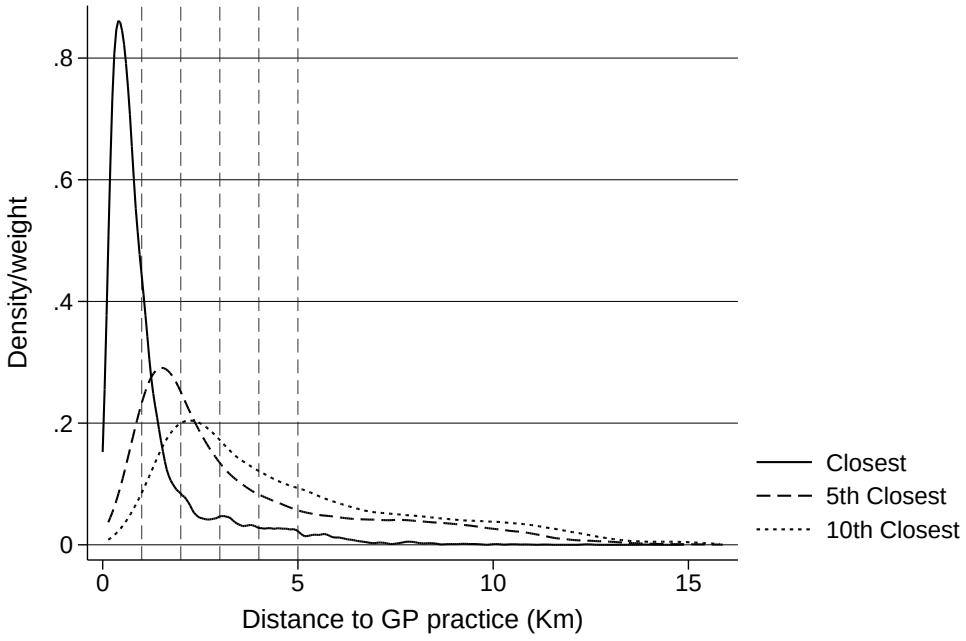
Figure 3.23: ELSA and QOF linkage example

The goal of this procedure was to stablish a link between ELSA respondents and nearby GP practices. This is possible as for waves where there is a nurse visit, ELSA has a record of the postcode of residence of the respondent. For other waves, the last available residence postcode was assumed. Under this criteria for 97% of the respondents (present in any ELSA wave) there is a valid postcode. For the case of GP practices, their postcode is publicly available in the directory *epraccur*, published by the HSCIC. As a result, it is possible to calculate the distance between a given postcode centroid and all its nearby practices' postcode centroids. Such calculations were made using the geodist routine (Picard et al., 2012) for STATA 13. This procedure was carried out in NatCen's secure data enclave in order to protect confidentiality of ELSA respondents. Finally, it was possible to provide a

measure for 87% of the respondents.

Table 3.10 tells us that the example in the data description is far from the usual case as expected given that it is based in central London. In general within 1 Km there is only 1 or 2 GP practices, a number that can go up to 19 in densely populated areas. Moreover, Table 3.11 shows that the mean distance to the closest practices is around 700 to 1200 meters, but it can go up and beyond the 15Km boundary in some rural areas. The most common distance to the furthest practice is beyond 3.7 Kms, if available. This can be. Back in Table 3.10, we can see that the average scenario is that the 10 closest practices are located within 3 Km. This is also reflected in Figure 3.24, which is a graphical version of Table 3.11.

Figure 3.24: Distance to the nearest GP practice



The selected weighting of practices' information is based on the previous descriptive information. The goal is to give more importance to the nearest practice but still considering the possibility of choice when it is available. Figure 3.25 presents the density of the 10 closest practices and the weight according to distance. By setting it fix to 1 Km, I am assuming that individuals will have a similar preferences for the average two practices that are quite close to their home, but will still consider those nearby. On the other hand, in rural areas, a considerable more weight will be given to a practice in 5 Kms rather than 10 Kms.

Given that all the elements (maximum distance, number of practices and weights) are arbitrary, I considered several alternative. Figure 3.25 shows how the weighted average list size measure varies under the different elements. While there are differences, in general all of them are highly correlated.

Figure 3.25: Weights allocation

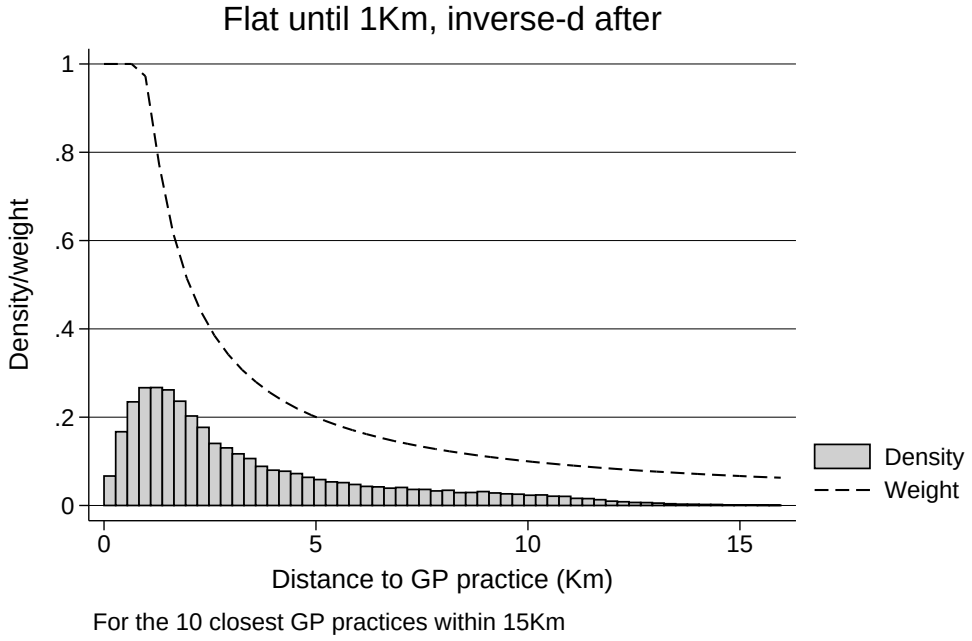


Table 3.10: Number of GP Practices Given a Distance Buffer

Practices within	Mean	Median	Max
1 Km	1.735	1	19
2 Km	5.264	3	50
3 Km	10.228	6	98
4 Km	16.468	9	151
5 Km	23.954	13	207

Source: own calculations based on ELSA respondents' postcode and *epraccur* GP practices' postcode.

Figure 3.26: Alternative criteria

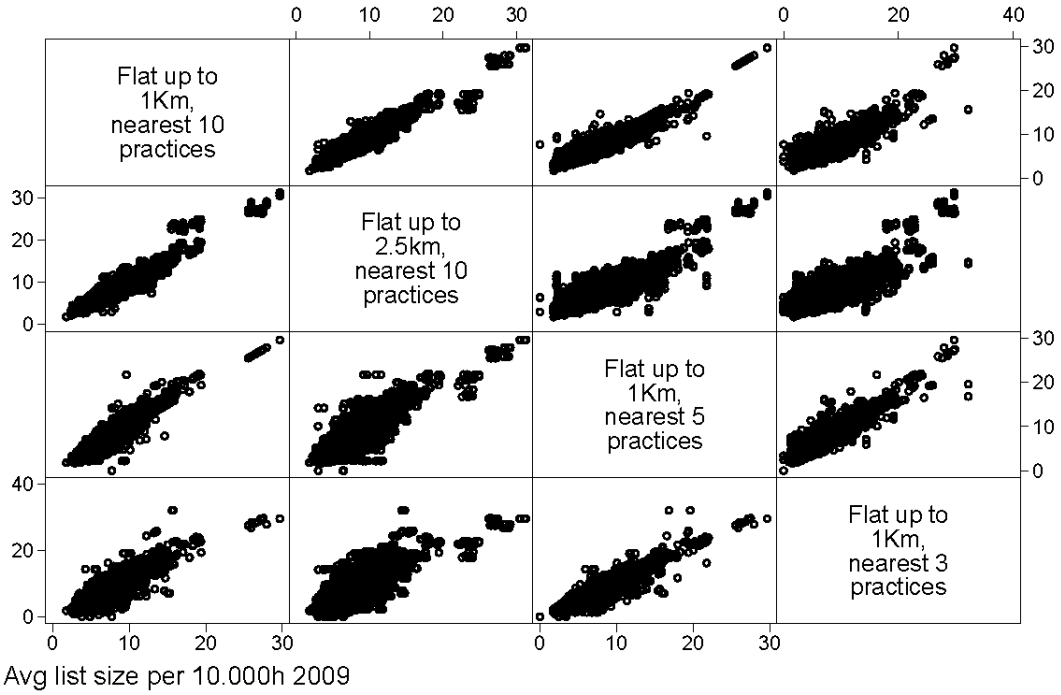


Table 3.11: Descriptive Statistics of the Distance to the 10 Nearest GP Practices

Practice Number	Number Available within 15 Km	Mean Distance	Percent 25 Distance	Percent 50 Distance	Percent 75 Distance	Minimum Distance	Maximum Distance
Closest	8878	1.171	0.408	0.714	1.269	0	14.886
2	8865	2.035	0.744	1.244	2.404	0	13.502
3	8823	2.680	1.027	1.686	3.475	0	14.916
4	8779	3.228	1.278	2.124	4.355	0.086	15.020
5	8706	3.610	1.484	2.446	4.934	0.143	15.878
6	8574	3.927	1.686	2.790	5.467	0.143	15.690
7	8467	4.233	1.869	3.118	5.892	0.143	15.833
8	8318	4.433	1.999	3.339	6.104	0.143	15.874
9	8178	4.597	2.149	3.526	6.358	0.143	15.963
10	8051	4.764	2.282	3.750	6.579	0.143	15.802

Source: own calculations based on ELSA respondents' postcode and *epraccur* GP practices' postcode.

3.C Cost-effectiveness details

Ward et al. (2007) summarised the main models for statins in the literature and developed a Markov model for the NHS framework, the ScHARR model. In order to provide a comparison as close as possible, I will follow their costs and benefits calculations. In their model, individuals are heterogeneous on their initial CHD-risk (CVD without stroke or TIA) and gender. Every year patients can transit into several CVD events or death, until they reach age 100. Age-dependent transition probabilities between those states are derived from the Health Survey of England, which is the same base population as this study. It is assumed that health-care or life-style advice do not differ between treatment and control scenarios. Also, that side-effects or the odds of adverse events associated to the medication are negligible. Both assumptions are also in place in the model presented in the previous sections.

With respect to the health-state *utilities* $\mathbb{U}$,[44] baseline levels were estimated based on the EQ-5D questionnaire and CVD-states taken from the literature. Given that Ward et al. (2007) considers several CVD illness separately while this model only have one CVD event state, an adjustment is required in order to obtain costs and associated utilities. I will get a single number by weighting according to the their figures on the distribution of primary events. Given these considerations, Equations 3.20 and 3.21 show how *utility* values and costs are assigned in the model in order to estimate the cost of gaining one QALY. One drawback is that the present model does not distinguish deaths from CVDs or other causes, as a result it is imposed that only deaths of previous CVD-survivors are linked to a different cost. This is likely to underestimate the true cost reduction of the intervention.

A final cost to consider is the value of statins treatment. At 2005 prices, Ward et al. (2007) costs of prescriptions are £281 per year, £127.5 for monitoring the first year and £33.42 for the subsequent ones. Their estimates are based on the mix of statin molecules used in the NHS by 2003, and of the usual tests required by guidelines at that point. These figures do overestimate the cost of the treatment as the mix of molecules and the intensity of generics prescriptions changed notoriously

[44] These *utilities* are different to the output of the utility function presented in the model as they are not designed to represent economic choices but quality of health. See footnote 20 for more details.

during the past decade. Nevertheless, these figures are going to be preserved in the present analysis in order to make both models comparable.

For this exercise, for both scenarios, with and without statins for primary prevention, a group of individuals is simulated without attrition under the conditions of ELSA wave 2 from age 50 until their death or until they are 100 years old. Initial conditions and shocks are the same in both groups, the sole difference is the governing prescription process. Given these datasets, *utilities* and costs are aggregated in each scenario. The resulting figures provide the values required for Equation 3.16 in the two-years cycle. Notice that apart from *utilities*, which give the QALY estimates, we can aggregate pure life years (LY), or any other outcome as number of working hours of labour-income.

$$\mathbb{U}_t = 2 \cdot (1.060 - 0.004 \cdot \text{age}_t) \cdot \varrho_t \cdot (S_t \neq 4) \tag{3.20}$$

$$\varrho = \begin{cases} 1 & \text{if } S_t \in \{1,2\} \\ 0.77 & \text{if } S_t = 3 \end{cases}$$

$$\text{cost}_t = \begin{cases} 6980 & \text{if } S_t = 3 \,\&\, S_{t-1} \in \{1,2\} \\ 1400 & \text{if } S_t = 3 \,\&\, S_{t-1} = 3 \\ 4000 & \text{if } S_t = 4 \,\&\, S_{t-1} = 3 \end{cases} \tag{3.21}$$

One clear limitation of ELSA data is that prescriptions are patient reports, hence we do not know if those who are not reporting it were prescribed with it. Also, we cannot infer the actual compliance of those who report to be taking the drug, as discussed in section 3.3.1.3. In order to understand its effect on the estimates, I will consider an scenario with a 50% non-compliance rate, the most extreme case presented in Ward et al. (2007), which effectively means than the cost of prescription is doubled.

Ward et al. (2007) performed a systematic literature review on costs and utilities for major CVD events, with emphasis on studies which are based on UK data. Their estimates are presented in Table 3.12. However, a major difference with this study is that separate states are defined based on the primary CVD event suffered by the patients. Hence, it is necessary to collapse their figures into a single number per state in order to use the same costs and utilities. This is done by weighting

each state according to the proportion of individuals who suffered it as primary event. Table 3.13 present such results by gender and age. Given these numbers, an utility factor of 0.77, a cost of £3490 for the first year and £700 for the subsequent ones are assumed. Deaths of CVD survivors will be given a cost of £2000.

Table 3.12: Ward et al. (2007) utilities and costs per health-state

	Health States				
	Stable angina	Unestable angina	MI	TIA	Stroke
Panel A: Utilities for Health					
Utility factor	0.808	0.77	0.76	1	0.629
Panel B: Costs					
First year	175.8	452.3	4572.5	1093.8	8271.3
Subsequent years	175.8	175.8	175.8	271.4	2223.6
Fatal event		1198.6			7238.1

Notes: Based on Table 55 (2005 prices) and Table 61 of Ward et al. (2007).

Table 3.13: Ward et al. (2007) utilities and costs per age

		I. Distribution of primary events for non-fatal post-CVD events					II. CVD-state utility and costs			
Gender	Age	Stable angina	Unestable angina	MI	TIA	Stroke	Utility factor	First year	Subseq. years	Fatal event
Males	45	28.7	10	37.4	7.2	16.6	0.7696	3257.6	522.4	1321.4
	55	37.2	8	36.2	4.3	14.2	0.7696	2978.4	470.5	1123.7
	65	31.2	12	32.1	7.5	17.2	0.7716	3081.6	535.2	1388.8
	75	29	12.4	30.5	4.8	23.3	0.7562	3481.4	657.5	1835.1
	Average						*0.7716*	*3481.4*	*657.5*	*1835.1*
Females	45	34.1	11.9	26.3	4.6	23	0.7577	3269.1	651.0	1807.4
	55	41.1	8.9	21.8	8.2	20	0.7741	2853.3	593.2	1554.3
	65	33.4	12.9	25.7	4.7	23.4	0.7587	3279.1	659.6	1848.4
	75	34.3	14.6	18.7	6.9	25.4	0.7605	3157.8	702.3	2013.5
	Average						*0.7741*	*3279.1*	*702.3*	*2013.5*

Notes: Panel I is based on Table 51 of Ward et al. (2007), and Panel II are derived using utilities and costs presented in Table 3.12. Each figure from such table is multiplied by the weights provided in Panel I for each age and gender. Costs are in 2005 constant prices.

3.D Model Solution and Estimation details

3.D.1 Detailed solution

Calculating the expectation in Equation 3.14 in each period below $T+1$ involves solving several integrals as shown in Equation 3.22, a problem that is solved by

both numerical and analytic methods. This section discuss them in more detail.

$$V_h = \underset{(c,l)}{\text{argmax}}\, u(c,l) + \beta E_{e',S',v',\epsilon'_h}\left[V'(A'(c,l),H'(h,e'),s,\Upsilon'(v');X)\right] \tag{3.22}$$

$$\begin{aligned} V_h = & \underset{(c,l)}{\text{argmax}}\, u(c,l) \\ & + \beta(1-p_{4,S}(H(h)))\cdot\sum_{s=1}^{3} p_{s,S}(H(h)) \int_{e'}\int_{v'}\int_{\epsilon'_d} V'(A'(c,l),H'(h,e'),s,\Upsilon'(v');X,\epsilon'_h)\cdot f(\epsilon'_h)d\epsilon'_h \cdot f(v')dv' \cdot f(e')de' \\ & + p_{4,S}(H(h))\cdot b(A'(c,l)) \end{aligned}$$

First, as ϵ'_h follows an extreme value type I distribution, the most internal integral can be analytically expressed in terms of the value of V'_1 and V'_0.[45] Second, the value function conditional on the state values is calculated over a grid of the potential permanent income Υ' that guarantees that each point is equally feasible (Tauchen discretisation of the AR(1) process). Third, the most outer integral is solved numerically by averaging over an equi-probable grid of shocks e'. Fourth, the analytical transition probabilities are calculated conditional on the choice h. Once the expectation is solved, the resulting V_h are functions of optimal (c,l). This optimization problem is discretised over 25 points for c, on top of the 3 points of l. A final consideration is that as $V(\cdot)$ is a function of continuous state variables A, Υ and H, in practice the function is evaluated over a grid that covers the potential values of such variables, conditional on the discrete states, and then interpolated when called.

Some details about the implementation of the solution:

- The structural estimation was programmed in JULIA 0.4.
- Linear interpolation of the Emax functions
- Grid on A: 24 points, log-spaced around 0.
- Grid on H: 10 points, constant-spaced
- Grid on log Y: 8 points, Tauchen version of the AR(1)

[45] The advantages and limitations of dynamic discrete choice problems where there are taste-shocks distributed as Extreme Value Type-I is discussed in detail by Arcidiacono and Ellickson (2011).

- Optimal consumption and labour are derived after populating a grid of 25 points for consumptions times the three labour-supply options.
- Optimization algorithm (estimation): Subplex (from NLopt), derived from Nelder-Mead simplex.

3.D.2 Detailed steps for the structural estimation of the model

1. Initial steps
    - EDUC and CVD risk index were constructed
    - Data profiles are derived as described in the data section
    - Initial values for the simulations are obtained for those aged 50 to 53 in waves 2 and 4, as we require biomarkers information. See Appendix 3.D.3 for more details.
    - Potential labour income per hour is imputed for those not working using the heckman selection model presented in Appendix 3.D.3
2. Estimate Markov process independently
    - For a given set of parameters, a dataset is simulated as discussed in Section 3.D.4. As in this model choice and transition probabilities can be computed directly from the functional forms, there is no solution step prior to the simulations.
    - Matched moments. Age cells go from 53-54 to 79-80 for the case of the CVD-risk index and medication intake, and from 53-54 to 97-98 for all the other variables.
        (a) Average CVD-risk index, the proportion of individuals in each health state, and the proportion under medication.
            i. For a given age cell (ex. all individuals aged 61 to 62 years old)
            ii. For a given age cell by health state (ex. all individuals aged 61 to 62 years old, who have been diagnosed with a CVD)
            iii. For a given age cell by waves' groups: 1-2, 3-4, 5-6 (ex. all individuals aged 61 to 62 years old, who were interviewed either in wave 3 or 4)

iv. For a given age cell by education level (ex. all individuals aged 61 to 62 years old, who have no formal qualification)

(b) Average CVD-risk index for a given age cell by medication status (ex. all individuals aged 61 to 62 years old, who report to be under lipid-lowering medication)

(c) Average CVD-risk index and the proportion of individuals under medication

i. For a given age cell by size of the nearby practice (ex. all individuals aged 61 to 62 years old, who live in an area where the average practice size has been classified as small)

ii. For a given age cell by size of the nearby practice and waves' groups: 1-2, 3-4, 5-6 (ex. all individuals aged 61 to 62 years old, who live in an area where the average practice size has been classified as small, and who were interviewed either in wave 3 or 4)

(d) Variance of CVD-risk index for a given age cell

(e) Autocovariances of CVD-risk index for a given age cell, of order 2 and 4 (ex. covariance of H at age 61 with H at ages 57 and 53), provided that an individual is observed as many times as required.

(f) Elements of the state transition matrix, for all age cells (ex. proportion of those aged 55 who were in risk of CVD -$S = 2$- two years ago, and by age 57 were diagnosed with a CVD -$S - 3$-)

(g) Average CVD-risk index according to current and future health status in 4 years. For example, the average for those who currently are in. There are 9 possible combinations.

3. Obtain starting values for the income model parameters

(a) A linear fixed effect model equivalent to equation 3.12 was estimated. As the model depends on earnings, it was approximated with the observed median labour income during all waves prior to SPA.

(b) Fixed-effects and idiosyncratic residuals were predicted, and then regressed into EDUC index in order to estimate persistent differences

(c) Parameters from Equation 3.11 were pre-estimated using OLS. A first regression takes into account EDUC, to have had a CVD event ever, and a constant. From such regression's residuals, an autoregressive process was fitted in order to get an approximation to the persistent innovations process.

(d) As the income process involves selection due to participation, these parameters are also included into the main model estimation.

4. Jointly estimate the utility function parameters and income model.

- Part of the parameters are calibrated as the model and data used do not allow for identification of all them. Bequest motive penalty was set to 8, $\beta = .9604$ (two-years equivalent of the common annual rate 0.98), and interest rate $r = 0.030225$ (annual rate of $r = 0.015$).
- For a given set of parameters, the model is solved as described in Section 3.D.1. Then, specific state-space policy rules are used for simulating a dataset as discussed in Section 3.D.4
- Matched moments. Age cells go from 53-54 to 97-98 for the case of assets, and from 53-54 to 75-76 for all the other variables.

  (a) Average assets, labour (if working) and non-labour/financial income; and the proportion of individuals in each health state, working part-time, working full-time.

    i. For a given age cell (ex. all individuals aged 61 to 62 years old). Average assets at age 53 is not matched as it is considered a starting condition.

    ii. For a given age cell by health state (ex. all individuals aged 61 to 62 years old, who have been diagnosed with a CVD)

    iii. For a given age cell by waves' groups: 1-2, 3-4, 5-6 (ex. all individuals aged 61 to 62 years old, who were interviewed either in wave 3 or 4)

    iv. For a given age cell by education level (ex. all individuals aged 61 to 62 years old, who have no formal qualification)

(b) Average assets and non-labour/financial income for a given age cell according to being working or not (ex. all individuals aged 61 to 62 years old, who are working)

(c) Variance of labour-income for a given age cell

(d) Autocovariances of labour-income for a given age cell, of order 1 to 4 (ex. covariance of income at age 61 with income at ages 59, 57, 55 and 53), provided that an individual is observed working as many times as required.

5. Calculate the compensating variation

- The treatment implies a different effective discount factor for the counter-factual as the odds of survival might change. Hence, in order to calculate the compensating variation (π), Equation 3.15 has to be computed.
- Equation 3.15 is solved by considering a grid of π between -50 and 50

3.D.3 Deriving Initial Conditions

An essential element of the estimation procedure is to obtain initial conditions of the state variables in order to perform the simulations. In a nutshell, observed data for individuals ages 50 to 53 in waves 2 and 4[46], for whom there is information on all variables, is randomly replicated until the desired number of simulated individuals. While this is straightforward for most variables, it is not the case for potential labour income as it is not observed for those who do not work. Given that the non-inclusion of individuals who are not working would generate a biased sample, I imputed wages based on a cross-sectional auxiliary income model. In order to take into account selection, a traditional Heckman selection model was implemented. Explanatory variables for income involve a comprehensive set of measures with respect to education, demographics, cognitive skills and health, and whether or not the respondent's partner is sick as an excluded variable in

[46]There are 589 respondents that meet these ages, and 322 (55%) for which there information on all variables. The main sources of missing data are biomarkers information (172, 30%) and assets (135, 23%). Lack of data for the first variable is not associated to the education level, but it significantly negative related to the second. Data for wave 6 is not considered as there are few individuals within the age range and some of the variables used for the income imputation model are not available.

the participation equation. Estimated coefficients for the auxiliary model are presented in Table 3.14 for men, and in Table 3.15 for women.

Table 3.14: Initial labour income per hour:males

	OLS	**Sel. Model**	
	(1)	(2)	
	Log-Wage per hour	Log-Wage per hour	
	_	Log-Wage per hour	Is working
Age	0.006	0.006	0.006
	(0.019)	(0.019)	(0.032)
Age finished full-time education	0.055***	0.055***	0.043
	(0.013)	(0.013)	(0.027)
Educ: Some medium qualif.	0.182**	0.184***	0.144
	(0.071)	(0.071)	(0.114)
Educ: Some high level or above qualif.	0.320***	0.325***	0.319**
	(0.074)	(0.074)	(0.129)
State 2: early signs of CVD	0.008	0.005	-0.179**
	(0.045)	(0.045)	(0.089)
State 3: suffered a CVD	-0.172	-0.187*	-0.954***
	(0.112)	(0.108)	(0.150)
Reg. Unemployment rate for the month	-0.004	-0.005	-0.058**
	(0.015)	(0.015)	(0.028)
BU total net (non-pension) wealth (1000£ of May2005)	0.000**	0.000**	0.000
	(0.000)	(0.000)	(0.000)
Married	-0.007	0.002	0.684***
	(0.052)	(0.051)	(0.088)
Partner sick (not sick if not have a partner)	0.000		-0.550***
	(0.079)		(0.147)
Observations	1499	1785	
Censored Obs.		286	
ρ_ε		0.0448	
χ^2 test on H0: $\rho\_\varepsilon = 0$		0.8888	
p-val		0.3458	

Males aged 50 to 53 years old from ELSA waves 1 to 6.
Standard errors clustered at individual level in parenthesis. This table reports coefficients from an OLS and and ML Heckman selection model.
Significance: * 10%, ** 5%, *** 1%.

Table 3.15: Initial labour income per hour: females

	OLS	**Sel. Model**	
	(1)	(2)	
	Log-Wage per hour	Log-Wage per hour	
	_	Log-Wage per hour	Is working
Age	0.034**	0.034**	-0.045*
	(0.017)	(0.017)	(0.025)
Age finished full-time education	0.034**	0.034**	0.049**
	(0.014)	(0.014)	(0.025)
Educ: Some medium qualif.	0.068	0.068	0.485***
	(0.052)	(0.053)	(0.087)
Educ: Some high level or above qualif.	0.311***	0.312***	0.762***
	(0.066)	(0.068)	(0.114)
State 2: early signs of CVD	0.011	0.010	-0.332***
	(0.038)	(0.041)	(0.072)
State 3: suffered a CVD	-0.166	-0.169	-0.975***
	(0.139)	(0.152)	(0.199)
Reg. Unemployment rate for the month	-0.009	-0.009	-0.040*
	(0.012)	(0.012)	(0.022)
BU total net (non-pension) wealth (1000£ of May2005)	0.000**	0.000**	-0.000***
	(0.000)	(0.000)	(0.000)
Married	0.002	0.003	0.133*
	(0.041)	(0.040)	(0.076)
Partner sick (not sick if not have a partner)	0.014		-0.392***
	(0.061)		(0.108)
Observations	1656	2177	
Censored Obs.		521	
ρ_ε		0.0081	
χ^2 test on H0: $\rho\_\varepsilon = 0$		0.0033	
p-val		0.9542	

Males aged 50 to 53 years old from ELSA waves 1 to 6.
Standard errors clustered at individual level in parenthesis. This table reports coefficients from an OLS and and ML Heckman selection model.
Significance: * 10%, ** 5%, *** 1%.

3.D.4 Detailed simulation

The goal of the simulations is to generate a dataset that have the same structure as the ELSA survey in terms of age profiles per wave. This implies that for a given set of simulated individuals with starting values at age 53, we will need to

decide when each of them is observed for the first time and for how long are they observed. In order to do this, the model replicates selection related to age of the observed data.[47] This section explains on detail how the data generating process accounts for both elements.

Step 1: take initial values and simulate the life-profiles of 10 times the number of individuals ever observed in ELSA from waves 1 to 6 (14967) from age 53 until their death for our gender-cohort-education groups. This is 21.620/24.680/21.960 men (without formal education/at most high school/college or above); 32.340/27.810/15.270 women born before 1954, and 1.060/2.990/1.940 born after such date. For the willingness-to-pay exercises, sample size is equivalent to 10 times ELSA sample. At the end, nearly K=29.772 individuals are included in the simulated dataset (almost two times the original ELSA sample). As we will see, the dataset is simulated twice the required size as some observations will not be included due to attrition.

Given the starting values and a set of random shocks, simulations operate by applying the correspondent law of motion of each state and choice variables. Most of the states laws are motion are straightforward given their functional form. For instance, for the transition between health states, the difference in shocks drawn from an extreme value type-I distribution are contrasted with the difference between latent indexes described in Equations 3.2. For the economic-choices model the policy rules $c^*(t,S,A,H,\mathrm{Y},h,X)$ and $l^*(t,S,A,H,\mathrm{Y},h,X)$, derived from the solution, are used to infer the progression of these choices and the state A.

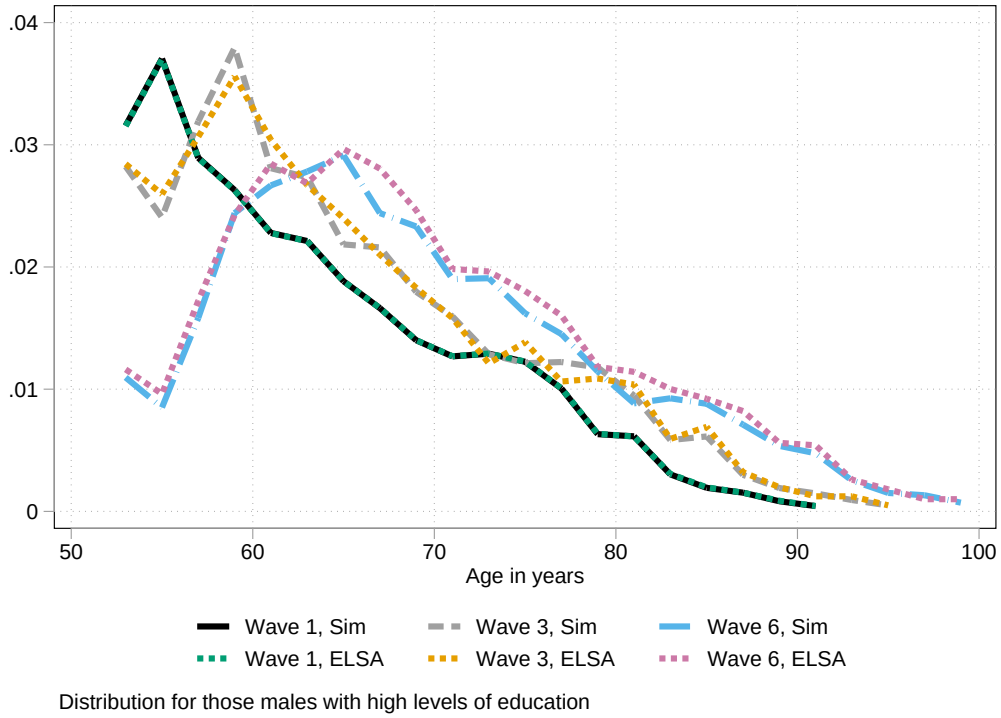
Step 2: individuals are assigned to be observed by the econometrician. This means that a given individual is observed at most for 6 consecutive periods, and that the resulting observed dataset resembles ELSA sampling structure according to age and wave. This involves deciding when an individual enters and quits the study. This is attained as shown in Figure 3.27 for males of high level of education in waves 1,2 and 3. This figure shows the proportion of individuals according to their age in each wave of ELSA and the simulated dataset. It takes into account both the initial sampling of the survey and the refreshment samples that are introduced.[48]

[47] For example, at age 73 we don't observe a random sample of the initial individuals at age 53. We do observed the healthiest and wealthiest ones, which are the one who survive and are more likely to stay for longer in the survey.

[48] Given that the data considers individuals from ages 52-53, younger ones are considered to be

Such process is done separately for the three groups of gender-cohort-education.

Figure 3.27: Age distribution of males by wave in ELSA and the simulated dataset



For step 2, the censoring of data due to mortality is modelled explicitly by the model. This is specially important with age, as oldest observed individuals are likely to be the healthiest of their cohort. The DGP replicates this selection pattern as 'weaker' individuals are less likely to be observed with age as they are more likely to die prior to be selected to be observed for the first time.

On top of mortality, there is non-random missing information related with age. This means that for some individuals, despite of being alive, we cannot observe the value of at least one of their choice or state variables, and that such event is correlated with age. The reasons for this are assumed to be independent of other choices and states. The probability of an observation being missing for individual i in wave w, $\Pr(M_i^w = 1)$, conditional on age is estimated outside the model using a logistic regression and, as shown in Equation 3.23 ($\Lambda(\cdot)$ is the inverse of the logistic function), its parameters are allowed to differ by wave and are gender-education-specific. Tables 3.16 and 3.17 presents the estimates for them.

observed for the first time only when they meet the age criteria.

Here we assume that individuals are always observed once they are death given the administrative link between ELSA and ONS mortality statistics.

$$\Lambda(\Pr(M_{i,w;g}=1)) = \psi_{1,w;g}(60 \leq \text{age}_{i,w;g} < 70) + \psi_{2,w;g}(70 \leq \text{age}_{i,w;g} < 80) + \psi_{3,w;g}(80 \leq \text{age}_{i,w;g}) \tag{3.23}$$

Table 3.16: Coefficients of missings model: Males

	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
Panel A: No Formal Education					
Age [60,70)	-0.158	-0.039	0.259	0.443**	0.309
	(0.171)	(0.175)	(0.192)	(0.203)	(0.236)
Age [70,80)	-0.343**	-0.163	0.120	0.433**	0.399*
	(0.173)	(0.174)	(0.189)	(0.201)	(0.233)
Age 80+	-0.174	-0.123	0.193	0.732***	0.697***
	(0.195)	(0.189)	(0.203)	(0.216)	(0.244)
Constant	-0.824***	-0.405***	-0.389**	-0.801***	-0.598***
	(0.138)	(0.148)	(0.166)	(0.181)	(0.217)
Observations	1534	1485	1374	1492	1437
Panel B: Up to high school					
Age [60,70)	0.001	0.115	0.244*	0.382***	0.358**
	(0.163)	(0.147)	(0.145)	(0.135)	(0.144)
Age [70,80)	-0.419**	-0.173	-0.003	0.092	0.063
	(0.188)	(0.161)	(0.155)	(0.147)	(0.155)
Age 80+	0.156	0.098	-0.048	0.152	0.163
	(0.218)	(0.198)	(0.200)	(0.186)	(0.182)
Constant	-1.334***	-0.950***	-0.763***	-0.960***	-0.844***
	(0.123)	(0.115)	(0.119)	(0.113)	(0.125)
Observations	1498	1564	1588	1927	1922
Panel C: College and above					
Age [60,70)	-0.144	-0.066	-0.027	0.031	-0.008
	(0.185)	(0.168)	(0.150)	(0.144)	(0.157)
Age [70,80)	-0.074	0.296	-0.029	-0.041	-0.105
	(0.215)	(0.189)	(0.177)	(0.167)	(0.171)
Age 80+	-0.197	-0.373	0.025	0.127	-0.329
	(0.337)	(0.297)	(0.223)	(0.215)	(0.224)
Constant	-1.662***	-1.489***	-1.070***	-1.242***	-0.962***
	(0.130)	(0.127)	(0.120)	(0.118)	(0.138)
Observations	1277	1388	1477	1820	1842

Notes: Coefficients of probability of not observing an individual conditional on being alive and being observed in at least one previous wave (see Equation 3.23). Standard errors in parenthesis. Significance: * 10%, ** 5%, *** 1%.

The following example illustrates on how the simulation algorithm works. Figure 3.28 presents a hypothetical simulated dataset. Each cell shows the number

Table 3.17: Coefficients of missings model: Females

	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
Panel A: No Formal Education					
Age [60,70)	-0.506***	0.205	0.010	0.318**	0.208
	(0.144)	(0.144)	(0.154)	(0.152)	(0.175)
Age [70,80)	-0.457***	0.011	-0.138	0.055	0.081
	(0.140)	(0.143)	(0.153)	(0.152)	(0.174)
Age 80+	-0.298**	0.228	0.109	0.595***	0.561***
	(0.149)	(0.147)	(0.156)	(0.154)	(0.175)
Constant	-0.760***	-0.756***	-0.366***	-0.717***	-0.539***
	(0.112)	(0.121)	(0.134)	(0.133)	(0.159)
Observations	2438	2422	2325	2570	2485
Panel B: Up to high school					
Age [60,70)	-0.240	0.061	-0.036	0.319***	0.190
	(0.147)	(0.131)	(0.122)	(0.118)	(0.130)
Age [70,80)	-0.521***	-0.198	-0.310**	0.001	-0.053
	(0.178)	(0.148)	(0.138)	(0.137)	(0.144)
Age 80+	-0.069	0.297*	-0.061	0.467***	0.488***
	(0.202)	(0.173)	(0.162)	(0.153)	(0.157)
Constant	-1.371***	-1.149***	-0.737***	-1.133***	-0.939***
	(0.104)	(0.098)	(0.097)	(0.099)	(0.115)
Observations	1868	2012	2136	2624	2624
Panel C: College and above					
Age [60,70)	-0.476**	0.178	0.110	-0.113	-0.092
	(0.230)	(0.202)	(0.176)	(0.163)	(0.169)
Age [70,80)	-0.385	0.295	0.080	-0.067	-0.404**
	(0.278)	(0.231)	(0.208)	(0.191)	(0.194)
Age 80+	0.194	0.497*	0.178	0.437*	0.376*
	(0.316)	(0.274)	(0.247)	(0.227)	(0.225)
Constant	-1.670***	-1.658***	-1.266***	-1.287***	-1.009***
	(0.153)	(0.157)	(0.139)	(0.129)	(0.144)
Observations	956	1035	1134	1432	1444

Notes: Coefficients of probability of not observing an individual conditional on being alive and being observed in at least one previous wave (see Equation 3.23). Standard errors in parenthesis. Significance: * 10%, ** 5%, *** 1%.

of wave assigned to each individual observation. For instance, individual 1 is observed in wave 1 when aged 53, while individual 6 is observed in the same wave but at age 55. In this case, the desired number of individuals is 10. As we will see, it is required to simulate 11 in order to fulfil the desired sample size. Moreover, let's assume that the observed dataset has the following structure:

1. 50% of the individuals are observed for the first time at age 53. Of those, 40% in wave 1, 40% in wave 2, and 20% in wave 3

2. 40% are observed at age 55 for the first time. Of those, 25% in wave 1, 50% in wave 2, and 25% in wave 3

3. The last 10% is observed for the first time at wave 3 when aged 57.

As it is required to replicate such structure, the first 5 individuals are assigned to be observed at age 53: two in wave 1, two in wave 2, and one in wave 3. Notice that we will follow them every period until wave 6. However, randomly we might be unable to observed them due to the missing information mode. For example, individual 4 was supposed to be observed ad ages 57 and 59, but it is not. Also, we fully observe them if they die, even if they have not been observed for some periods, as happens with individual 4. At age 55, we need 4 additional individuals to be observed in order to meet the data age-structure. That is one in wave 1, two in wave 2, and one in wave 3. However, notice that individual number 8 is already death, so it has to be replaced. For this reason individual 10 is considered while number 8 is completely discarded from the analysis. Finally, the sole observation left is number 11, this one is assigned to wave 3.

Figure 3.28: Example of the simulation procedure

Table: Wave if observed

					Age				Observed:
Individual	**53**	**55**	**57**	**59**	**61**	**63**	**65**	**67**	**Alive/death**
1	1	2	3	4	5	6			*6/0*
2	1	2	3	4	D	D			*4/2*
3	2	3	4	5	6				*5/0*
4	2	3	X	X	D				*2/1*
5	3	4	5	6					*4/0*
6		1	X	3	X	5	D		*4/0*
7		2	3	4	5	6			*5/0*
8		*D*	*D*	*D*	*D*	*D*	*D*	*D*	*0/0*
9		2	3	4	5	6			
10		3	4	5	6				
11			3	4	5				*6/0*
Total obs	*5*	*9*	*8*	*9*	*8*	*5*	*1*	*0*	

Observed 1st time D Death X Attrition

3.E Health Investments

Receiving a diagnosis of hypertension or the onset of a stroke might have an impact on the CVD-risk index progression and the transition probabilities. Equation 3.5 shows that there health-status has an effect on the level of the CVD-risk index. Also, there is full heterogeneity on τ according to health-status in Equations 3.2. However, the model is agnostic on why there is such heterogeneity, and more explicitly, individuals' investment in their own health is not disentangle from health care treatments. This is important as one potential mechanism to consider is that drug treatments might imply different investments in a dimension that is not already considered by the model.

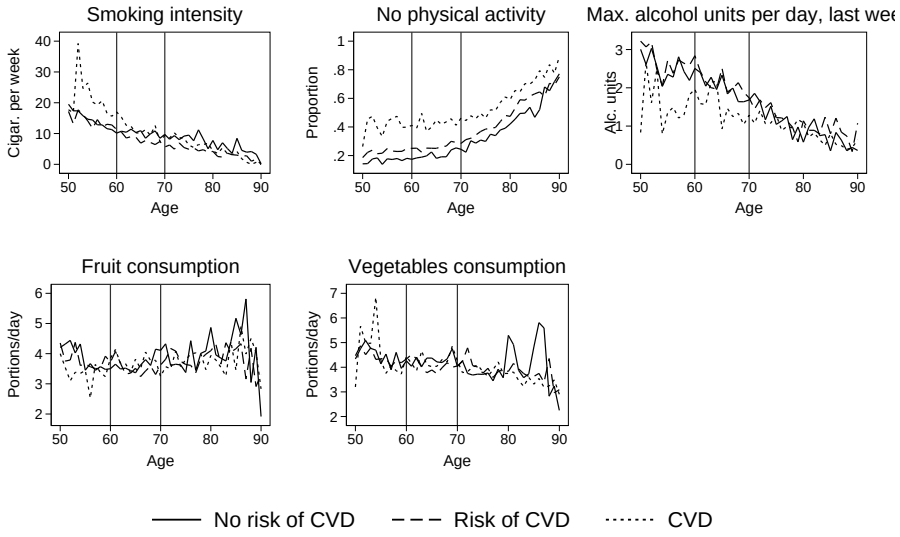
Preventive health investments come in two main groups: lifestyle and medical care. The first one involves personal choices as diet, physical activity, or smoking. Individuals normally know that their CVD-risk is affected by such habits. Figure 3.29 shows cohort-adjusted profiles, by age and health status, of some behaviours that might be relevant for CVD-risk. Both smoking and alcohol consumptions seem to decline with age, while the opposite is observed for physical activity. However, other choices as fruit and vegetables consumption are more stable on time. However, while it is true that selection is in place, it is difficult to observe a clear difference between profiles according to health state. The sole exception is physical activity, but this is a variable that might be a bad proxy for health investments as the capacity to perform certain exercises might be seriously affected by a CVD event.

The second group is related to demand of health care: check blood pressure and cholesterol, x-rays, blood tests and visits to the doctor. Figure 3.30 presents data averages per age-group and CVD status from the British Household Panel Survey (BHPS). It is clear that the increasing risk of further health complications imply more demand of preventive medical services as regular BP and cholesterol checks, blood tests, and even simple interaction with the family doctor.[49] While once again these profiles involve selection, they suggest that main mechanisms behind the different trends in CVD-risk accumulation are related to health-care

[49]The question refers about the amount of times the respondent has talked or visited a GP for the same reference period as the check-up questions. They are specifically instructed to not take into account visits to the hospital.

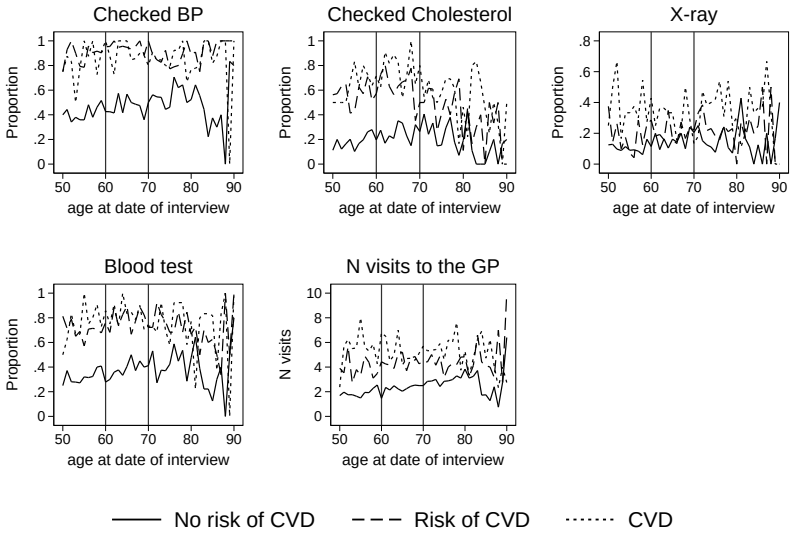
procedures rather than lifestyle modification.

Figure 3.29: Lifestyle



Smoking intensity of 0 for non-smokers

Figure 3.30: Demand for Health-Care (BHPS)



BHPS data from waves 2001/2 and 2006/7

3.F Financial Incentives

The QOF introduced a pay for performance scheme into primary care services in the UK. Since 2009/10, a new clinical area was introduced with the aim of preventing the onset of CVD for the first time (primary prevention). The first indicator promotes the use of risk assessment for patients recently diagnosed with hypertension without pre-existing diagnosis of CVD-related conditions (PP01), and the second refers to provide lifestyle recommendations to the same patients (PP02). These indicators were suggested by NICE and are based on the Clinical Guideline 67 (NICE, 2008). The first indicator, which was paid at most £1016 in that year for the average practice, is the main focus of this analysis. The precise definition of indicator is as follows:

> **PP01:** In those patients with a new diagnosis of hypertension (excluding those with pre-existing CHD, diabetes, stroke and/or TIA) recorded between the preceding 1 April to 31 March: the percentage of patients who have had a face to face cardiovascular risk assessment at the outset of diagnosis using an agreed tool.

In other words, GPs are paid for meeting with their new hypertense patients and apply a standard CVD risk calculator, initially the Framingham equation (Anderson et al., 1991), or more recently the QRISK2 model (Hippisley-Cox et al., 2008). These epidemiological models predict the odds of developing a CVD within the following 10 years, given a set of risk factors that typically involve biomakers as blood pressure and cholesterol. Their prediction is the standard for NICE guidelines on preventive care. For instance, guideline 67 recommended that patients with a 10-year risk of 20% or above should be under lipid-lowering medication. More recently, the threshold was reduced to 10% (NICE, 2014) and the indicator became explicit about the statins prescription.

QOF payment in clinical indicators is given according to a number of points obtained by a GP practice. They follow a non-linear function of indicator achievement, which is the reported usage of the services described. In this case, the proportion of new hypertense patients who received the risk assessment. Practices start to earn them if the achievement is above 40%, with a maximum number reached at 70%. The exact amount of money derived from each point depends

on the size of the practice and how common are new cases of hypertension in the practice, with respect to the national average. Hence, practices that before the intervention did not use such tools have incentives to do it. After the first year, nearly 86% of all GP practices with at least one new case of hypertension[50] attained the maximum amount of points.

As discussed in Section 3.2, there is evidence of differential performance in the QOF according to the number of patients registered in the practice, a variable known as the *list size*. While there is not conclusive evidence on the superiority of bigger practices (Ng and Ng, 2013), they have advantages as economies of scale for hiring staff and adopting information technologies in their favour. Such advantage seems to be clear given the increasing administrative pressures that family doctors are facing. This is a feature that motivates policy recommendations suggesting a move towards bigger primary care provider institutions (Goodwin et al., 2011; Smith et al., 2013). For the PP01 indicator, nearly 90% of the big practices (above 8300 patients) obtained the maximum amount of points for PP01 in 2009/10, against 83% of the small ones (less than 6000 patients). Rather than QOF achievement, what is central for this study is the odds of being prescribed with statins.

The essential question is if the introduction of affected changed the odds of being prescribed with statins between small and non-small practices. Guideline 67 suggests the usage of the risk assessment tool for people aged 40-74, so individuals aged 52 to 74 will be considered for our analysis. Equation 3.24 presents the main specification for testing this hypothesis. Here the variation in lipid-lowering $Y_{ir,w} - Y_{ir,w-1}$ medication from an individual i between 2010 and 2012 is compared with such increase between 2006 and 2008, between areas r with small practices ($SL_r = 1$) relative to areas with non-small practices ($SL_r = 0$). As a condition, the individual cannot be previously diagnosed with a CVD by the first year of the respective variation. This is conditional on second order polynomials on age and biomarkers measured in 2008 (for 2010 to 2012) and in 2004 (for 2006 to 2008), gender, smoking status, PCT covariates and government region dummies ($X_{i,w-2}$). BMI, systolic blood pressure, total and HDL cholesterol are considered two waves

[50] 92 practices did not have one case.

before in order to avoid contamination from current medication status.

$$Y_{ir,w} - Y_{ir,w-1} = \alpha_1 SL_r + \alpha_2 \mathbb{1}\{w = 6\} + \alpha_3 SL_r \cdot \mathbb{1}\{w = 6\} + \alpha X_{i,w-2} + u_{irt}, \quad w \in [4, 6] \tag{3.24}$$

Column 1 of Table 3.18 presents the results for this analysis. It shows that in , those living in *small* GP practices areas it was around 2 pp. less likely to be under lipid-lowering medication conditional on age and objective health measures. However, this difference is removed by 2012.

An additional concern is that QOF policy is intended to include lifestyle advice. As we have seen, doctors are also paid for providing such advice. Therefore, the benefits of the program might be understated. More worrying is that individuals might increase their unhealthy behaviour due to the medication, or at least fail to reduce it as much as those who are not under the drug. This will imply an overestimation of the effects of the policy. Table 3.18 also shows that there is no evidence of different lifestyle changes as smoking, physical activity of fruit and vegetables consumption. Also, there is no evidence that other essential medication for CVD-risk reduction, hypertension treatments, are modified.

Table 3.18: QOF impact on primary care prescription of health investments according to list size

	(1) LIPID PILL	(2) BP PILL	(3) SMOKE	(4) LOW PA	(5) FRUIT VEGT
(=1) Living in a Small Practice (SP)	-0.025*	-0.003	-0.002	-0.001	0.241
	(0.014)	(0.010)	(0.010)	(0.024)	(0.537)
(=1) Interviewed in 2012/13 $(\mathbb{1}\{w=6\})$	-0.023*	-0.013	-0.004	-0.038	0.043
	(0.012)	(0.009)	(0.007)	(0.024)	(0.393)
(=1) Interaction $(SP)\cdot(\mathbb{1}\{w=6\})$	0.035**	0.002	0.016	0.009	-0.036
	(0.017)	(0.012)	(0.010)	(0.032)	(0.549)
Observations	5338	5339	5337	5275	4620
Clusters	150	150	150	150	150

Notes: This table presents coefficients associated to be in a *small* GP practice area, to be surveyed in 2012/13 instead of 2008/09, and their interaction , for five different dependent variables (columns) The sample consist on ELSA core individuals aged 52 to 74 in 2012 or 2008, for which there is information on their lipid-lowering medication status in the current and previous wave. All models include a second order polynomial on age and 4 year lagged biomarkers (BMI, systolic blood pressure, total and HDL cholesterol), blood pressure medication, previous signs of CVD status (State 2), and smoking status. Governement office region and education level fixed effects, and PCT level controls are also included. LIPID PILL: to be under lipid-lowering medication. BP PILL: to be under any type of blood pressure medication; lagged version of this variables is not included as a control. SMOKE: to report to be an active tabacco smoker. LOW PA: to be sedentary or to have a low physical activity level. This is, either not working or at sedentary occupation, and at most engages in mild exercise less than three times a month or less. FRUIT VEG: portions o fruit and vegetables per week. Standard errors clustered at individual level in parenthesis. Significance: * 10%, ** 5%, *** 1%.

Chapter 4

Identifying complementarities across tasks using two-part contracts. An application to family doctors

Joint work with Marcos Vera-Hernández

4.1 Introduction

Principal-agent relationships are widespread in economics. Since Holmstrom and Milgrom (1991)'s seminar article, it is well understood that the agent's cost function plays a crucial role in a multitask environment, that is, when the agent must carry out more than one task. If tasks complement each other, rewarding one task will be enough to increase the production of an unrewarded task. If, however, tasks substitute for each other, rewarding one will reduce the effort exerted on the unrewarded task. Complementarities/substitutions across tasks not only play a role in the structure of incentive contracts, but also in job design. Whenever possible, tasks that are substitutes should be performed by different agents, each of them carrying out tasks that complement each other.

In this paper, we show how to recover from the data whether tasks are complements or substitutes when the agent faces a two-part linear contract, essentially a contract with two different piece-rate levels. Our approach exploits a change in the incentives faced by the agents, but in contrast to the literature, we can exploit nationwide incentive changes, and do not need that a "control" group, that is, a group of agents not eligible for the change of incentives.

Our main insight is that when agents face two-part linear contracts, a group

of agents will naturally choose to produce at the level of the kink of the two-part contract, the level at which the piece rate changes. We show that these agents are insensitive to local changes in the incentives of other tasks, independently of how the tasks interact in the cost function. Hence, the individuals at the kink will work as a "control" group. Because linear two-part contracts are quite prevalent (and we do not need an explicit control group), our method greatly expands the situations in which we can test for complementarities/substitutions in the agent's cost function.

We apply our method to identify whether different activities that family doctors perform are complements or substitutes in their cost function. Examples of the activities that we analyse include carrying out certain tests on diabetic patients, recording smoking history in *at risk* patients, or reviewing asthmatic patients with some minimum frequency, among others.

The types of activities that we analyse contrast with much of the existing empirical literature that has focused on much simpler activities.[1] This literature has focused on studying a specific case of multitasking: the trade-off between quantity and quality within a *single* activity. Because of the very nature of it, quantity and quality are either substitutes or independent at best, but complementarity is rightly dismissed.[2] Because we study genuinely different tasks (rather than the quantity and quality of a single task), the possibility of complementarities across them is real. It might well be, for instance, that the marginal cost of carrying out a test is smaller if another test is also being conducted during the same visit.

Monetary incentives are also used amongst professions with a large pro-social component, such as teachers and doctors. Although crowding out of intrinsic motivation is usually cited as a concern, multitasking is another one. Unsurprisingly, there is a reasonably large body of literature for "teaching to the test", and more generally whether teachers shift effort from unrewarded tasks to rewarded ones (see Neal (2011) for a review of US focused studies).[3] The evidence on health care

[1] See for instance, Lazear (2000), Shearer (2004), Kosfeld and Neckermann (2011), Bradler et al. (2013).

[2] Al-Ubaydli et al. (2012) finds that higher piece rates leads to higher quality when stuffing envelopes, but this is explained because the piece-rate mechanism signals to the agent that the principal has a good monitoring technology rather because there are complementarities in the cost function.

[3] Muralidharan and Sundararaman (2011), and Glewwe et al. (2010) are examples of developing country studies).

probably lags behind that on education. Dumont et al. (2008) found that Canadian physicians who voluntary signed up to a contract that paid less for a specific quantity of consultations, increased the average time per consultation (an indicator of quality) as well as other activities unremunerated at the margin (i.e. teaching). Feng Lu (2012) exploited a mandatory quality disclosure policy and found that nursing homes improved scores on quality measures for the reported dimensions, but deteriorated in regard to unreported ones.

In this paper, we exploit the *Quality Outcomes Framework* (QOF), a programme established in 2004 that remunerated all family doctors in England according to their performance in a large battery of indicators. There is a remuneration schedule for each rewarded indicator, which has a lower and an upper limit. The doctor's remuneration increases linearly as long as the indicator is between the lower and upper limit, and flattens out if the upper limit is passed. The programme was rolled out simultaneously across England, and any changes to the remuneration schedule also apply nationally. This makes it an ideal setting to apply the method that we develop in this paper.

The QOF is the largest primary care pay for performance programme worldwide, and has already received some attention. Sutton et al. (2010)compared incentivised and unincentivised measures before and after the introduction of the program, a improvements in both measures which were higher for incentivised ones.[4] This approach relies on the assumption that incentivised and unincentivised measures would follow a common trend in the absence of the program. We are able to overcome these limitation thanks to the two-part linear payment scheme of QOF where there is an upper limit for the increasing payments according to performance. Also, as there is a period in which there are changes to rewards (2010/11), preceded by a period without them (2009/10), we are able to distinguish effort response to the new rewards from variations linked to year-to-year variation, which are correlated with performance. We found that there is no evidence of substitutability between tasks in the system, and if anything, several of them are complements.

[4]Kaarboe and Siciliani (2011) motivate their multitasking model using the QOF. They argue, based on the results of Sutton et al. (2010), that quality dimensions in primary care might be complements.

After this introduction, we present a basic model of multitasking with a two-part linear reward function for agents. Given that, we show the conditions under which we are able to identify complementarities/substitution in the cost function empirically. This is followed by a description of the QOF and the results of using our test on it. Finally, the conclusions are presented.

4.2 Model

Our test is based on the existence of a two-part linear tariff on a principal-agent relationship. In order to understand the intuition behind the test, we will start by presenting a simple version of the model without uncertainty. In this model, we will introduce the kink produced by a two-linear tariff and examine its implications. Later, we will consider how this main ideas would be affected by introducing uncertainty.

Consider a principal-agent relationship with two distinct tasks. The principal hires the agent to to exert task-specific efforts (e_1,e_2). The principal benefits increasingly from the output of the two tasks (x_1,x_2). The agent is paid according to $P(x_1,x_2;a_1,a_2) = T + a_1x_1 + a_2x_2$, where T represents a lump-sum payment, and a_i is the piece rate associated to x_i. The agent's cost function is given by $C(e_1,e_2;z)$. characterised by a parameter z. We assume that for $i \in 1,2$ we have that $\frac{\partial C}{\partial e_i} = C_i > 0$, $\frac{\partial^2 C}{\partial e_{ii}^2} = C_{ii} > 0$, $\frac{\partial C}{\partial z} > 0$, $\frac{\partial^2 C}{\partial z^2} > 0$, and that C is a convex function, but we do not restrict the sign of the cross-derivatives $C_{ij} = \frac{\partial^2 C}{\partial e_j \partial e_i}$, $i \neq j$. That is, while we know that it is increasingly costly to exert effort, we do not know if increasing effort in one task, increases or reduces the marginal cost of exerting effort on the other task. In the former case, the tasks are said to be substitutes, and in the latter they are complements. Our main goal is to estimate the sign C_{ij} to ascertain whether the tasks are complements or substitutes.

The agent takes the contract $P(x_1,x_2)$ as given, and decides optimal levels of effort in order to maximize his surplus, that is:

$$\max_{e_1,e_2} U = E\left[P(x_1,x_2;a_1,a_2) - C(e_1,e_2;z)\right] \tag{4.1}$$

4.2.1 Model without uncertainty

Let's assume that $x_{i=}e_i$ and that providers are heterogeneous only on an efficiency parameter z which is assigned in the population following a pdf $g(\cdot)$, or CDF $G(\cdot)$. Specifically, let's assume that $C(e_1,e_2;z) = \frac{1}{z}C(e_1,e_2)$. As a result, given a contract specified by $\{T,a_1,a_2\}$, the provider will solve:

$$\max_{e_1,e_2} U = (T + a_1 \cdot e_1 + a_2 \cdot e_2) - \frac{1}{z}C(e_1,e_2). \tag{4.2}$$

The first order conditions (FOC) of the problem are given by:[5]

$$a_i - \frac{1}{z}C_i = 0,\ i \in \{1,2\} \tag{4.3}$$

Essentially, the marginal benefit (a_i) of exerting effort has to be equal to the marginal cost ($\frac{1}{z}C_i$). If we differentiate these FOC, we obtain:

$$da_i - \frac{1}{z}C_{ii}de_i - \frac{1}{z}C_{ij}de_j = 0,\ i \neq j,\ i,j \in \{1,2\} \tag{4.4}$$

This system of equations allows us to explore how optimal allocation of effort in each task would be adjusted as a response to variations in the piece-rates a_i and to the efficiency parameter z.

Proposition 1. *With a linear payment and without uncertainty, we have that* $\frac{de_1}{da_2} = \frac{-z \cdot C_{12}}{C_{11}C_{22}-C_{12}^2} > 0$, *and hence that the sign of* $\frac{de_1}{da_2}$ *is opposite to the sign of* C_{12}. *If the tasks are substitutes* $(C_{12} > 0)$, *we will have that* $\frac{de_1}{da_2} < 0$. *On the contrary, if the tasks are complements* $(C_{12} < 0)$ *then* $\frac{de_1}{da_2} > 0$.

Proof:

If we set $da_1 = 0$, that is, a_1 as the unchanged P4P incentive, we can obtain that

$$de_2 = -\frac{C_{11}}{C_{12}}de_1 \tag{4.5}$$

And hence, the impact of modifying the reward a_2 on e_2 is obtained by substituting (4.5) in the FOC of e_2 :

[5]The second order condition (SOC) is given by $C_{11}C_{22} - C_{12}^2 > 0$, which we assume to hold.

$$\frac{de_1}{da_2} = -\frac{z \cdot C_{12}}{C_{11}C_{22} - C_{12}^2} \tag{4.6}$$

Q.E.D.

If we consider that $da_2 = 0$ but $da_1 \neq 0$, we can derive the response on optimal effort for task 1, given variations in its own price. As expected, it is unambiguously positive:

$$\frac{de_1}{da_1} = \frac{z \cdot C_{22}}{C_{11}C_{22} - C_{12}^2} > 0 \tag{4.7}$$

Assumption 1. *We assume that* $\frac{de_1}{dz} = \frac{a_1 C_{22} - a_2 C_{12}}{C_{11}C_{22} - C_{12}^2} > 0$, *for any value of z.*

Note that this is a very natural assumption: if the agent becomes more efficient and its costs decreases, he will exert more effort. It is indeed guaranteed for the case of complements, because $C_{12} < 0$. For the case of substitutes, we need to assume that C_{22} is not too small compared to C_{12}. Otherwise, the agent might greatly increase e_2 and decrease e_1.

4.2.2 The role of kinks

Now, let's consider a two-part linear payment function, with a kink at $e_1 = UL$.[6] We consider a piece-rate for a given task varies at UL from $\underline{a}_1$to $\bar{a}_1$, as shown in Equation 4.8 below. As a notation convention, all objects denoted with a lower bar will be related to the contract when the output is below UL, and those with an upper bar for the contract when the output is above such a value. Following our specific application,[7] we will consider $\underline{a}_1 > \bar{a}_1$, so the marginal benefit of e_1 decreases discontinuously at $e_1 = UL$. Notice that this payment function also implies that the fix income jumps in order to maintain the total payment continuous

[6] As will be described in the application section, the QoF is a three-part linear contract. It has a zero piece-rate below a first threshold, the *lower limit*, and above a second threshold, the *upper limit*. We will concentrate on what happens around the *upper limit* given that most of the agents are situated around or above it. Nevertheless, the model and empirical test detailed in this paper could potentially be formulated to the lower limit if there was enough information.

[7] The QoF presents an extreme scenario: $\underline{a}_1 > \bar{a}_1 = 0$. The results that we present here do not require a zero marginal benefit for unit of effort after the upper threshold. An alternative interpretation is that $\bar{a}_1$ represents the altruistic marginal benefit that the physicians obtain for improving their patients' health.

at UL.

$$P(x_1,x_2;a_1,a_2) = \begin{cases} \underline{a}_1 x_1 + a_2 x_2 + T & if\, x_1 < UL \\ \bar{a}_1 x_1 + a_2 x_2 + T + (\underline{a}_1 - \bar{a}_1)\cdot UL & if\, x_1 \geq UL \end{cases} \tag{4.8}$$

Proposition 2. *Without uncertainty, the presence of a kink at $e_1 = UL$ implies that those providers with a $z \in [\underline{z},\bar{z}]$ choose $e_1^* = UL$. Moreover, $\frac{de_1}{da_2} = 0$ for them.*

Proof:

Below the threshold UL, for a given z there is a optimal level of effort $\underline{e}_1(z) = e_1^*(z,\underline{a}_1,a_2)$. In particular, we assume that $\exists z = \underline{z}\ st\ \underline{e}_1(\underline{z}) = UL$. Above the threshold, $e_1 > UL$, there is also an optimal allocation $\bar{e}_1(\bar{z}) = e_1^*(z,\bar{a}_1,a_2)$, and we also assume that $\exists \underline{z}\ st\ \bar{e}_1(\bar{z}) = UL$.

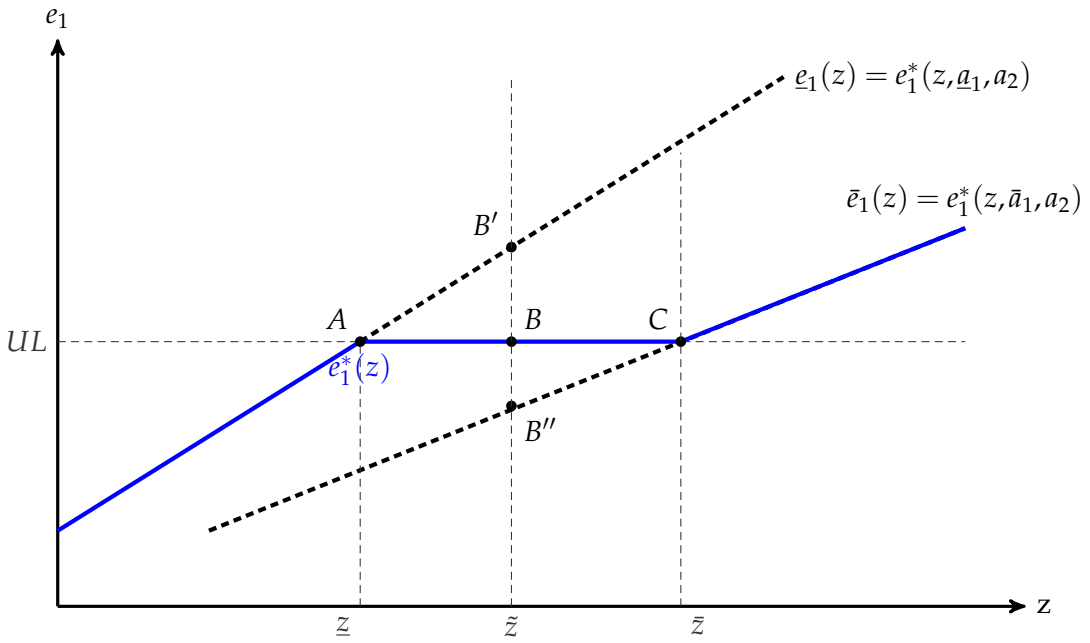
Given that $\underline{a}_1 > \bar{a}_1$, the optimal effort above UL, $\bar{e}_1(z) = e_1^*(z,\bar{a}_1,a_2)$, has to be smaller than the corresponding decision if there were no kink: $\underline{e}_1(z) > \bar{e}_1(z)\ \forall z$. In particular, $UL = \underline{e}_1(\underline{z}) > \bar{e}_1(\underline{z})$. This is due to Equation 4.7. Notice that it has to be the case that $e_1^*(z+\epsilon,a_1,a_2) > e_1^*(z,a_1,a_2)\ \forall \epsilon > 0$, which holds because of Assumption 1 ($a_1C_{22} - a_2C_{12} > 0$).[8] As a result, given that $UL > \bar{e}_1(\underline{z})$, it is required that $\bar{z} > \underline{z}$.

Those providers with a $z \in [\underline{z},\bar{z}]$ have to choose $e_1^* = UL$, even though the FOC is not satisfied, because any deviation would be detriment of their utility. Let us consider the diagram on Figure 4.1 to illustrate the argument. Point A represents the decision of a provider with productivity $\underline{z}$, which is $e_1^* = UL$ as stated before. Point C does the same for the typical $\bar{z}$ provider, which also chooses $e_1^* = UL$.

Let us consider a provider with a productivity in between, $\tilde{z} \in (\underline{z},\bar{z})$. Without the kink, the optimal decision under $\underline{e}_1(z)$ would have been point B'; however under the kinked payment function it is not optimal. At this point the marginal cost of exerting effort is larger than the marginal benefit of doing so, $\frac{C_1}{C_2}a_2 > a_1$ (from the FOC), so it is a better idea to reduce effort in order to enhance utility. An alternative scenario is to consider a world where $a_1 = \bar{a}_1\,\forall e_1$; in such a scenario B'' would have been the choice. Once again, under the actual kinked function this is suboptimal. The provider is better off if effort is increased, as at that point

[8] If the assumption does not hold, $e_1^* = 0$ as discussed before. A milder version would be when $\underline{a}_1 \cdot C_{22} - a_2C_{12} > 0$ but $\bar{a}_1 \cdot C_{22} - a_2C_{12} < 0$. In such a case $e_1^* = UL$ will always be preferred for all $z \geq \underline{z}$. This implies that there should not be no provider above UL, regardless of the value of z.

Figure 4.1: The effect of a kink on rewards at $e_1 = UL$



Note: Providers' payment for task 1 effort below UL is $\underline{a}_1$, and above it is $\bar{a}_1 < \underline{a}_1$. It produces a piecewise optimal effort function $e_1^*(z) = \underline{e}_1(z) \times \mathbb{1}(e_1 \leq UL) + \bar{e}_1(z) \times \mathbb{1}(e_1 > UL)$, where $\underline{e}_1(z) = e_1^*(z,\underline{a}_1,a_2)$ and $\bar{e}_1(z) = e_1^*(z,\bar{a}_1,a_2)$. This diagram assumes constant second derivatives of function $C(e_1,e_2)$. It is also assumed that both tasks are substitutes, so the slope above UL is smaller than below it (see Assumption 1). Nevertheless, in the diagram $\bar{a}_1 C_{22} - a_2 C_{12} > 0$, hence the values of e_1^* above UL are feasible.

$\frac{C_1}{C_2} a_2 < a_1$. As a result, due to the non-smoothness of the optimization problem, the provider is better off at point B, even though the FOCs do not hold. Notice that as $\frac{C_1}{C_2} a_2 \neq a_1$, the effect of a small variation in a_2 would have no impact on the allocation of a_1^*. As a result, $\frac{de_1}{da_2} = 0$ for those providers with a $z \in [\underline{z},\bar{z}]$.

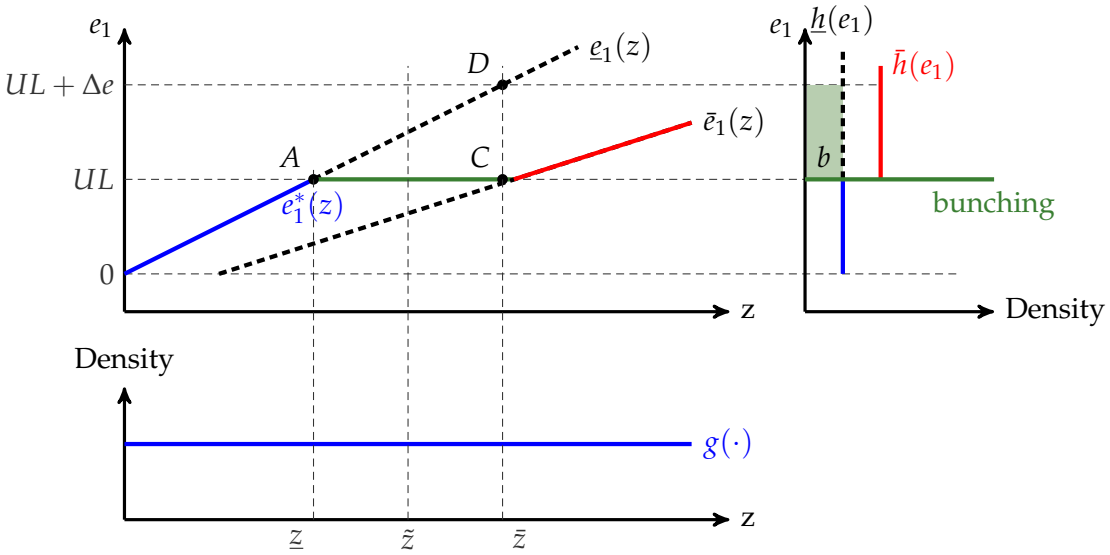
Proposition 3. *Without uncertainty, the presence of a kink at $e_1 = UL$ generates bunching on the distribution of effort on task one, $H(e_1)$.*

Proof:

This we can follow Saez (2010).[9] We define $H(\tilde{e}_1) = \Pr[e_1^*(z,a_1,a_2) \leq \tilde{e}_1] = \Pr\left[z \leq e_1^{*-1}(\tilde{e}_1;a_1,a_2)\right] = G\left[e_1^{*-1}(\tilde{e}_1;a_1,a_2)\right]$, where $e_1^{*-1}(\cdot)$ is the inverse function of $e_1^*(z)$. As explained above, $e_1^*(z)$ is piecewise defined, which is also the case for $H(\tilde{e}_1)$. Below UL we have $\underline{H}(\tilde{e}_1) = G\left[e_1^{*-1}(\tilde{e}_1;\underline{a}_1,a_2)\right]$, and above it the relevant function is $\bar{H}(\tilde{e}_1) = G\left[e_1^{*-1}(\tilde{e}_1;\bar{a}_1,a_2)\right]$. Given that all providers with a $\tilde{z} \in [\underline{z},\bar{z}]$ have to choose $e_1^* = UL$, an entire mass that would have exerted an effort $\underline{e}_1(\tilde{z}) >$

[9] See Kleven (2016) for a good review.

Figure 4.2: The effect on e_1 density of a kink on rewards at $e_1 = UL$



Note: Providers' payment for task 1 effort below UL is $\underline{a}_1$, and above it is $\bar{a}_1 < \underline{a}_1$. It produces a piecewise optimal effort function $e_1^*(z) = \underline{e}_1(z) \times \mathbb{1}(e_1 \leq UL) + \bar{e}_1(z) \times \mathbb{1}(e_1 > UL)$, where $\underline{e}_1(z) = e_1^*(z, \underline{a}_1, a_2)$ and $\bar{e}_1(z) = e_1^*(z, \bar{a}_1, a_2)$. This diagram assumes constant second derivatives of function $C(e_1, e_2)$. It is also assumed that both tasks are substitutes, so the slope above UL is smaller than below it (see Assumption 1). Nevertheless, in the diagram $\tilde{a}_1 C_{22} - a_2 C_{12} > 0$, hence the values of e_1^* above UL are feasible.

UL if there were no kink is now collapsed at that single point and has a value of $b = h(UL) = \underline{H}(\underline{e}_1(\bar{z})) - \underline{H}(UL)$. Above $\bar{z}$, the distribution will follow $\bar{h}(e_1)$

Figure 4.2 extends the previous example and considers a uniform density $g(z)$ and how it transforms into $h(e_1)$. For $z < \underline{z}$, the kink makes no difference at all: $h(e_1) = \underline{h}(e_1)$. However, for those $z \in [\underline{z}, \bar{z}]$ there is a clear change. Without the kink, such provider would have exerted $e_1 \in [UL, UL + \Delta e]$, between points A and D in the figure, which would have followed the density $\underline{h}(e_1)$. Because of the kink, AD became AC and the entire area b is now collapsed into a unique spike at $e_1 = UL$. Finally, for $z > \bar{z}$ we have that optimal effort is given by $\bar{e}_1(z)$, which is reflected by density $\bar{h}(e_1)$. Notice that it is required that $1 - \bar{H}(UL) = 1 - \underline{H}(UL)$, so the final $H(e_1)$ is a valid CDF. This is reflected in the fact that all observations that would have covered $e_1 \in [UL + \Delta, \infty)$, are now spread into $e_1 \in [UL, \infty)$.[10]

[10]For the uniform example in Figure 4.2, this means that the maximum value of e_1 will fall, but the density at any point will be larger ($\bar{h}(e_1) > \underline{h}(e_1)$ for $e_1 \in [UL, \bar{e}_1(z^{max})]$). See the example in the Appendix for more details.

4.2.3 Uncertainty

A common characteristic of multitasking models is the role of uncertainty.[11] In particular, Holmstrom and Milgrom (1991) discuss the role of using noisy signals for rewarding agents. Let us consider $x_1 = e_1 + \varepsilon_1$, where ε_1 is distributed according to $F(\cdot)$, which is a twice differentiable CDF, with PDF $f(\cdot)$. Let us assume that uncertainty has an impact of $-\Omega < 0$ on utility.[12] Hence, we can write their problem as follows.

$$\begin{aligned}\max_{e_1,e_2\in[0,1]} U =& \Pr[e_1 < UL - \varepsilon_1] \cdot \left\{ E\left[(T + \underline{a}_1 \cdot e_1 + \underline{a}_1 \cdot \varepsilon_1 + a_2 e_2) - \frac{1}{z} C(e_1, e_2) \right] - \Omega \right\} \\ &+ \Pr[e_1 \geq UL - \varepsilon_1] \cdot \left\{ (T + p_1 UL + \bar{a}_1 e_1 + a_2 e_2) - \frac{1}{z} C(e_1, e_2) \right\}\end{aligned}$$

The FOC for e_2 is still the same as before, but for e_1 it is different. First, part of the marginal financial return $p_1 = \underline{a}_1 - \bar{a}_1$ is now subject to uncertainty; so as long as $e_1 < UL$ it will be obtained. However, by exerting more effort, the probability of loosing such a financial reward decreases, but also there is a reduction in the uncertainty penalty Ω. This is reflected in the term in brackets of Equation 4.11.

$$FOC_1 := \bar{a}_1 - \frac{1}{z} C_1 + \{ F(UL - e_1) \cdot p_1 - f(UL - e_1) \cdot [(e_1 - UL) \cdot p_1 - \Omega] \} = 0 \tag{4.11}$$

We can obtain the marginal variation in optimal effort on task 1 with respect

[11] For our particular application, the model without uncertainty is not necessarily too simplistic. This is because the payment is based on the aggregate outcome of the doctor's patients, and hence the noise might be averaged out.

[12] This would be the case with preferences that exhibit absolute risk aversion η. For example:

$$\begin{aligned}\max_{e_1,e_2\in[0,1]} U &= E\left[u(\alpha B(e_1,e_2) + (T + \phi(p_1, e_1 + \varepsilon_1) + a_2 e_2 - \frac{1}{z} C(e_1,e_2)) \right] \\ &= E[-e^{-\eta(\alpha B(e_1,e_2) + (T + \phi(p_1, e_1+\varepsilon_1) + \bar{a}_1 e_1 + a_2 e_2) - \frac{1}{z} C(e_1,e_2))}]\end{aligned} \tag{4.9}$$

With a linear tariff $\phi_1(x_1) = p_1 x_1 = p_1 e_1 + p_1 \varepsilon_1$ the problem can be expressed in terms of the certainty equivalent $\hat{U}$. Where, despite risk aversion, the noise plays no role in the allocation of effort. This is because a provider's choices do not affect the expected value of the reward for attaining a certain level of performance.

$$\max_{e_1,e_2\in[0,1]} \hat{U} = \alpha B(e_1,e_2) + (T + a_1 e_1 + a_2 e_2) - C(e_1,e_2) - \frac{1}{2}\eta(p_1^2 \sigma_1^2) \tag{4.10}$$

to the reward on task 2 following the same procedure as in the case without uncertainty.

$$\frac{de_1}{da_2} = -\frac{z \cdot C_{12}}{C_{11}C_{22} - C_{12}^2 + p_1 \cdot z \cdot C_{22} \cdot f(UL - e_1) \cdot \left\{2 + \frac{f'(UL-e_1)}{f(UL-e_1)} \cdot \left[\frac{1}{p_1}\Omega + UL - e_1\right]\right\}} \tag{4.12}$$

This expression is equivalent to Equation 4.6, but with D instead of C_{11}. As before the sign is determined by C_{12}, but the magnitude is a function of current effort with respect to e_1. Hence, Proposition 1 is not affected by the presence of either risk or uncertainty. However, Proposition 2 requires further analysis.

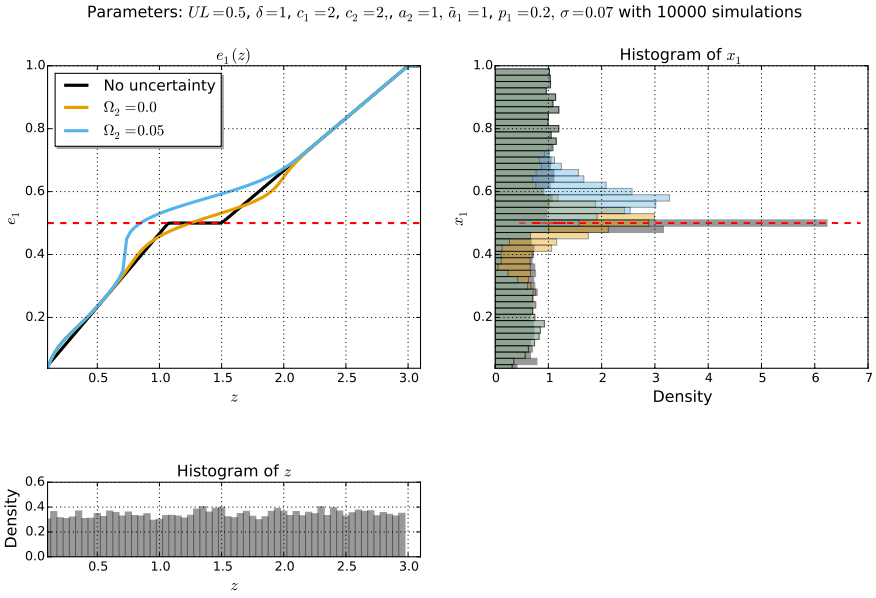
As with the no-uncertainty scenario, we can derive how general efficiency z is related to e_1^*.

$$\frac{de_1}{dz} = \frac{\left\{\bar{a}_1 + p_1 \cdot \left[F(UL - e_1) + f(UL - e_1) \cdot \left(\frac{1}{p_1}\Omega + UL - e_1\right)\right]\right\} C_{22} - a_2 C_{12}}{C_{11}C_{22} - C_{12}^2 + p_1 \cdot z \cdot C_{22} \cdot f(UL - e_1) \cdot \left\{2 + \frac{f'(UL-e_1)}{f(UL-e_1)} \cdot \left[\frac{1}{p_1}\Omega + UL - e_1\right]\right\}} \tag{4.13}$$

The simulation exercise in Figure 4.3 will be useful to illustrate how Equations 4.12 and 4.13 compare with the ones in the no-uncertainty case. This Figure follows the same configuration as the diagram presented in Figure 4.2. In this simulation, a cost function with constant second order derivatives is assumed. the noise on the task's result is assumed to follow a normal distribution. The provided parameters imply that both tasks are substitutes, and parameter z is drawn from a uniform distribution. The figure considers three cases: first, in black, the policy rules for e_1^* derived with no-uncertainty (black); second, with uncertainty but without risk aversion (orange), and finally including risk aversion (light blue).

Let us consider the case without risk aversion, $\Omega = 0$. As shown in the graph, uncertainty essentially smooths out the corners of optimal effort $e_1^*(z)$. Moreover, the slope $\frac{\partial e_1}{\partial z}$ is always positive, as predicted by Equation 4.13. While introducing noise removes the idea of corner solution, it still generates bunching at UL as the slope becomes smaller rapidly near this threshold.

Figure 4.3: Simulation exercise



Note: Parameters z drawn from a beta distribution with parameters $(5,2)$ multiplied by 3. The cost function is defined as $C(e_1,e_2;z,c_1,c_2,\delta) = \frac{1}{z} \cdot (\frac{1}{2}(c_1e_1^2 + c_2e_2^2) + \delta e_1e_2)$. For the cases with uncertainty, $x_1 = e_1 + v_1$ where $v_1 \sim N(0,\sigma)$

Let us consider first the denominator of Equation 4.13, and in particular, its last term: $\left(\frac{f'(UL-e_1)}{f(UL-e_1)} \cdot [UL - e_1] < 0\right)$. When $e_1^* < UL$, it is implied that $(UL - e_1) < 0$ which also means that $f'(UL - e_1) > 0$. Hence, the entire term is negative $\left(\frac{f'(UL-e_1)}{f(UL-e_1)} \cdot [UL - e_1] < 0\right)$, so the denominator will become smaller as e_1 moves away from UL. When $e_1^* > UL$, exactly the same happens as when $f'(\cdot) < 0$ and $(UL - e_1) > 0$. Hence, the further e_1 is from UL, the larger the derivative, at least until it becomes equal to the no-uncertainty case when $f(UL - e_1) \to 0$.

Risk aversion plays an important role as observed in the example in Figure 4.3 ($\Omega = 0.05$). In the denominator, the term $\left(\frac{f'(UL-e_1)}{f(UL-e_1)} \cdot \left[\frac{1}{p_1}\Omega + UL - e_1\right]\right)$ changes the sign near UL three times. First, below UL, it makes the slope even larger, as it goes in the same direction as $UL - e_1$and the denominator becomes smaller. Second, in the interval $e_1^* \in [UL, \frac{1}{p_1}\Omega + UL]$, the term $f'(\cdot)$ becomes positive so the denominator is larger and then the derivative $\frac{de_1}{da_2}$ is smaller. Finally, when $e_1^* \geq$

$\frac{1}{p_1}\Omega + UL$, the derivative starts to grow again. The implication for the distribution of x_1 is that the bunching will be centred above UL.

While the numerator of Equation 4.13 is also a function of $f(\cdot)$ and risk aversion, it plays a less important role in the graph of $e_1^*(z)$. The term $\left[F(UL - e_1) + f(UL - e_1) \cdot \left(\frac{1}{p_1}\Omega + UL - e_1\right)\right]$ decreases as e_1 departs from 0. This is because $F(UL - e_1)$ decreases with e_1, and so does $\left(\frac{1}{p_1}\Omega + UL - e_1\right)$. This effect is present both above and below UL.

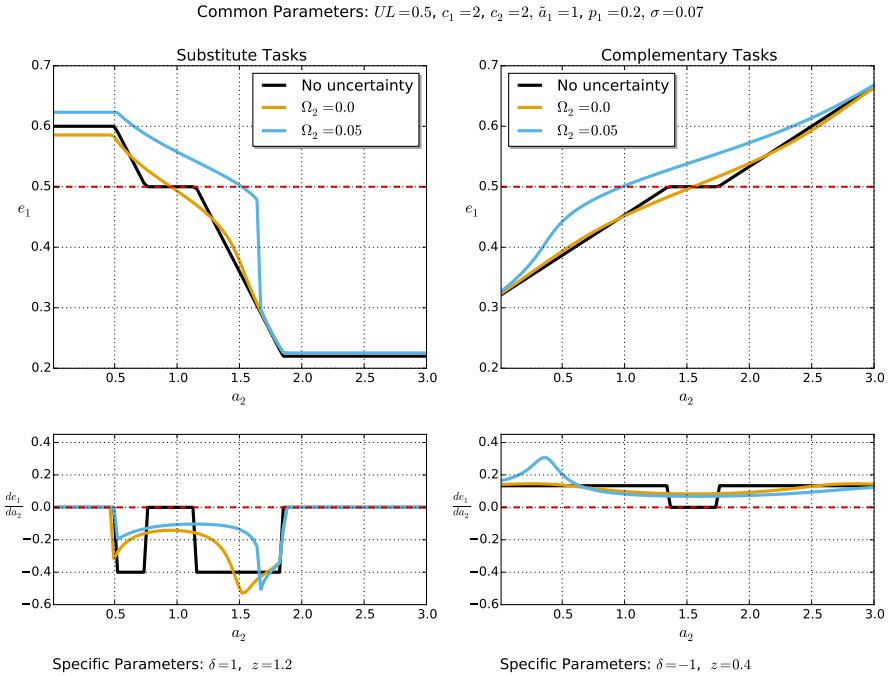
Proposition 4. *In the presence of uncertainty on the task result, and if $f(\cdot)$ corresponds to a symmetric unimodal distribution with mean 0, $\frac{de_1}{da_2}$ becomes larger in absolute value as e_1 moves away from $\frac{1}{p_1}\Omega + UL$.*

This proposition replaces Proposition 2, as $\frac{de_1}{da_2}$ is not required to be 0 at UL anymore. The denominator in Equation 4.6 is the same as in Equation 4.13, so the same attenuation pattern when e_1 is just above UL can be expected. The main difference is that the sign is given by parameter C_{12} and that $f(\cdot)$ and risk aversion are present only in the denominator. Figure 4.4 presents two additional examples. The graphs on the left correspond to a cost function that exhibits substitution between tasks, while the ones on the right come from complementary tasks. The top graphs show optimal effort exerted on task 1 as a function of the price of task 2, for each of the cost functions and considering no-uncertainty (black), uncertainty (orange) and risk aversion (light blue). In the second row, the figure presents the first derivative of the graphs above, $\frac{de_1}{da_2}$. In both types of cost function, the derivatives are closest to zero when $e_1 = UL$ or is above it. For the case of substitutes, there are two additional cases in which the derivative is zero; those are corner solutions in which either $e_2^* = 0$ or $e_2^* = 1$.

4.3 Empirical Test

In this section we present how we implement the test for determining whether a specific task is a complement or a substitute of a set of tasks for which there was an observed variation on the reward per unit of effect. The general concern is that if such a shock to the system occurs, normally it should affect all agents who are under the same contract. It involves two steps. First, two tests are presented in order to determine if there is bunching at the upper limit of a given indicator.

Figure 4.4: Simulation exercise: $x_1(a_2)$



Note: The cost function is defined as $C(e_1,e_2;z,c_1,c_2,\delta) = \frac{1}{z} \cdot (\frac{1}{2}(c_1e_1^2 + c_2e_2^2) + \delta e_1e_2)$. For the cases with uncertainty, $x_1 = e_1 + v_1$ where $v_1 \sim N(0,\sigma)$

It this is the case, for this specific indicator we can establish a set of agents that will not react to a variation in the reward per unit of effort in other tasks. These agents, who bunch themselves above UL, constitute a control group that motivates a difference-in-differences (DiD) approach (Equation 4.20 below). As a treatment group, agents that originally reported a level of output below the kink point UL are selected.

In the subsections below, the motivation and identification arguments for the DiD are discussed first, and the tests for bunching afterwards.

4.3.1 Test Specification

Our object of interest is the sign of C_{12}. According to proposition 1, the sign of C_{12} is the same as the sign of $\frac{de_1}{da_2}$. In this section, we explain how we can use data from a random sample of agents to estimate the sign of $\frac{de_1}{da_2}$ (and hence the sign of C_{12}).

Assume that we have available a random sample of N agents, observed consecutively for three time periods ($t = 1,2,3$). For each agent and time period we observe their task 1 output, that is, $\{x_{1it}\}_{i=1,t=1}^{N,3}$. Assume that the payment func-

tion for output x_1 is exactly as (4.8) and is the same in the three time periods. On the contract, assume that the piece rate for task 2 output is the same in the first two time periods, but changes in the third time period: a_{2t}=$a_2^{'}$ if $t = 1,2$; and a_{2t}=$a_2^{''}$ if $t = 3$. Without loss of generality, we assume that $a_2^{''}$<$a_2^{'}$.

We will represent agent i's observed level of task 1 output at time t by:

$$x_{1it} = e_1^* (a_{2t}, z(i)) + \theta_{1i} + \lambda_{1t} + \epsilon_{1it}, \tag{4.14}$$

where $e_1^* (a_{2t}, z(i))$ represents agent i's effort choice on task 1 when he faces a_{2t} as task 2 piece rate, and $z(i)$ is agent i's efficiency parameter.[13]We allow for the measured level of x_{1it} to differ from the agent's optimal choice due to a agent fixed component, θ_{1i}, a time component common across agents, λ_{1t}, and an independent and identically distributed random error term ϵ_{1it}, which exhibits zero mean and finite variance.

Using the above, the change in agent i's observed task 1 output between the third and second time period is given by:

$$x_{1i3} - x_{1i2} = e_1^* \left(a_2'', z(i)\right) - e_1^* \left(a_2', z(i)\right) + \lambda_{13} - \lambda_{12} + \epsilon_{1i3} - \epsilon_{1i2},$$

where we are using that the task 2 piece rate, a_{2t}, changed from $a_2^{'}$ to $a_2^{''}$ between these two time periods. We will specialise the above expression according to whether the agent's efficiency parameter, $z(i)$, is such that $z(i) \in [\underline{z}, \bar{z}]$, and hence agent i's optimal effort corresponds to the kink ($e_1^* = UL$), or when $z(i) < \underline{z}$, and hence the exerted effort is higher. Moreover, we assume that $a_2^{'}$, and $a_2^{''}$ are sufficiently close, so that $e_1^* (a_2', z(i)) = e_1^* (a_2'', z(i)) = UL$ if $z(i) \in [\underline{z}, \bar{z}]$ (see proposition 2). This means that $e_1^* (a_2'', z(i)) - e_1^* (a_2', z(i)) = 0$ for the group of agents for which $z(i) \in [\underline{z}, \bar{z}]$. Hence, we have that:

$$x_{1i3} - x_{1i2} = \lambda_{13} - \lambda_{12} + \epsilon_{1i3} - \epsilon_{1i2} \quad \text{if} \quad z(i) \in [\underline{z}, \bar{z}] \tag{4.15}$$

[13]For ease of notation, we do not make explicit that the agent's optimal choice of task 1 effort, $e_1^* (\cdot)$, also depends on the payment function of x_1 as well as agent $i's$ cost function. These elements are assumed to be constant along the sample period.

$$x_{1i3} - x_{1i2} = e_1^*\left(a_2'', z(i)\right) - e_1^*\left(a_2', z(i)\right) + \lambda_{13} - \lambda_{12} + \epsilon_{1i3} - \epsilon_{1i2} \quad \text{if} \quad z(i) < \underline{z} \tag{4.16}$$

Taking expectations of (4.15) and (4.16) over the relevant group of agents, and subtracting one from the other, we have that:

$$\Delta = E_{i\epsilon\{i:z(i)<\underline{z}\}}[x_{1i3} - x_{1i2}] - E_{i\epsilon\{i:z(i)\in[\underline{z},\bar{z}]\}}(x_{1i3} - x_{1i2}) = E_{i\epsilon\{i:z(i)<\underline{z}\}}[e_1^*\left(a_2'', z(j)\right) - e_1^*\left(a_2', z(j)\right)] \tag{4.17}$$

Note that the left hand side of (4.17), $E_{i\epsilon\{i:z(i)<\underline{z}\}}[e_1^*(a_2'', z(j)) - e_1^*(a_2', z(j))]$, is the discrete approximation to $(-\frac{de_1}{da_2})$ (averaged over the set of agents i for which $z(i) < \underline{z}$), whose sign is the same as the sign of C_{12}, our object of interest, and hence the sign of C_{12}. [14] We can estimate the sign of $E_{i\epsilon\{i:z(i)<\underline{z}\}}[e_1^*(a_2'', z(j)) - e_1^*(a_2', z(j))]$, by estimating the sign of the coefficient γ_1 in the following difference-in-difference regression:

$$x_{1i3} - x_{1i2} = \gamma_1 \mathbb{1}(z(i) < \underline{z}) + v_{ijt} \tag{4.18}$$

which implicitly uses the idea that those agents whose $z(i)$ is between $[\underline{z}, \bar{z}]$ can be used as a control group, because they choose to be at the kink of the payment function of $x_1$and hence are insensitive to small changes in a_2, the piece rate of the other task: x_2.

A problem with implementing (4.18) is that neither $z(j)$ nor $\underline{z}$ will generally be observable to the econometrician. To address this problem, one could estimate the following regression diff-in-diff regression:

$$x_{1i3} - x_{1i2} = \beta_1 \mathbb{1}(x_{1i2} < UL) + v'_{ijt} \tag{4.19}$$

where we are using the idea that those agents whose $z(i) < \underline{z}$ are those that have a output level below the kink ($x_{1i2} < UL$), because the individuals that choose to produce at the kink (UL) are those with $z(i) \in [\underline{z}, \bar{z}]$. While it is feasible to estimate (4.19), a problem is that x_{1i2} depends on the random component $\varepsilon_{1,i2}$,

[14] Note that we place a minus in front of $\frac{de_1}{da_2}$ because we assumed that $a_2'' < a_2'$.

which introduces a bias due to mean reversion. That is, there might be agents for which $e^*_{1i2} > UL$ but due to a large negative transitory shock, $\varepsilon_{1,i2} < 0$, they end up with $x_{1i2} < UL$. In the following time period, $t = 3$, we expect x_{1i3} to be larger or equal to UL, even if $a_{2t}$was the same in both $t = 2$ and $t = 3$. To net out this mean reversion bias, we need to estimate the following regression:

$$x_{1it} - x_{1it-1} = \alpha_1 \mathbb{1}(x_{1it-1} < UL) + \alpha_2 \mathbb{1}(t=3) + \alpha_3 \mathbb{1}(x_{1it-1} < UL) \cdot \mathbb{1}(t=3) + v''_{ijt},\ t = 2,3 \tag{4.20}$$

where the estimate of α_1 absorbs the mean reversion effect, and the sign of the estimate of α_3 will have the same sign as $E_{i\epsilon\{i:z(i)<\underline{z}\}}[e^*_1(a''_2, z(j)) - e^*_1(a'_2, z(j))]$, and hence the same sign as C_{12}.

4.3.2 Detection of Bunching

It is necessary to construct a counterfactual distribution of achievement in order to detect the existence of bunching. First, we consider the basic strategy for bunching developed by Kleven (2016): fit a parametric model on the observed distribution excluding an interval around UL, and compare it with the observed distribution. Moreover, if financial rewards play a big role in effort allocation, they will affect the entire shape of the distribution above UL, not only an interval around the threshold. For this reason, we borrow a concept from regression discontinuity design. Essentially, if agents' effort is the main driver of achievement, this will produce not only bunching at UL but a discontinuity on the density at that point. By running a standard McCrary (2008) test, we can determine if this is the case for a given estimator without imposing an assumption on the endogenous shape of the density.

In both exercises, our output variables are the histograms of the indicators. For this purpose we define bins on achievement following McCrary's procedure $(\tilde{x}_h)$ and count the number of agents in each bin (n_{hj}).[15]

[15]More precisely,

$$n_{jh} = \sum_{i=1}^{N} \mathbb{1}\left\{ \frac{\tilde{x}_h - \tilde{x}_{h-1}}{2} \leq x_{ij} < \frac{\tilde{x}_{h+1} - \tilde{x}_h}{2} \right\},\ \tilde{x}_h \in \{0.5, 1, 1.5, ..., 99.5\}$$

Bunching strategy We fit restricted cubic splines on the histogram excluding the interval $[UL_j, UL_j + L]$.[16] This strategy essentially splits the domain into segments defined by K knots (joint points) in order to fit the histogram (n_{hj}) of indicator j with a piece-wise cubic polynomial in the middle segments, and a linear function in the first and last ones. It requires the transformation of the domain variable (the midpoint of the bins, $\tilde{x}_h$) into $K-1$ constructed variables $\left(X_{jh}^{(k)}\right)$ that ensure that the resulting function's first and second derivatives are the same.[17] Such variables are included in the linear expression presented in Equation 4.21 which also considers dummy variables that indicate the presence of an excluded bin $\left(\mathbb{1}\{\tilde{x}_h = l\}, \forall l \in [UL_j, UL_j + L]\right)$. The error term, u_{jh}, is assumed to be i.i.d. and normally distributed.

$$n_{jh} = \sum_{k=1}^{K} \omega_k X_{jh}^{(k)} + \sum_{l=UL}^{UL+L} \gamma_l \mathbb{1}\{\tilde{x}_h = l\} + u_{jh} \tag{4.21}$$

After the vector of parameters $\{\omega, \gamma\}$ is estimated, the counterfactual density is the predicted value of this equation without the dummies for the excluded range's contribution: $\hat{n}_{jh} = \sum_{k=1}^{K} \hat{\omega}_k X_{jh}^{(k)}$. Then, the excess number of observations that bunch above UL relative to the calculated counterfactual is the difference between the observed and counterfactual histograms in the excluded range. This is equivalent to the sum of the omitted dummies γ:

$$\tilde{b}_j = \sum_{l=UL}^{UL+L} \hat{\gamma}_l = \sum_{l=UL}^{UL+L} \left(n_{jh} - \hat{n}_{jh}\right)$$

Following Chetty et al. (2009), we compare the amount of excess bunching with the average density per 1 pp. in the excluded range

$$b_j = \frac{\tilde{b}_j}{\frac{1}{L+1}\sum_{l=UL}^{UL+L} \hat{n}_{jh}}$$

In case there is bunching, the estimated b_j overestimates the amount of it. The

[16] While Kleven (2016) recommends polynomials, such functions might produce poor approximations in certain cases (Harrell, 2015, Chap 2.4.2). Spline interpolation is a parametric approach that is as easy to implement as a polynomial, without several of its limitations.

[17] The procedure was implemented in STATA 13 using mkspline command, using 5 to 7 knots determined by percentiles recommended in Harrell (2015, Chap 2.4.6).

reason is that it does not consider that some of the bunched observations in the interval $\left[UL_j, UL_j + L\right]$ should be above $UL_j + L$ in the counterfactual distribution, as predicted by the model.[18] As our goal is to determine whether or not there is bunching, we perform a joint significance test of the omitted dummies from Equation 4.21:

$$H_0 : \sum_{l=UL}^{UL+L} \hat{\gamma}_l = 0 \tag{4.22}$$

RDD strategy In the context of the regression discontinuity design (RDD), McCrary (2008) introduced a test for the continuity of the log-density $g(x)$ at a given point:

$$\iota = \ln \lim_{\tilde{x} \downarrow UL} g(\tilde{x}) - \ln \lim_{\tilde{x} \uparrow UL} g(\tilde{x})$$

The basic idea behind it is that if a treatment were assigned according to being above or below such a point, individuals would try to 'choose' their position in the domain in order to obtain or avoid the treatment. Such self-selection would induce a discontinuity on the density. In the bunching literature a discontinuity is not necessary as it allows for a noisy relationship between individual choices and observed outcomes. However, if such a noise is not present, the excess of density at one point will induce a drastic change in the density at such a point.

The estimation of the jump on the log-density, $\hat{\iota}$, is undertaken following McCrary's procedure. First, the bin size is determined according to the standard deviation of the indicator and the total number of indicators. Second, a bandwidth is selected based on the non-parametric estimator literature.[19] Given the bandwidth, local linear regressions are fitted to both sides of UL. Finally, the estimator tests whether the fitted function is continuous at UL.

4.3.3 The importance of bunching

The presence of the kink at UL is essential for the test. If there were no corner solution near this point, the expression in Equation 4.17 would deliver mislead-

[18] Chetty et al. (2009) correct for this using an iterative procedure in which the area above $UL_j + L$ is artificially increased in such a way that the area under both the observed and counterfactual densities is the same.

[19] In a few cases, the suggested optimal bandwidth is beyond the domain of the indicator (i.e. upper limit above 100%). In such case, we set the bandwidth to be equal to $100 - UL$.

ing results. From Equation 4.6, and assuming as before that the sole source of heterogeneity is the efficiency parameter z, we can derive the predicted sign of Equation 4.17 if the UL does not produce bunching. As shown in Figure 4.2, a higher level of z implies a higher level of e_1. As a result, when we compare the e_1^* response to variation in a_2 for an agent with high x_1 with one with low x_1, we are comparing an agent with a high vs. low value of z. Then, the essential question here is how $\frac{de_1}{da_2}$ changes along z. Equation 4.23 answers that question, and shows that its sign is determined by the sign of C_{12}, just like the derivative itself. For substitutes ($C_{12} > 0$), the derivative is negative ($\frac{de_1}{da_2} < 0$) and becomes even more negative with higher values of the productivity parameter $\left(\frac{d^2e_1}{dzda_2} < 0\right)$. For complements the opposite is true.

$$\frac{d^2e_1}{dzda_2} = -\frac{C_{12}}{C_{11}C_{22} - C_{12}^2} \tag{4.23}$$

Equation 4.23 has a strong implication for the test described above. Essentially, if the sorting is based on overall productivity, z, and there is no bunching, the term Δ presented in Equation 4.17 will produce a result that is opposite to the test result. In order to illustrate this, let us compare the response of two practices, one below UL with a productivity $\underline{z}$ and the other above such a cut-off with $\bar{z}$. The sorting of e_1^* implies that $\bar{z} = \underline{z} + \iota$, where $\iota > 0$. As shown below, if we approximate Equation 4.17 with derivative, it is clear that the sign of Δ is the same as the sign of C_{12}, exactly the opposite result from the one stated in the test description.

$$\begin{aligned}
\Delta &= E_{i\epsilon\{i:z(i)<\underline{z}\}}[x_{1i3} - x_{1i2}] - E_{i\epsilon\{i:z(i)\in[\underline{z},\bar{z}]\}}(x_{1i3} - x_{1i2}) \\
&\approx \frac{de_1(\underline{z})}{da_2} - \frac{de_1(\bar{z})}{da_2} \\
&= -\frac{\underline{z} \cdot C_{12}}{C_{11}C_{22} - C_{12}^2} + \frac{\bar{z} \cdot C_{12}}{C_{11}C_{22} - C_{12}^2} \\
&= (\bar{z} - \underline{z})\frac{C_{12}}{C_{11}C_{22} - C_{12}^2} \\
&= \frac{\iota \cdot C_{12}}{C_{11}C_{22} - C_{12}^2}
\end{aligned}$$

The previous derivation was based on particular sorting with respect to overall efficiency z. However, sorting might be along other dimensions so no reliable test can be derived based on such a difference. For instance, if heterogeneity is only based on the efficiency of task 2, Δ might always be negative regardless of the sign of C_{12}. See the example in Appendix 4.A.1 for more details.

4.4 An application: The Quality and Outcomes Framework

4.4.1 Background

The program that we analyse, the *Quality and Outcomes Framework* (QOF), was introduced in 2004 as part of major reform with the aim of improving service and reducing inequality in the quality of care received. It is a financial reward system for achieving a set of administrative and clinical goals. The level of achievement of these goals is monitored by a regional commissioner. Every year, the NHS and the physicians trade union, the *British Medical Association*, negotiate which indicators should be included and how much money should be paid for each one. Rewards are defined according to a point system, which is based on indicators. Administrative indicators are usually binary questions, where the practice obtains all of the points assigned to an indicator if a certain requirement is fulfilled. On the other hand, most clinical indicators are a non-linear function of the proportion of patients that received a certain standard of care. This will be explained in detail in the next section. Changes to the system have been proposed by the *National Institute for Health and Care Excellence* (NICE), but still have to be negotiated by the interested parties. These indicators are one of the most significant contributions of the program, as they provide an image of the quality of primary care services that was not available before. All the information is published yearly by the NHS at GP practice level and is the main source of data for the present study.[20]

Clinical indicators are related to management of chronic diseases and public health concerns. They cover chronic patients that require specific treatments such as those with coronary heart disease, heart failure or diabetes. Moreover, it involves lifestyle advice for smoking, obesity and primary prevention of cardiovascular diseases in general. Since their introduction, several areas have been removed or introduced or indicators replaced.

[20]Currently date is archived by *NHS Digital* at `http://digital.nhs.uk/qof`.

Analysis of multitasking on the QOF starts with the introduction of the system. The first order concern was to determine whether the programme had a negative impact on unmeasured (thus, unrewarded) indicators of care, one of the possible outcomes predicted by Holmstrom and Milgrom (1991) and Baker (1992). Sutton et al. (2010) studied a panel of medical records collected before and after the introduction of the programme in Scotland, which included both rewarded and unrewarded outcomes. They claim that after the introduction of the programme there was an improvement in record-keeping for both type of outcomes with respect to the pre-programme trend, but this was larger for those rewarded measures. This was the case for recordings on blood pressure, cholesterol and smoking, which were rewarded, against BMI and alcohol consumption, which were not. Doran et al. (2011) did a similar exercise for a sample of practices in England, but in this case they had access to prescription and biomarkers data, and they obtained similar results. In both studies, as unrewarded measures are affected by the reallocation of effort generated by the introduction of rewards, the identification of the effects of multitasking relies on the validity of using extrapolated pre-treatment trends as a counter-factual. This has also motivated theoretical work on the optimal design of the system. such as Eggleston (2005) and Kaarboe and Siciliani (2011).

As the QOF is adjusted almost every year, a second generation of the analysis followed these innovations. A first set of changes was introduced in 2005/06, where the payment thresholds were revised for some indicators making it more difficult to achieve the maximum number of points. Feng et al. (2015) compared the evolution of the modified and unmodified indicators in Scotland, and showed that performance increased for the affected measures.

A final element to consider is gaming of the system. The main concern is called exception reporting for clinical indicators, which consists of declaring that a patient should not be treated according to the QOF guidelines due to specific health conditions. By increasing the number of excepted patients, the relevant indicator will increase without providing extra services. Gravelle et al. (2010) showed that GP practices exempt relatively more patients from being considered for some of the clinical indicators if the overall achievement in the previous year

was below UL, than if it was above this threshold. For our purposes, cheating implies that some practices with productivity $z_0 - \eta$ would report having productivity z_0. This would be a problem for our estimates if those cheating above UL adjusted their reported effort in response to changes in the price of alternative tasks.

Panel A of Table 4.1 presents the number of practices in the financial years 2009, 2010 and 2011 and their average number of patients (list size). There are around 8000 GP practices covering on average 7000 patients. Panel B shows the mean achievement per domain in each year, which is very close to 100% in all years. The big increase from 2010 to 2011 is due to the removal of some of the indicators, which will be discussed in the next section. Panel C presents the total clinical points (2009) assigned to those conditions with the highest prevalence in the population, according to the QOF data reports. Such points assignments provide an idea on the areas where the NHS considered it a priority to improve and standardize health care. In 2009, diabetes was the most rewarded clinical area with 100 points out of 697 available for the clinical indicator, followed by hypertension and CHD. While these are also some of the most common chronic conditions, relevance is not the sole criteria. For instance management of new cases of depression in the previous years received more points than asthma, even though the latter was the second most common chronic disease after hypertension.

4.4.2 Payment system

In our analysis we will consider that for a GP practice, the marginal benefit of exerting effort on a task is a linear function that involves both altruism and monetary payments. Hence, the marginal reward above UL for task j, which we called $\bar{a}_1$ in subsection (4.2.2), refers to the altruistic motive.[21] Our analysis is based on data from the years 2009 to 2011. In 2009 and 2010, GPs could obtain up to 1000 points: 697 for the clinical domain, 167.5 for the organizational domain, 91.5 for patient experience, and 44 for additional services. In 2011, the clinical domain was reduced to 661 and patient experience to 33, and 262 points were rellocated to organizational indicators. Points are translated into income depending on the

[21] The assumption of a linear benefit to patients' welfare is relaxed by Kaarboe and Siciliani (2011). In such a scenario, the relevant function is not $C(\cdot)$ but $B(\cdot) - C(\cdot)$, hence our results will signal complementarity or substituiability of this function.

Table 4.1: GP Practices and QOF Descriptives

Panel A: Main Characteristics	Average by practice and year		
	2009	2010	2011
Number of patients (list size)	6602.84	6691.28	6835.62
Number of practices	8305	8359	8124
Panel B: QOF achievement	Average by practice and year		
	2009	2010	2011
Clinical	95.86	96.75	97.01
Organisational	96.34	97.36	96.37
Patient Experience	71.47	72.60	98.95
Additional Services	95.35	97.13	97.02
Total	93.69	94.66	96.91
Panel C: Selected Raw Prevalences and QOF points for 2009			
	Points	Mean	Std Dev
Diabetes †	100	4.28	1.85
Hypertension	81	13.53	4.79
Asthma	45	5.95	2.29
Coronary Heart Disease	87	3.45	1.49
Depression new cases †	53	0.76	0.80

Notes: Own calculations based on QOF data published in NHS Digital. † Diabetes raw prevalence is underestimated as it is calculated as the number of individuals aged 17 and over with diagnosed types I or II, over the total list size (without age distinction). New cases of depression are those patients diagnosed with the disease during the last financial year (April 1 to March 31).

size of the practice and how common the underlying health condition is in the practice's population.[22]

Monetary payments in the QOF are determined by achievement according to a set of indicators, of which there are two main types: binary and ratios. The former gives a fixed amount of points if a condition is attained.[23] For instance, indicator BP1 gives 6 points if there is a register of people with established hypertension, or 0 points if there is not. On the other hand, the awarded points for ratio based indicators depend on the number of patients that should potentially receive a given treatment (denominator), and the number of those who effectively receive it (numerator) during a specific period of time.[24] For instance, the definition below for indicators DM17 and ASTHMA6.

> **Indicator ASTHMA6:** The percentage of patients with asthma who have had an asthma review in the previous 15 months

[22]See Appendix 4.B for further details.

[23]Some administrative indicators also involve ratios. For instance, if there are less than 5 years of records of the blood pressure of patients for 80% of the patients aged 45 and over (indicator RECORD17). In those cases, the number of points allocated follow a binary allocation instead of a piece-rate reward system.

[24]In principle, payment is retrospective, but it is possible to obtain advance payments based on previous year's performance, which are known as *aspiration payments*. More details are available from the BMA (2013).

> **Indicator DM17:** The percentage of patients with diabetes whose last measured total cholesterol within the previous 15 months was 5 mmol/l or less

If achievement is below a lower limit (LL_j) zero points are awarded, and if it above the upper limit (UL_j) the maximum amount of available points for indicator j are awarded.

Returning to the DM17 indicator example, the lower limit is $LL = 40\%$ and the upper limit is $UL = 70\%$. Then, if at least 70 out of every 100 patients with diabetes have total cholesterol of 5mmol/l or less in the last 15 months, the practice will receive 6 points, the total number of points allocated to this indicator. For ASTHMA6 there are 20 points available and it has the same thresholds $LL = 40\%$ and $UL = 70\%$. A graphic representation of such an assignment rule is presented in the top diagrams in Figure 4.5, where the horizontal axis presents the possible levels of achievement and the vertical axis represents the number of points that would be awarded according to the QOF rules. Figure 4.5 also presents histograms for the actual achievement attained by GP practices in each indicator for the 8301 practices in the 2009/10 financial year.[25] From these densities, there are two main points to remark on. First, there are few practices at or close to the lower limit LL; and in fact, most of the distribution is above the UL. The mean achievement for ASHTMA6 was 80% and 83% for DM17 (see Table 4.2). Less than 6% of the practices attained a level below UL for ASTHMA6, while for DM17 this figure was 2.5%. This is a common element in all indicators that initially exceeded the expectations of the policymakers (Gregory, 2009). As a result, the main focus of this project is the role of the UL, hence the LL will not be discussed.

Second, as seen for the case of ASHTMA6, there is a sudden increase in the density at UL; in other words, there is *bunching* above the threshold, which is an usual feature of the data produced by discontinuities in budget constraints (Saez, 2010). However, this is not the case for all of the indicators. This seems to be the case of indicator DM17. According to the model discussed before, this might be either because the financial reward has a minimum impact on the motivation of physicians for accomplishing the goal or due to substantial noise between effort

[25] This includes practices without any cases of hypertension (5 practices) or asthma (8 cases). In those scenarios, zero points are given.

and the measured achievement indicator.[26] Another typical reason for not detecting bunching, the measurement error (Kleven, 2016), is a problem for the present study as the QOF data are based on administrative records for a large number of GP practices.

The other main source of variation in the data is time. Given that between 2009/10 and 2010/11 there were not changes to the QOF indicators, we can understand how achievement changes from period to period. First, while achievement is persistent, there is substantial year-to-year variation. The autocorrelation coefficients are 0.54 for ASTHMA6 and 0.6 for DM17. Second, practices below the UL in one year tend to increase their achievement in the next one. The mean variation for ASTHMA6 is 11 pp. (SD = 14.8 pp.) for those practices below the UL in 2009, but it is -0.2 pp. (SD = 6.7 pp.) for those above it. Such a mean difference is different from 0 at the 99% level. The same happens for DM17, but with a difference of means of 9 pp. Descriptive statistics for the other indicators are presented in Table 4.7 in the appendix as the pattern is the same.

Table 4.2: DM17 and ASTHMA3 QOF indicators descriptives for 2010/11

Indicator	UL	(1) Number	(2) $E[x_t]$	(3) $\mathrm{P}[x_t < UL]$	(4) $\rho(x_t)$	(5) $E[x_t - x_{t-1} \mid x_{t-1} < UL]$	(6) $E[x_t - x_{t-1} \mid x_{t-1} > UL]$
ASTHMA06	70%	8245	79.58	5.29	0.54	11.03	-0.19
DM17	70%	8245	82.73	2.43	0.60	8.70	-0.55

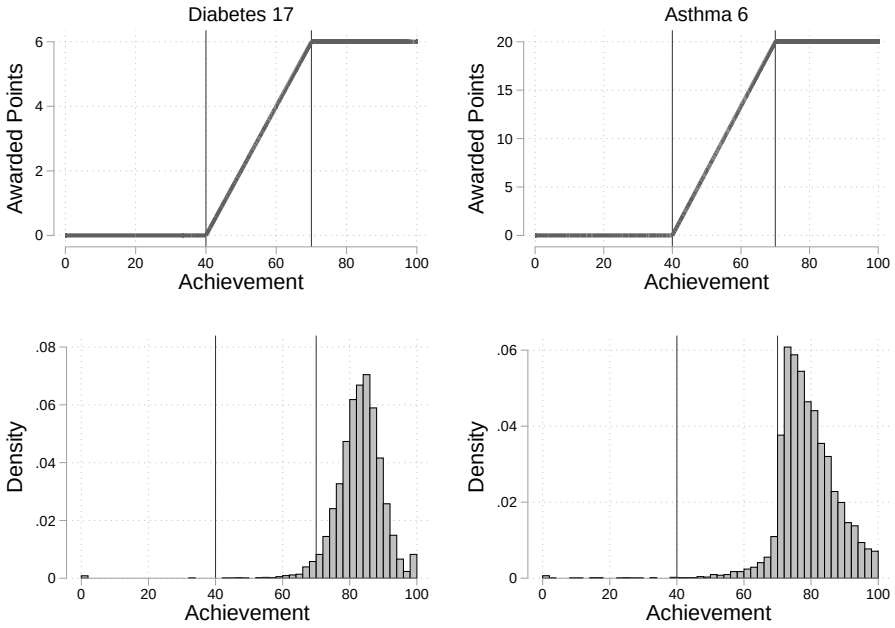
Notes: Own calculations based on QOF data. **Number:** Number of GP practices, including those with 0 elegible patients for the given indicator. $E[x_t]$ **:** Average achievement per indicator. $\mathbf{P}[x_t < UL]$ **:** Proportion of practices with an achivement below UL. $\rho(x_t)$ **:** Correlation between 2010 and 2009 achivement.

4.4.3 The 2011 changes

While QOF is normally revised every year, there was no change between 2009 and 2010 after an agreement between BMA and NHS during the H1N1 vaccination program (NHS Employers, 2010). However, between 2010 and 2011 there were major changes that we will interpret as a net reduction in the financial reward per

[26] For instance, the staff of the GP practice might have complete control in keeping records of tests or ensuring that patients with a given condition are prescribed a given drug. However, ensuring that the levels of cholesterol of their patients are within certain range, as required by indicator DM17 discussed before, might depend on many actions not controlled by providers. Indeed, Fichera et al. (2014) present a game in which physicians and doctors interact using their available tools, prescriptions and lifestyle, in response to QOF incentives.

Figure 4.5: Points reward function and achievement density for Diabetes 17 (DM17) and ASHTMA6 (2009/10)



Notes: Own calculations based on QOF data archived at NHS Digital.

unit of effort for part of the clinical indicators. This time-frame between 2009 and 2011 will be the main source of data for our analysis.

There are in total 1000 QOF points in all three years, but several indicators were either removed, modified or replaced by new ones. We have summarized them in three broad categories presented in Table 4.3. First, those that imply a reduction in the financial reward per unit of effort; second, those that we interpret as an increase in the marginal benefit; and third, those whose nature is ambiguous. A more detailed explanation of these changes is presented in Table 4.8 in the appendix.

In the first category (reduction in the financial reward per unit of effort), we include indicators that are withdrawn,[27] increases in UL (which will obviously flatten the slope of the reward function)[28] or changes that consisted of a reduction

[27]Clinical retired indicators were almost a requirement for measuring other QOF indicators. For instance, indicator CH5 was about having a recent blood pressure record for patients who suffered from coronary heart disease but CHD6 rewards practices for keeping the blood pressure of these patients controlled.

[28]See Equation 4.25 in the appendix. While the initial proposal was to redefine the UL and make them a function of the underlying indicator distribution in 2011 (match the 75th percentile), the negotiations delivered a slow-paced plan. By 2011 two ULs had increased by one pp. However in

in the number of points allocated to the indicator. In total 143 of the original clinical points are affected. A different type of change also implied a reduction in the financial reward per unit of effort: these were wording amendments in which the goal definition changed to require either additional tasks or reduce the reference time of the indicator.

The second category (ambiguous change) covers several word amendments that are not straightforward to classify. In these cases typically a more precise definition of the goal to be accomplished is accompanied by additional points in compensation. In total 51 of the original points are in this category.

The third category (increase in the financial reward per unit of effort) includes new indicators as well as old ones with goals that are easier to achieve. The new indicators, covering 12 points, refer to tasks that were not financially rewarded before. Also, for one indicator (17 points) the new wording relaxed the goal defined in the original version.

As we can see, in terms of clinical indicators, the total amount of points related to a reward drop are larger than those associated with an increase, even if we consider all ambiguous changes as increases. Hence, we interpret the overall changes in 2011 as an overall reduction in the marginal payment per unit of effort.

Administrative indicators suffered a major modification in 2011. Two thirds of the *patient experience* domain were removed in favour of the new *quality and productivity* indicators. Practices had to agree a plan with the primary care organisations consisting of three main goals for prescribing (28 points), outpatient referrals (21 points) and emergency admissions (47.5 points). The exact indicator definition and its upper threshold was defined at local level. The objective of the indicators was to reduce costs for the PCT by improving the cost-efficiency of prescribing and by treating more patients at primary care level, reducing both referrals and emergency admission rates.

For the reasons given above, we consider that the main objective of the changes was to *tighten-up* the requirements for obtaining rewards, at least on the clinical side. We will not discuss the administrative indicators, given that almost an entire domain was replaced with an other: the perceived time for getting an

2012 both the lower and upper limits were increased by between 4 to 10 pp. for 13 indicators (Doran et al., 2014).

Table 4.3: Changes in QOF 2011 with respect to 2009-2010

Panel A. Clinical Indicators

Price Interpretation (Total Points)	**Status**	**Description**	**Points**
Reduction (143 to 87)	Withdrawn	No longer rewarded tasks	32
	Points reduced	Number of assigned points per indicator was reduced.	26 to 22
	Upper Limit Increased	Increase on UL	22
	Replacement I	New wording with more strict definition of a goal or a reduced time-frame for accomplishing it	18
	Replacement II	Decrease in points and new wording is more detailed	45 to 25
Ambiguous (51 to 59)	Replacement III	Harder to accomplish or more detailed goals but compensated with extra points	51 to 59
Increase (29)	Replacement IV	Reference cutoff relaxed	17
	New	New tasks to be rewarded	12
NA (486)	Replacement V	Similar or same wording, but expressed in new units or highlight recent changes on diagnostic procedures.	32
	Unchanged	No change on points, thresholds or wording	454

Panel B. Non-Clinical Indicators

Price Interpretation	**Status**	**Description**	**Points**
Reduction	Retirements	No longer rewarded tasks	60.5
Increase	New	New tasks to be rewarded	96.5
NA	Unchanged	No change on either points or wording	242.5

Note: Authors' interpretation based on NHS Employers public documents.

appointment was replaced with meetings related to prescribing and other supervised improvement plans designed by the PCT. Because these are administrative tasks, we assume that they were not carried out by doctors themselves and hence that they do not alter the marginal cost of clinical effort.

4.5 Results

The results are presented in two steps. First, we assess the validity of the test by checking for a discontinuity and/or for bunching at the upper limit (UL). Second, we test the sign of the response on effort to a price drop in alternative tasks, on those indicators that were not affected by the QOF 2011 changes.

For the bunching analysis, we pool data from both years 2009 and 2010 and set 10 pp. an estimation window below and above UL. We also discard the bins

corresponding to 100%, which is hard to fit with a continuous density function. Figure 4.6.A presents a graphical representation of the McCrary test for continuity on the density at UL for indicators DM17 and ASTHMA6, our examples discussed in the previous section. Both graphs present the histogram (n_{hj}), and the fitted models to both sides of UL . For ASTHMA6 there is clear evidence of the existence of a discontinuity as the null hypothesis that both approximated log-densities are the same at UL is rejected. In both cases the test suggest the presence of a discontinuity on the density at UL. For DM17 such a null cannot be rejected at the 95% level, but it is at the 90% level. Table 4.4 presents this exercise (Column 4) for each indicator (rows) given a McCrary's default calculations for bin size (Column 2) and bandwidth (Column 3).

The calculation of the amount of excess bunching for both indicators is presented in Figure 4.6.B. Apart from the histogram (n_{hj}), these figures present the fitted model including dummies γ covering $[UL, UL + 5\text{pp.}]$ (orange line) and excluding them from the prediction (black line). For DM17, the difference between the histogram and the counterfactual difference is of 42% of the average density in the interval; and for ASTHMA6 it is 107%. Both estimates are significant at the 95% level. However, such estimates are sensible to the number of knots in the spline, the excluded range size L, and the estimation window. Varying the configuration of such parameters we obtain very different point estimates. Columns 5 to 9 in Table 4.4 present several configurations of an excluded range from $L = 2$ to $L = 7$, 5 and 7 knots, and estimation windows of 10 and 20. For DM17 an estimate of b between -90% and 43%; and for ASTHMA6 it is around 60% to 417%. Despite such large differences, the null in Equation 4.22 is not rejected for ASTHMA6. On the other hand, for DM17 the null is rejected in 3 out of 5 of the explored specifications.

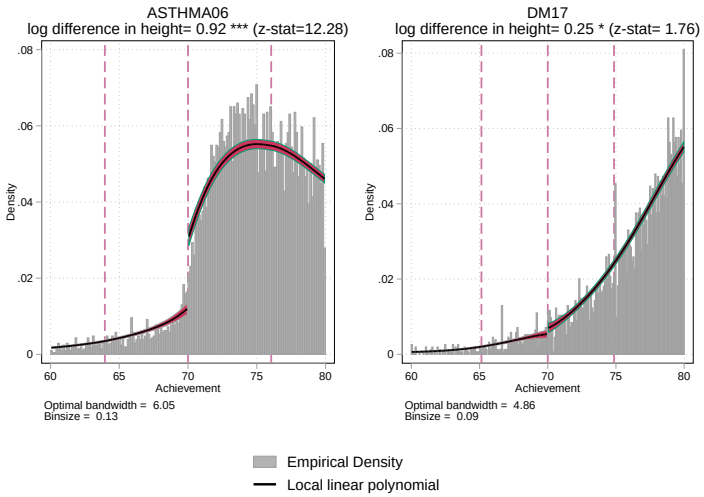
Given the results stated above, there is clear evidence that the upper limit has an effect on practices, effort allocation for ASTHMA6, but this is not as clear for DM17. Therefore the test is likely to be informative for the first but not the second indicator. Table 4.4 also suggests that for indicators DM22,[29] SMOKE3[30]

[29]Based on having a record of glomerular filtration rate (GFR), which measures kidney function.

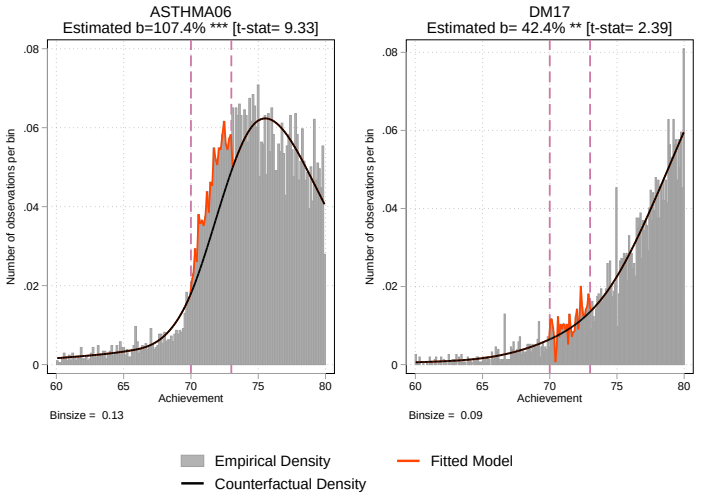
[30]Proportion of individuals affected by several chronic conditions who are referred to smoking cessation advice.

Figure 4.6: Testing for Bunching

A. McCrary Test



B. Excess bunching estimate



Note: In sub-figure (a), the empirical densities to both sides of the threshold UL are smoothed using a local linear regression within the given bandwidth of UL (vertical lines). These smoothed functions are presented with a 95% CI. In sub-figure (b), the empirical density is fitted with a restricted cubic spline based on 5 knots. Domain was restricted to a 10 pp. window around UL, and the excluded range is $[UL_j, UL_j + 5pp.]$

and THYROI02[31] there is no evidence of bunching. Table 4.9 in the appendix presents definitions and graphs equivalent to Figures 4.6.A and 4.6.B for these indicators.

[31] Record on thyroid function tests.

Table 4.4: QOF indicators corner test

Indicator	(1) UL	(2) BS	(3) BW	(4) DC Test	(5) w=10, h=2, k=5	(6) w=10, h=3, k=5	(7) w=10, h=3, k=7	(8) w=20, h=3, k=5	(9) w=20, h=5, k=5
AF03	90	0.06	6.38	1.82 ***	123.1 ***	242.4 ***	183.3 ***	176.9 ***	345.4 ***
				[30.72]	[6.53]	[10.96]	[6.11]	[9.34]	[9.73]
AF04	90	0.19	10.00	2.50 ***	171.4 ***	184.7 ***	237.4 ***	137.1 ***	183.7 ***
				[26.28]	[4.10]	[3.88]	[2.67]	[3.82]	[3.17]
ASTHMA03	80	0.11	10.00	1.76 ***	200.5 ***	154.5	89.1	150.4 **	295.5 ***
				[20.65]	[3.31]	[1.41]	[0.51]	[2.13]	[3.32]
ASTHMA06	70	0.13	6.05	0.92 ***	49.3 ***	107.4 ***	57.1 ***	125.4 ***	416.9 ***
				[12.28]	[4.68]	[9.33]	[5.19]	[6.36]	[19.66]
ASTHMA08	80	0.14	10.00	1.93 ***	205.8 ***	254.8 ***	193.4	256.6 ***	492.7 ***
				[27.49]	[4.98]	[3.42]	[1.59]	[5.11]	[8.02]
BP5	70	0.10	5.26	0.31 ***	13.9 *	37.5 ***	42.5 ***	24.1 ***	18.8 **
				[3.60]	[1.92]	[3.99]	[2.83]	[3.94]	[2.32]
CANCER03	90	0.28	10.00	1.41 ***	160.2 ***	235.6 ***	195.5 **	229.2 ***	352.1 ***
				[26.48]	[3.36]	[4.05]	[2.44]	[6.30]	[5.83]
CHD08	70	0.10	6.59	0.30 ***	37.8 ***	39.1 **	68.8 **	14.7	-56.2 ***
				[2.59]	[4.27]	[2.10]	[2.54]	[0.85]	[-2.91]
CHD09	90	0.05	4.57	1.20 ***	77.1 ***	136.2 ***	85.4 ***	57.7 ***	208.4 ***
				[18.55]	[8.68]	[11.54]	[5.81]	[5.48]	[9.06]
CHD10	60	0.15	7.69	1.30 ***	91.9 ***	128.3 ***	112.7 ***	112.4 ***	237.5 ***
				[12.46]	[7.18]	[8.89]	[4.60]	[6.19]	[9.17]
CHD12	90	0.08	4.82	1.21 ***	118.2 ***	222.2 ***	153.4 ***	161.5 ***	323.7 ***
				[22.56]	[10.79]	[20.70]	[17.18]	[16.06]	[14.33]
CKD02	90	0.04	3.42	0.83 ***	289.9 ***	141.9 ***	-1279.0 ***	398.2 **	-166.1 **
				[3.52]	[6.71]	[2.83]	[5.35]	[2.30]	[-2.02]
CKD03	70	0.13	7.73	0.69 ***	82.9 ***	136.9 ***	100.9 ***	164.0 ***	343.5 ***
				[16.24]	[10.73]	[11.64]	[6.81]	[11.02]	[18.86]
CKD05	80	0.16	10.00	2.40 ***	466.9 ***	249.9	83.9	243.7 *	413.4 **
				[18.73]	[3.21]	[1.11]	[0.29]	[1.72]	[2.48]
CKD06	80	0.22	8.89	0.85 ***	77.3 ***	113.6 ***	109.9 ***	109.8 ***	259.6 ***
				[16.84]	[5.82]	[5.70]	[3.40]	[5.11]	[8.42]
CVD01	70	0.28	10.00	1.11 ***	164.8 ***	254.3 **	200.1	179.7 **	499.0 ***
				[13.83]	[4.00]	[2.32]	[1.34]	[2.01]	[5.17]
CVD02	70	0.22	10.00	1.02 ***	97.4 ***	118.1 **	97.6	79.6 *	71.4
				[10.77]	[3.84]	[2.50]	[1.40]	[1.78]	[0.89]
DEM02	60	0.19	10.00	2.31 ***	366.9 ***	359.9 ***	717.4 *	282.1 **	338.2
				[13.60]	[4.56]	[4.71]	[1.92]	[2.40]	[1.30]
DM2	90	0.06	4.22	0.20 **	4.9	-34.2 ***	39.3 ***	-106.8 ***	-141.5 ***
				[2.33]	[0.53]	[-3.78]	[3.16]	[-10.34]	[-6.49]
DM10	90	0.12	4.98	0.74 ***	48.9 ***	108.5 ***	57.6 ***	58.9 ***	214.8 ***

Continued on next page

Table 4.4: (Continued)

Indicator	(1) UL	(2) BS	(3) BW	(4) DC Test	(5) w=10, h=2, k=5	(6) w=10, h=3, k=5	(7) w=10, h=3, k=7	(8) w=20, h=3, k=5	(9) w=20, h=5, k=5
				[13.74]	[4.80]	[9.03]	[6.26]	[5.46]	[9.35]
DM13	90	0.14	4.95	0.75 ***	74.0 ***	175.6 ***	113.8 ***	159.1 ***	400.7 ***
				[16.51]	[5.69]	[11.93]	[14.40]	[12.83]	[19.51]
DM15	80	0.13	10.00	1.45 ***	306.7 ***	223.7 *	172.0	194.6 **	262.3 ***
				[19.09]	[4.69]	[1.83]	[0.84]	[2.57]	[2.70]
DM17	70	0.09	4.86	0.25 *	0.0	42.4 **	42.6	42.5	-93.2 ***
				[1.76]	[0.49]	[2.39]	[1.60]	[1.54]	[-3.44]
DM18	85	0.09	5.60	0.46 ***	30.0 ***	27.7 ***	42.1 ***	14.5 *	65.9 ***
				[6.41]	[4.70]	[3.08]	[3.19]	[1.90]	[6.50]
DM21	90	0.12	5.30	1.07 ***	119.7 ***	222.1 ***	156.1 ***	186.2 ***	376.1 ***
				[22.39]	[9.26]	[17.40]	[14.53]	[15.90]	[21.36]
DM22	90	0.05	4.21	0.27 *	76.2 ***	34.1	601.1 ***	-9.8	-143.7 ***
				[1.87]	[5.44]	[1.63]	[5.28]	[-0.32]	[-2.82]
EPILEP06	90	0.11	8.54	1.38 ***	84.6 ***	66.9 **	181.1 ***	-3.9	-72.3 *
				[19.48]	[2.70]	[1.99]	[2.69]	[-0.18]	[-1.66]
EPILEP08	70	0.21	10.00	1.18 ***	145.7 ***	245.2 ***	208.6 **	250.8 ***	288.7 **
				[23.80]	[4.44]	[3.43]	[2.15]	[3.75]	[2.33]
HF02	90	0.17	10.00	2.41 ***	222.9 ***	259.5 ***	257.5 **	232.6 ***	336.4 ***
				[29.80]	[3.90]	[4.31]	[2.65]	[5.14]	[6.12]
HF03	80	0.11	10.00	2.19 ***	399.7 ***	240.8 *	246.3	221.1 **	293.6 **
				[21.12]	[4.66]	[1.72]	[0.95]	[2.40]	[2.50]
HF04	60	0.19	10.00	2.51 ***	662.3 ***	675.1 ***	-4475.9 *	580.8 **	274.0
				[11.53]	[3.67]	[4.94]	[1.92]	[2.39]	[0.68]
SMOKE03	90	0.04	3.15	0.14	-9.7	-91.0 ***	104.7 ***	-165.7 ***	-238.9 ***
				[1.17]	[-0.58]	[-6.60]	[5.32]	[-12.89]	[-10.12]
SMOKE04	90	0.08	4.86	1.32 ***	132.9 ***	299.2 ***	165.9 ***	272.4 ***	662.6 ***
				[24.79]	[9.08]	[18.04]	[14.88]	[18.08]	[26.30]
STROKE07	90	0.10	6.58	1.12 ***	98.1 ***	179.4 ***	125.2 ***	142.5 ***	308.2 ***
				[25.10]	[6.41]	[8.71]	[4.76]	[8.74]	[11.01]
STROKE08	60	0.13	9.60	0.47 ***	89.9 ***	116.4 ***	114.5	114.5 **	90.4
				[4.25]	[2.99]	[2.85]	[1.30]	[2.49]	[1.12]
STROKE10	85	0.11	7.79	0.98 ***	64.6 ***	103.9 ***	95.3 ***	113.3 ***	262.9 ***
				[17.87]	[4.18]	[5.84]	[3.64]	[7.16]	[11.52]
STROKE12	90	0.07	5.96	1.51 ***	88.3 ***	155.6 ***	128.2 ***	84.2 ***	217.0 ***
				[24.03]	[4.39]	[5.91]	[3.21]	[3.75]	[5.30]
STROKE13	80	0.19	10.00	2.32 ***	450.8 ***	267.3 *	165.3	259.9 **	455.9 ***
				[21.42]	[5.27]	[1.72]	[0.68]	[2.55]	[3.52]
THYROI02	90	0.05	4.48	0.04	-20.3 *	-50.7 ***	33.8	-122.2 ***	-144.6 ***
				[0.31]	[-1.97]	[-4.09]	[1.35]	[-10.19]	[-5.84]

Continued on next page

Table 4.4: (Continued)

Indicator	(1) UL	(2) BS	(3) BW	(4) DC Test	(5) w=10, h=2, k=5	(6) w=10, h=3, k=5	(7) w=10, h=3, k=7	(8) w=20, h=3, k=5	(9) w=20, h=5, k=5

Notes: Own calculations based on QOF data. McCrary test on the continuity of the density at the threshold. Optimal bin sizes (BS) and bandwidths (BW) for each indicator are chosen following McCrary implementation of the test. Significance: * 10%, ** 5%, *** 1%.

Table 4.5 presents the second part in which we estimate regression (4.20), where x_1 refers to indicators whose rewards remained unchanged throughout the three years that we consider (2009-2011). We exclude from the analysis those indicators in which the test would not be not valid because bunching was not detected. For each indicator (rows), the table reports the number of observations above and below the threshold within a 5 pp. window to both sides of the UL (Columns 1 and 2). Such is the selected sample for estimating the parameters of regression (4.20): columns 3 to 5 of the table presents estimates for α_1, α_2, and α_3. In order to estimate the model, first differences with respect to time are obtained for each GP practice between 2010 and 2011, and between 2009 and 2010. Such a variable is the outcome of the equation. We also construct a binary variable that indicates whether the practice was below the UL in years 2009 and 2010 $\left(\mathbb{1}(x_{1j,t-1} < UL_j)\right)$, and another that indicates whether we are observing data from the variation 2010 to 2011 $(\mathbb{1}(t = 2011))$. The sample is restricted to a window of $[UL - 5, UL + 5]$. In particular, we are interested in the sign of α_3. Given that we observed a net reduction in the marginal benefit of alternative tasks, a negative sign of $\hat{\alpha}_3$ indicates a positive cross-derivative $\left(\frac{de_1}{da_2} > 0\right)$ which indicates that the analysed task are complementary to the tasks affected by the 2011 changes. This does not mean that the task is a complement of all modified indicators, but that overall, the net response is equivalent to complements. Another possibility is that the task is a substitute only of those tasks for which the marginal reward was increased instead of reduced. This is less likely as the majority of changes correspond to a decrease, rather than an increase, but we cannot rule out such a possibility.

We also note that, for some indicators, we might not be able to reject the hypothesis that$\alpha_3 = 0$ because of lack of power. In particular, there are some

indicators that have a very small number of practices below the threshold. For instance, for HF04 there are only 45 practices below UL in comparison with 669 above it.

We find that AF04,CKD06, DM10, DM13 and EPILEP08 are complements of the overall modified indicators: effort was reduced in response to the net reduction in incentives in other indicators. The first (AF04) is the percentage of patients with atrial fibrillation (a rapid and irregular heartbeat) who had their diagnosis confirmed by an specialist or with a specialised test. The second (CKD06) is the percentage of patients with chronic kidney disease who have a record for a test that checks their kidney status. The third (DM10) is on having records of neuropathy testing (nerve disorders) and the fourth (DM13) is records of micro-albumuria testing (kidney's status) for diabetic patients. The last one, EPILEP08, is on having records on the percentage of epileptic patients under drug treatment who have been seizure free

Alternative estimation windows are considered in Table 4.6. In this table, each cell presented is an estimate of α_3 considering a sample of $[UL-l, UL+k]$. This table is restricted to those cases in which the hypothesis $\alpha_3 = 0$ is rejected at least once. This means that Column 3 of Table 4.5 corresponds to the fifth column $(l = 5, k = 5)$ of Table 4.6. Estimates for AF04, CKD06 and DM13 are stable across the different specifications. Table 4.10 in the appendix presents definitions and graphs with the bunching test for these indicators.

The diabetes mellitus (DM) area suffered several changes. There were changes in payments for keeping blood pressure of patients controlled and on records of foot examination. Also, financial rewards for keeping records of plasma glucose concentration, blood pressure and cholesterol were removed. Given that both DM10 and DM13 are also records of recent tests, it seems plausible that such tasks are complements.

Neither the chronic kidney disease nor the atrial fibrillation indicators were modified in 2011. Nevertheless, AF04 and CKD06 are affected by other indicators' changes. AF04 measures the proportion of individuals diagnosed with ECG or by a specialist. CKD06 rewards keeping a record of albumin creatinine ratio, which is a specific measure related to kidney disease.

Table 4.5: QOF indicators results: Window of 5 pp.

		(1)	(2)	(3)	(4)	(5)	
		Descriptives		Estim. Regression Coefficients			Classif.
Indicator	UL	N Below	N Above	BELOW α_1	AFTER α_2	INTER α_3	
AF03	90%	771	5747	0.027***	0.001	−0.001	
				(0.002)	(0.001)	(0.003)	
AF04	90%	433	3001	0.039***	−0.002*	−0.015***	Comp
				(0.003)	(0.001)	(0.005)	
ASTHMA03	80%	336	2505	0.031***	−0.004	0.003	
				(0.007)	(0.003)	(0.011)	
ASTHMA06	70%	512	3214	0.032***	−0.008***	−0.009	
				(0.005)	(0.002)	(0.007)	
ASTHMA08	80%	436	3389	0.032***	−0.006***	−0.007	
				(0.006)	(0.002)	(0.008)	
BP5	70%	554	2122	0.021***	−0.002	−0.005	
				(0.003)	(0.002)	(0.005)	
CANCER03	90%	1098	2892	0.019***	−0.003	−0.006	
				(0.004)	(0.002)	(0.007)	
CHD08	70%	317	1255	0.028***	−0.021***	−0.012	
				(0.006)	(0.003)	(0.009)	
CHD09	90%	711	5724	0.014***	0.000	0.002	
				(0.003)	(0.001)	(0.003)	
CHD10	60%	210	1621	0.038***	−0.001	−0.012	
				(0.009)	(0.003)	(0.012)	
CHD12	90%	1412	5769	0.016***	−0.004***	0.002	
				(0.002)	(0.001)	(0.003)	
CKD02	90%	95	1212	0.044***	0.001	−0.014	
				(0.006)	(0.003)	(0.009)	
CKD03	70%	1688	3778	0.016***	0.009***	0.001	
				(0.003)	(0.002)	(0.003)	
CKD05	80%	243	1487	0.038***	−0.012***	−0.010	
				(0.010)	(0.005)	(0.016)	
CKD06	80%	1235	3387	0.027***	−0.015***	−0.018***	Comp
				(0.004)	(0.002)	(0.006)	
CVD01	70%	548	1699	0.020	−0.014**	−0.003	
				(0.013)	(0.007)	(0.017)	
CVD02	70%	347	1458	0.029**	0.001	−0.023	
				(0.014)	(0.006)	(0.018)	
DEM02	60%	79	1008	0.015	0.014**	−0.000	
				(0.025)	(0.007)	(0.035)	

Continued on next page

Table 4.5: (Continued)

		(1)	(2)	(3)	(4)	(5)	
		Descriptives		Estim. Regression Coefficients			Classif.
Indicator	UL	N Below	N Above	BELOW α_1	AFTER α_2	INTER α_3	
DM2	90%	700	4130	0.020***	0.001	−0.001	
				(0.002)	(0.001)	(0.003)	
DM10	90%	1585	5411	0.014***	0.001	−0.006*	Comp
				(0.002)	(0.001)	(0.003)	
DM13	90%	2297	5154	0.011***	−0.001	−0.005**	Comp
				(0.002)	(0.001)	(0.002)	
DM15	80%	472	1988	0.024***	0.003	−0.007	
				(0.007)	(0.003)	(0.009)	
DM18	85%	812	3437	0.014***	−0.004**	0.007	
				(0.003)	(0.001)	(0.004)	
DM21	90%	1687	5417	0.010***	−0.002**	0.004	
				(0.002)	(0.001)	(0.003)	
EPILEP06	90%	639	3745	0.031***	−0.002	−0.006	
				(0.004)	(0.001)	(0.006)	
EPILEP08	70%	1254	3262	0.014***	0.014***	−0.013*	Comp
				(0.005)	(0.003)	(0.007)	
HF02	90%	620	3028	0.025***	−0.001	0.001	
				(0.004)	(0.001)	(0.005)	
HF03	80%	271	2175	0.039***	−0.001	0.001	
				(0.010)	(0.003)	(0.012)	
HF04	60%	45	669	0.054*	0.022**	−0.050	
				(0.028)	(0.009)	(0.038)	
SMOKE04	90%	1186	6007	0.020***	0.001	−0.005	
				(0.002)	(0.001)	(0.003)	
STROKE07	90%	1751	5483	0.010***	−0.003***	0.002	
				(0.002)	(0.001)	(0.003)	
STROKE08	60%	228	705	0.024***	−0.007	0.005	
				(0.009)	(0.006)	(0.013)	
STROKE10	85%	995	4049	0.014***	−0.002	0.006	
				(0.003)	(0.002)	(0.005)	
STROKE12	90%	839	5322	0.023***	0.001	−0.005	
				(0.002)	(0.001)	(0.004)	
STROKE13	80%	290	2084	0.020*	−0.006*	−0.003	
				(0.011)	(0.003)	(0.014)	

Notes: Own calculations based on QOF data. **BELOW:** To have attained below the respective upper thershold in the first year of the variation (2009 for 2009-2010 and 2010 for 2010-2011). **AFTER:** 2010 to 2011 variation. **AFTER:** Interaction between INTER and AFTER. Clustered at PCT-level standard errors in parenthesis. Significance: * 10%, ** 5%, *** 1%.

Table 4.6: QOF indicators results: Multiple windows

Estimate of α_3 under the sample in $[UL-l, UL+k]$

Presents only indicators for which $\alpha_3 = 0$ is rejected in at least one specification.

	k=3 pp. above UL			k=5 pp. above UL		
Indicator	l=2	l=5	l=8	l=2	l=5	l=8
AF04	−0.007	−0.015***	−0.017***	−0.007	−0.015***	−0.017***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.006)
ASTHMA06	−0.007	−0.005	−0.007	−0.012	−0.009	−0.011*
	(0.008)	(0.007)	(0.006)	(0.008)	(0.007)	(0.006)
CHD08	−0.015	−0.017*	−0.017**	−0.010	−0.012	−0.013*
	(0.011)	(0.009)	(0.009)	(0.010)	(0.009)	(0.008)
CKD06	−0.008	−0.015**	−0.016***	−0.011	−0.018***	−0.019***
	(0.008)	(0.006)	(0.005)	(0.008)	(0.006)	(0.005)
CVD02	0.005	−0.020	−0.025	0.001	−0.023	−0.028*
	(0.024)	(0.020)	(0.018)	(0.023)	(0.018)	(0.016)
DM10	−0.001	−0.004	−0.002	−0.002	−0.006*	−0.003
	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
DM13	−0.007**	−0.005*	−0.005**	−0.007**	−0.005**	−0.006***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)
DM15	−0.018*	−0.005	−0.007	−0.020**	−0.007	−0.009
	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.008)
DM18	0.001	0.006	0.007	0.002	0.007	0.008**
	(0.006)	(0.005)	(0.004)	(0.006)	(0.004)	(0.004)
EPILEP08	−0.000	−0.005	−0.002	−0.008	−0.013*	−0.009
	(0.009)	(0.007)	(0.006)	(0.008)	(0.007)	(0.006)
HF04	−0.112**	−0.059	−0.055	−0.103**	−0.050	−0.045
	(0.044)	(0.039)	(0.036)	(0.043)	(0.038)	(0.035)
STROKE10	0.011*	0.003	0.005	0.013**	0.006	0.007*
	(0.006)	(0.005)	(0.004)	(0.006)	(0.005)	(0.004)

Notes: Own calculations based on QOF data. Clustered at PCT-level standard errors in parenthesis. Significance: * 10%, ** 5%, *** 1%.

4.6 Conclusion

This paper introduces a test for complementarities/substitutions in the agent's cost function in a multitasking setting when there is a two-part linear contract. It works by considering as a "control" group those agents who self-select into levels of effort that corresponds to the "kink" in the reward function, that is, at the threshold where there is a sudden change in the marginal benefit for exerting effort in a given task. For these agents, there is a wedge between the marginal benefit and marginal cost of effort, and hence, small changes in incentives will not

alter their effort allocation (and hence can be used as a control group). The test consists of two steps: first, determining whether the kink produces "bunching" in the distribution of achievement at the threshold, and if that is the case, a difference in differences estimator identifies the desired characteristic of the cost function.

As a case of study we have analysed a pay for performance scheme for family doctors in the UK, the Quality and Outcomes Framework (QOF). We have shown that changes introduced in 2010/11, which we understand as a net price drop in a set of modified indicators, revealed that several indicators are in fact complements. This might be because most clinical indicators refer to chronic patients, who not unusually have several co-morbidities.

4.A Model Examples

A simple cost function that captures both substitutability and complementarity is presented in Bolton and Dewatripont (2005): $C(e_1,e_2;\theta = \{z,c_1,c_2,\delta\}) = \frac{1}{z} \cdot (\frac{1}{2}(c_1e_1^2 + c_2e_2^2) + \delta e_1e_2)$ under the assumption that $\delta < \sqrt{c_1c_2}, c_i > 0 \forall i$. As a result we can characterize the second derivatives with each parameter $C_{ii} = \frac{1}{z} \cdot c_i$ and $C_{ij} = \frac{1}{z} \cdot \delta$, $\forall i \neq j$.

4.A.1 No uncertainty

Given our function $\phi_i(x_i)$, for an optimal level of effort below UL_1 , the optimal levels of effort are given by

$$e_1^* = z \cdot \frac{a_1c_2 - \delta a_2}{c_1c_2 - \delta^2}, e_2^* = z \cdot \frac{a_2c_1 - \delta a_1}{c_1c_2 - \delta^2}$$

Hence, Equation 4.6 becomes:

$$\frac{de_1}{da_2} = z \cdot \frac{-\delta}{c_1c_2 - \delta^2}$$

Where it is clear that the sign of δ dominates the response to the incentives: if it is negative, then the tasks are complements as the marginal cost of one of the tasks is reduced when the effort of the other is increased (similar to the concept of economies of scope). However, notice that if we are above the threshold UL_1, two options should be considered

$$e_1^* = z \cdot \frac{\tilde{a}_1 c_2 - \delta a_2}{c_1 c_2 - \delta^2}, e_2^* = z \cdot \frac{a_2 c_1}{c_1 c_2 - \delta^2} \quad \text{and} \quad e_1^* = UL_1, e_2^* = \frac{z \cdot a_2 - \delta UL_1}{c_2}$$

As a result:

1) If $\delta > 0$ (substitutes), at most, it is optimal to exert an effort level $e_1 = UL_1$, so it is expected that $\frac{de_1}{da_2}|_{e_1^* \geq UL_1} = 0$. Below that level, effort in task 1 it is decreasing with respect to the other task price: $\frac{de_1}{da_2}|_{e_1^* \geq UL_1} \leq 0$

2) If $\delta < 0$ (complements), below a cutoff $\bar{a}_2$ it is optimal to exert an effort level $e_1 = UL_1$, but above such a price cutoff, $\frac{de_1}{da_2} > 0$.

The result is a three section supply of effort 1. For substitutes it is flat, and then it decreases until it is optimal not to do any effort; and for complements it is increasing, flat and then increasing.

Kink With our current restrictions, it is straightforward to obtain the density of e_1^*. Here, $\bar{H}(\tilde{e}_1) = G\left[e_1^{*-1}(\tilde{e}_1; \bar{a}_1, a_2)\right] = G\left[\tilde{e}_1 \frac{c_1 c_2 - \delta^2}{\bar{a}_1 c_2 - \delta a_2}\right]$. Then, $\bar{h}(\tilde{e}_1) = g\left[\tilde{e}_1 \frac{c_1 c_2 - \delta^2}{\bar{a}_1 c_2 - \delta a_2}\right] \cdot \frac{c_1 c_2 - \delta^2}{\bar{a}_1 c_2 - \delta a_2}$ and similarly $\underline{h}(\tilde{e}_1) = g\left[\tilde{e}_1 \frac{c_1 c_2 - \delta^2}{(\bar{a}_1 + p_1) \cdot c_2 - \delta a_2}\right] \cdot \frac{c_1 c_2 - \delta^2}{(\bar{a}_1 + p_1) \cdot c_2 - \delta a_2}$.

Let us consider the point $\hat{e} = \tilde{e}_1 \frac{(\bar{a}_1 + p_1) \cdot c_2 - \delta a_2}{\bar{a}_1 c_2 - \delta a_2}$. If we consider the density without kink $\underline{h}(\hat{e}) = g\left[\tilde{e}_1 \frac{c_1 c_2 - \delta^2}{\bar{a}_1 c_2 - \delta a_2}\right] \cdot \frac{c_1 c_2 - \delta^2}{(\bar{a}_1 + p_1) \cdot c_2 - \delta a_2}$. We can re-express it as $g\left[\tilde{e}_1 \frac{c_1 c_2 - \delta^2}{\bar{a}_1 c_2 - \delta a_2}\right] = \underline{h}(\hat{e}) \cdot \frac{(\bar{a}_1 + p_1) \cdot c_2 - \delta a_2}{c_1 c_2 - \delta^2}$. Replacing this term in the density above UL, we can express the density of e_1^* in terms of $\underline{h}(\cdot)$, as shown below:

$$h(\tilde{e}_1) = \begin{cases} \underline{h}(\tilde{e}_1) & \text{if } \tilde{e}_1 < UL \\ b & \text{if } \tilde{e}_1 = UL \\ \underline{h}\left(\tilde{e}_1 \frac{(\bar{a}_1 + p_1) \cdot c_2 - \delta a_2}{\bar{a}_1 c_2 - \delta a_2}\right) \cdot \frac{(\bar{a}_1 + p_1) \cdot c_2 - \delta a_2}{\bar{a}_1 c_2 - \delta a_2} & \text{if } \tilde{e}_1 > UL \end{cases} \tag{4.24}$$

Notice that near UL, there is a discontinuity on the density even if we do not consider the bunching mass at UL. Below UL the density is $\underline{h}(\tilde{e}_1)$, but above it, the density is larger for a constant $\underline{h}(\tilde{e}_1)$. This is evident in the example of figure 4.3, where $\underline{h}(\cdot)$ is a constant as $g(\cdot)$ is uniformly distributed.

Comparative Statics What can generate the distribution over e_1? Let us consider only interior solutions ($a_1 c_2 - \delta a_2 > 0$ and $a_2 c_1 - \delta a_1 > 0$) and let Δ_T be the difference between the slopes of e_1^* with respect to a_2 above and below a given point T

$$\Delta_T = \frac{de_1}{da_2}|_{e_1^*<T} - \frac{de_1}{da_2}|_{e_1^*\geq T}$$

Heterogeneity on c_1 If the distribution on e_1 is due to efficiency on task 1, the sign of Δ is informative about the sign of δ. The resulting sorting on e_1 due to variation in e_1 is the same regardless of the nature of the cost function, while the size of the e_1^* slope with respect to a_2 depends on it.

	$\frac{\partial e_1}{\partial c_1}$	$\frac{\partial^2 e_1}{\partial a_2 \partial c_1}$	Below - Above (Δ_T)
	$-\left(a_1c_2 - \delta a_2\right)\cdot z^{-1}\cdot\left(c_1c_2-\delta^2\right)^{-2}c_2$	$\delta\cdot z^{-1}\cdot\left(c_1c_2-\delta^2\right)^{-2}c_2$	$\frac{-\delta}{z\cdot\left(c_1c_2-\delta^2\right)} - \frac{-\delta}{z\cdot\left(\underline{c}_1c_2-\delta^2\right)}$ First term is smaller in abs. val as its denominator is larger
$\delta<0$ (Complements)	<0	<0	<0
$\delta>0$ (Substitutes)	<0	>0	>0

Heterogeneity on c_2 If the distribution on e_1 is due to the efficiency on task 2, the sign of Δ is not informative about the sign of δ. In this case, both the sorting and the size of the e_1^* slope with respect to a_2 depend on the nature of the costs function.

	$\frac{\partial e_1}{\partial c_2}$	$\frac{\partial^2 e_1}{\partial a_2 \partial c_2}$	Below - Above (Δ_T)
	$\delta\left(a_2c_1 - a_1\delta\right)\cdot z^{-1}\cdot\left(c_1c_2-\delta^2\right)^{-2}$	$\delta\cdot z^{-1}\cdot\left(c_1c_2-\delta^2\right)^{-2}c_1$	Depends on δ
$\delta<0$ (Complements)	<0	<0	<0
$\delta>0$ (Substitutes)	>0	>0	<0 !!!!

Heterogeneity on δ If the distribution on e_1 is due to the degree of complementarity/sustituibility, the sign of Δ can only detect substitutes. Here, the size of the e_1^* slope with respect to a_2 depend on the nature of the costs function but the sorting depends on the value of other parameters. If tasks are substitutes and $e_1^* > \frac{a_2}{2\delta z^2}$, the sorting will be positive. In that case it is possible to say that the tasks are substitutes by observing a positive Δ, but if this term is positive it is not possible to deduce the sign of δ.

	$\frac{\partial e_1}{\partial \delta}$	$\frac{\partial^2 e_1}{\partial a_2 \partial \delta}$	Below - Above (Δ_T)
	$\left(-a_2\left(c_1c_2-\delta^2\right)+2\delta\left(a_1c_2-\delta a_2\right)\right)\cdot z^{-1}\cdot\left(c_1c_2-\delta^2\right)^{-2}$ $\left(-a_2+2\delta z^2\frac{a_1c_2-\delta a_2}{z\left(c_1c_2-\delta^2\right)}\right)\cdot z^{-1}\cdot\left(c_1c_2-\delta^2\right)^{-1}$ $\left(-a_2+2\delta z^2 e_1^*\right)\cdot z^{-1}\cdot\left(c_1c_2-\delta^2\right)^{-1}$	$-\left(\delta^2+c_1c_2\right)\cdot z^{-1}\cdot\left(c_1c_2-\delta^2\right)^{-2}c_1$	Depends on δ
$\delta<0$ (Complements)	<0	<0	<0
$\delta>0$ (Substitutes)	If $-a_2\left(c_1c_2-\delta^2\right)+2\delta\left(a_1c_2-\delta a_2\right)>0$, then >0	<0	>0
	If $-a_2\left(c_1c_2-\delta^2\right)+2\delta\left(a_1c_2-\delta a_2\right)<0$, then <0	<0	<0 !!!!

4.A.2 With uncertainty

Adding the functional form $C = \frac{1}{z} \cdot (\frac{1}{2}(c_1 e_1^2 + c_2 e_2^2) + \delta e_1 e_2)$. Also, assume $\varepsilon_1 \sim N(0,\sigma_1)$, which will allow us to work with the standard normal distribution. An additional element is the inclusion of the penalty Ω for uncertainty. For instance, this term will be equal to $\frac{1}{2}\eta(a_1^2\sigma_1^2)$ if we consider an exponential utility function $u(p) = -exp(-\eta \cdot p)$, where $\eta$is the absolute risk aversion coefficient.

$$FOC_1 := z \cdot \left\{ a_1 \Phi\left(\frac{UL - e_1}{\sigma_1}\right) + \frac{1}{\sigma_1}\phi\left(\frac{UL - e_1}{\sigma_1}\right) \cdot [(UL - e_1) \cdot a_1 + \Omega] \right\} - c_1 e_1 - \delta e_2 = 0$$

$$FOC_2 := z \cdot a_2 - c_2 e_2 - \delta e_1 = 0$$

And the equivalent of Equation 4.12

$$\frac{de_1}{da_2} = z \cdot \frac{-\delta}{c_1 c_2 - \delta^2 + z \cdot c_2 \left\{ \frac{1}{\sigma^2}\phi'\left(\frac{UL-e_1}{\sigma_1}\right) \cdot [(UL - e_1) \cdot a_1 + \Omega] + 2a_1 \frac{1}{\sigma}\phi\left(\frac{UL-e_1}{\sigma_1}\right) \right\}}$$

Given that for the standard normal pdf it holds that $\phi'(x) = -x\phi(x)$

$$\begin{aligned}\frac{de_1}{da_2} &= z \cdot \frac{-\delta}{c_1 c_2 - \delta^2 + z \cdot c_2 \left\{ -\frac{1}{\sigma^3}\phi\left(\frac{UL-e_1}{\sigma_1}\right) \cdot [UL - e_1] \cdot [(UL - e_1) \cdot a_1 + \Omega] + 2a_1 \frac{1}{\sigma}\phi\left(\frac{UL-e_1}{\sigma_1}\right) \right\}} \\ &= z \cdot \frac{-\delta}{c_1 c_2 - \delta^2 + a_1 \cdot z \cdot c_2 \cdot \frac{1}{\sigma} \cdot \phi\left(\frac{UL-e_1}{\sigma_1}\right) \left\{ 2 - \frac{1}{\sigma^2} \cdot [UL - e_1] \cdot \left[\frac{1}{a_1}\Omega + UL - e_1 \right] \right\}}\end{aligned}$$

As in the general case, being far from UL implies a larger slope (in absolute value). This is an effect that is attenuated by risk aversion below UL. Above such cut-off, risk aversion makes the derivative larger in absolute value. In this particular case, being very far from UL implies that the derivative will be equivalent to the non-uncertainty case.

4.B QOF Payment

Equation 4.26 shows how ratio indicators are translated into income for a practice i. Essentially, achievement x_i of indicator j is translated into points, and such points into yearly income. First, points are allocated according to a non-linear tariff that depends on two indicator specific thresholds. Below the lower limit (LL_j) zero points are awarded, and above the upper limit (UL_j) the maximum amount of available points for indicator j is awarded (Equation 4.25). The resulting figure is adjusted with respect to the relative size of the practice (*contractor population index*, CPI_i), and to the relative prevalence of the specific condition rewarded for clinical indicators (PF_{ij}). The achievement factor is multiplied by the CPI index and the prevalence factors, and by the price per point (Equation 4.26). The CPI captures the size of the practice, and is calculated as the number of patients in the practice relative to the figure 5891, which was the 2003 average list size.[32] The prevalence factor measures how commonly the condition is treated in indicator j, relative to the national average.

$$
\begin{aligned}
x_{ij} &= \frac{\text{Numerator}_{ij}}{\text{Denominator}_{ij}} \\
AF_{ij} &= \begin{cases} 0 & \text{if } x_i \leq LL_j \\ \left(x_i - LL_j\right) \cdot \frac{\text{Avail. Points}_j}{UL_j - LL_j} & \text{if } x_i > LL_j \\ \text{Avail. Points}_j & \text{if } x_i \geq UL_j \end{cases} \qquad (4.25) \\
CPI_i &= \frac{list_i}{5891} \\
PF_{ij} &= \frac{denom_{ij}/list_i | X_{ij}}{E[denom/list|X]} \text{ , where } X \text{ are specific conditions} \\
P_{ij} &= \Big(\text{Value per point in Â£}\Big) \cdot AF_{ij} \cdot CPI_i \cdot PF_{ij} \qquad (4.26)
\end{aligned}
$$

[32] Since 2013 this figure has been updated annually. More details are available from BMA (2013).

4.C Additional Tables

Table 4.7: QOF indicators descriptives for 2010/11

Indicator	UL	(1) Number	(2) $E[x_t]$	(3) $\mathrm{P}[x_t < UL]$	(4) $\rho(x_t)$	(5) $E[x_t - x_{t-1} \mid x_{t-1} < UL]$	(6) $E[x_t - x_{t-1} \mid x_{t-1} > UL]$
AF03	90%	8245	93.82	7.14	0.50	9.88	-0.55
AF04	90%	8245	95.28	6.43	0.51	26.20	-1.39
ASTHMA03	80%	8245	90.00	4.69	0.41	18.64	-0.76
ASTHMA06	70%	8245	79.58	5.29	0.54	11.03	-0.19
ASTHMA08	80%	8245	87.89	6.37	0.46	15.63	-0.98
BP5	70%	8245	79.68	5.17	0.64	5.87	-0.09
CANCER03	90%	8245	92.75	17.84	0.34	18.18	-2.69
CHD08	70%	8245	81.90	3.51	0.56	11.30	-0.41
CHD09	90%	8245	93.58	7.56	0.46	4.30	-0.66
CHD10	60%	8245	74.91	2.60	0.67	13.30	-0.70
CHD12	90%	8245	92.73	16.53	0.48	5.17	-0.31
CKD02	90%	8245	97.26	1.29	0.41	26.83	-0.37
CKD03	70%	8245	74.86	21.73	0.52	5.58	-1.72
CKD05	80%	8245	90.78	6.03	0.46	40.20	-2.70
CKD06	80%	8245	82.35	24.29	0.53	14.80	-1.33
CVD01	70%	8245	80.12	14.71	0.44	26.06	-5.50
CVD02	70%	8245	82.61	7.94	0.37	34.13	-5.68
DEM02	60%	8245	80.54	3.04	0.42	36.15	-0.92
DM2	90%	8245	94.87	7.00	0.54	4.47	-0.36
DM10	90%	8245	91.39	22.84	0.58	4.93	-0.69
DM13	90%	8245	88.80	37.48	0.65	3.38	-1.38
DM15	80%	8245	89.28	8.07	0.53	20.31	-1.66
DM17	70%	8245	82.73	2.43	0.60	8.70	-0.55
DM18	85%	8245	91.19	9.76	0.47	5.99	-0.17
DM21	90%	8245	91.08	24.33	0.52	5.46	-0.97
DM22	90%	8245	96.95	2.44	0.44	7.78	-0.02
EPILEP06	90%	8245	95.62	6.95	0.27	13.07	-0.64
EPILEP08	70%	8245	73.96	26.14	0.56	7.72	-3.09
HF02	90%	8245	95.46	8.02	0.51	17.25	-1.03
HF03	80%	8245	90.26	4.24	0.46	27.42	-1.10
HF04	60%	8245	83.15	3.26	0.47	41.65	-1.39
SMOKE03	90%	8245	95.61	2.66	0.52	5.69	0.00
SMOKE04	90%	8245	93.07	12.48	0.44	4.98	-0.72
STROKE07	90%	8245	91.49	23.91	0.43	5.01	-1.08
STROKE08	60%	8245	77.18	3.07	0.50	20.72	-0.57
STROKE10	85%	8245	90.09	13.45	0.40	8.97	-0.49

Continued on next page

Table 4.7: (Continued)

(7) Indicator	(8) UL	(1) (9) Number	(2) (10) $E[x_t]$	(3) $P[x_t < UL]$	(4) $\rho(x_t)$	(5) $E[x_t - x_{t-1} \mid x_{t-1} < UL]$	(6) $E[x_t - x_{t-1} \mid x_{t-1} > UL]$
STROKE12	90%	8245	93.79	8.98	0.45	9.97	-0.93
STROKE13	80%	8245	88.90	7.51	0.58	25.92	-1.87
THYROI02	90%	8245	95.81	3.24	0.41	10.46	-0.11

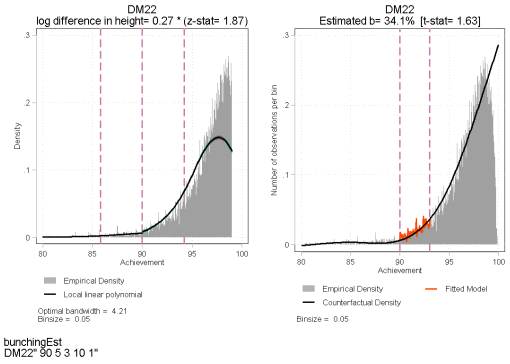
Notes: Own calculations based on QOF data. **Number:** Number of GP practices, including those with 0 elegible patients for the given indicator. $E[x_t]$ **:** Average achievement per indicator. $\mathbf{P}[x_t < UL]$ **:** Proportion of practices with an achivement below UL. $\rho(x_t)$ **:** Correlation between 2010 and 2009 achivement.

Table 4.8: Detailed Changes in QOF 2011 clinical indicators with respect to 2009-2010

Status	Description	Affected Indicators	Price Interpretation	Points
Retirements	These tasks are not rewarded anymore. Clinical indicators are about having a recent record of certain physical measures, or reviews.	CHD5, CHD7, DM5, DM11, DM16, EPILEPSY7, MH7, STROKE5	Reduction	32
Points reduced	Number of assigned points per indicator was reduced.†	BP4, DEP1	Reduction	26 to 22
Upper Limit Increased	Small increase from 70% to 71%. ♠	CHD6, STROKE6	Reduction	22
Replacement I	For indicators PP01, MH04, MH05, the time for accomplishing a given goal was reduced. For CHD2, the optional specialist referral was made compulsory.	PP01, MH04, MH05, CHD2	Reduction	18
Replacement II	Decrease in points and new wording is more precise and requires actions at the moment of diagnosis instead of treatment starting point.	DEP2, DEP3	Reduction	45 to 25
Replacement III	Most of these indicators were replaced by versions which are harder to accomplish. In a few of them this was compensated with extra points, but in some others there was a reduction as well: • For CHD11/CHD14 there is an increase from 7 to 10 points in exchange for prescribing aspirin and statins on top of an ACE inhibitor or alternative blood pressure treatments. • Requirements for DM9 were increased from checking peripheral pulses to a more comprehensive foot examination. It was also increased from 3 to 4 points. • Indicator DM12 was split into DM30 and DM31, keeping the same number of points. It asked for a percentage of patients below a given blood pressure target (145/85). It was replaced by two targets, one slightly below the original (140/80), and one notoriously above (150/90). • Indicator MH09 was split into MH11, MH12, MH13, MH14, MH15 and MH16. It moved from 23 to 27 points. The original indicator was general and imprecise ("routine health promotion and prevention advice appropriate to their age and health status"), while the replacements ask for specific measurements depending on age and gender.	CHD11/CHD14, DM9, DM12 (DM30,DM31), MH09 (MH11, MH12, MH13, MH14, MH15 and MH16)	Ambiguous	51 to 59
Replacement IV	The cutoff was relaxed from last HbA1C to be 7% or less, to HbA1C to be 7.5% or less	DM23/DM26	Increase	17
Replacement V	Similar or the same wording, but the recoding was done in order to highlight recent changes in diagnostic procedures. For diabetes indicators the wording is explicit about new measurement standards.	COPD1/COPD14, COPD12/COPD15, MH6/MH10, DM24/DM27, DM25/DM28	-	32
New	These are tasks that were not considered before. Three new clinical indicators, on dementia, epilepsy and learning disabilities.	DEM3, EPILEPSY 9, LD2	Increase	12
Unchanged	No change on points, thresholds or wording		-	454

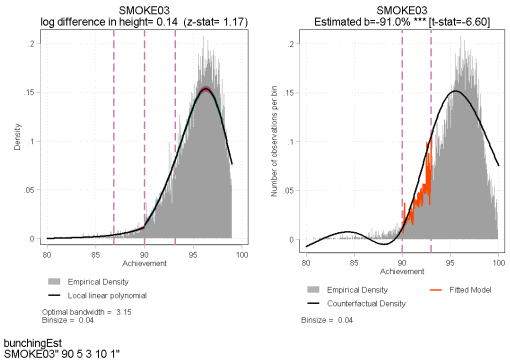
Note: This corresponds to our interpretation based on NHS Employers public documents. † Does not include indicators which wording was amended as DEP2 and DEP3. ♠ Does not include DM12/DM30, which is an indicator that its wording was also amended.

Table 4.9: Bunching tests for selected indicators I



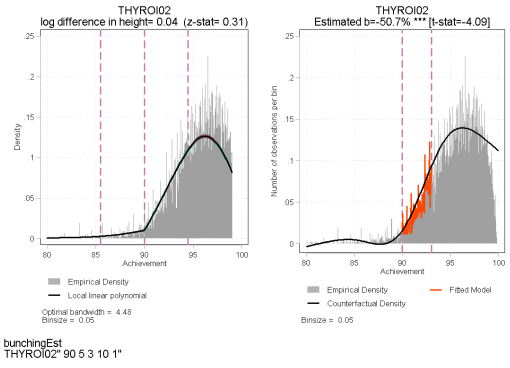
bunchingEst
DM22" 90 5 3 10 1"

DM22: The percentage of patients with diabetes who have a record of estimated glomerular filtration rate (eGFR) or serum creatinine testing in the previous 15 months.
3 points. LL=40, UL=90.



bunchingEst
SMOKE03" 90 5 3 10 1"

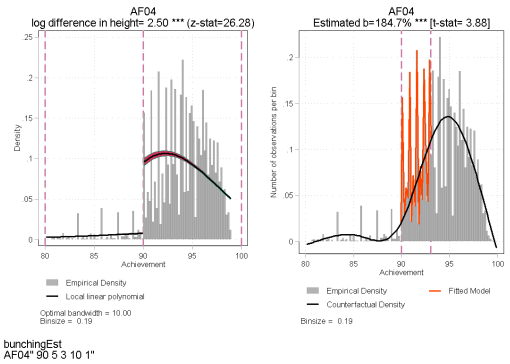
SMOKE3: The percentage of patients with any or any combination of the following conditions: coronary heart disease, stroke or TIA, hypertension, diabetes, COPD, CKD, asthma, schizophrenia, bipolar affective disorder or other psychoses whose notes record smoking status in the previous 15 months (except those who have never smoked where smoking status need only be recorded once since diagnosis)
30 points. LL=40, UL=90.



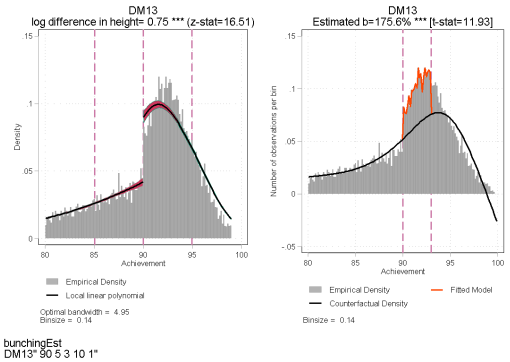
bunchingEst
THYROI02" 90 5 3 10 1"

THYROID2: The percentage of patients with hypothyroidism with thyroid function tests recorded in the previous 15 months
6 points. LL=40, UL=90.

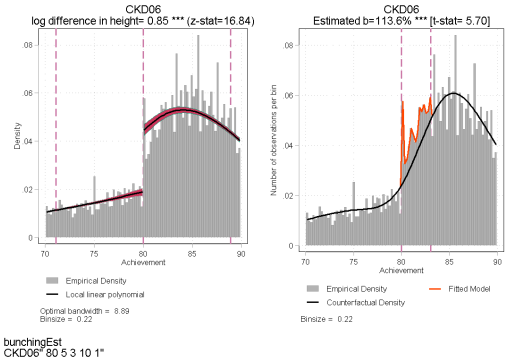
Table 4.10: Bunching tests for selected indicators II



AF04: The percentage of patients with atrial fibrillation diagnosed after 1st April 2008 with ECG or specialist confirmed diagnosis.
10 points. LL=40, UL=90.



DM13: The percentage of patients with diabetes who have a record of micro-albuminuria testing in the previous 15 months (exception reporting for patients with proteinuria)
3 points. LL=40, UL=90.



CKD06: The percentage of patients on the CKD register whose notes have a record of an albumin:creatinine ratio (or protein:creatinine ratio) value in the previous 15 months
6 points. LL=40, UL=80.

Chapter 5

Conclusions, future work and policy recommendations

This thesis has explored three broad topics in the area of the economics of health. While each of them deals with different research questions, the common element among them is the role of government policies in shaping health and, as a consequence, individual choices. As a result, the conclusions of each chapter are able enrich our understanding of how to design health-related policies in order to improve the welfare of a society.

In chapter 2, behavioural responses to routine health checks were considered. Specifically, as a result of being advised to visit a family doctor due to a potential risk of suffering hypertension, I found an increase in the probability of being on medication to lower blood pressure but also differences in lifestyle. Interestingly, there were responses in opposite directions. For instance there was a reduction in alcohol intake frequency but also an increase in the odds of being obese. This adds to current evidence that suggests that information-based treatments should consider behavioural responses in order to assess the full impact of these programs, and that more research is needed to understand which lifestyles are more responsive and in which direction to such interventions.

As a policy lesson, this chapter tells us that health checks type of programmes might work in identifying certain individuals who are at risk of developing chronic diseases. The effect on medication use was substantial, especially for those individuals with a high risk of developing cardiovascular diseases, even if it was temporal. However, it is important that this type of policy is also accompanied by a

lifestyle follow-up in order to check whether this is deteriorating.

Chapter 3 presented a novel framework for assessing the value and the cost-effectiveness of the adoption of a health intervention. It combines a health progression model with a standard life-cycle model with endogenous labour supply. As a result, it is possible to construct counterfactual scenarios with diverse policies, featuring not only health technology or social protection but also labour market characteristics. This allows for determining a willingness-to-pay value that incorporates how households' consumption and leisure is affected by the treatment. With this too is also possible to improve standard cost-effectiveness calculations by incorporating the gains on additional labour income, a potential source of resources for the health system.

This chapter developed a tool that would be useful for institutions around the world that have to decide whether to include certain treatments in a health insurance contract. The general design of the model allows potential users to analyse preventive care innovations in areas different to cardiovascular disease. For instance, valuing innovations in mental health, cancer and disability benefits among others are feasible applications. The sole requirements are to be able to map such conditions into reported diagnoses, and to have information on markers that signal a higher risk of developing certain conditions. This tool can also introduce a more general analysis of the heterogeneity in the value of a treatment, which might also help to target some treatments to specific populations.

And last, in the context of contracting schemes where there are rewards for specific tasks, in chapter 4 a novel test for complementarity/substitutability across tasks was introduced. It takes advantage of two-piece linear tariffs such as the *Quality and Outcomes Framework* clinical indicators. Essentially, for a given task, agents - GP practices in our example - who self-select into the level of effort at which the marginal benefits change become insensitive to small exogenous variations in the marginal benefit of alternative tasks. Effectively, those agents near the threshold become a control group that allow us to understand how agents react to such alternative task prices variations.

While this last chapter's contribution is more technical than the previous ones, it has direct implications for our understanding of financial incentive schemes. In

this precise case, it shows that the UK primary care incentive programme does not have tasks that are clear substitutes. It also suggests that these schemes should include kink points in order to allow for a recurrent evaluation of changes in their design. This is a sensible alternative when it is not possible to introduce experimental variation.

As a general conclusion, this dissertation has shown that there is potential to extend our knowledge of preventive care interventions both from the supply and demand side. First, there is empirical evidence that certain interventions do affect individuals' decisions; however work on this topic is scarce and more detailed evidence of the mechanisms behind is still required. In my research I found that some lifestyles are improved while others deteriorate, but knowing more about the heterogeneity in beliefs regarding the contribution of each input of the health production function is still required. Understanding this would motivate a better analysis on how individuals value medical innovations. In my research, rational agents know perfectly the benefit of a drug and have beliefs about their future health that match realised events for the previous generation. This is an assumption that could be relaxed if my proposed framework were to be combined with information that includes subjective beliefs about current and future health.

The previous recommendations, both in terms of policy and research, are direct contributions from this dissertation. I hope that they could be implemented in order to improve both welfare and knowledge.

Bibliography

Al-Ubaydli, O., S. Andersen, U. Gneezy, and J. A. List (2012). Carrots that look like sticks: toward an understanding of multitasking incentive schemes. *NBER Working Paper 18453*.

Anderson, K. M., P. WOLSON, P. M. Odell, and W. B. Kannel (1991). An updated coronary risk profile: a statement for health professionals. *Circulation 83*(1), 356–362.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association 103*(484).

Arcidiacono, P. and P. B. Ellickson (2011). Practical methods for estimation of dynamic discrete choice models. *Annual Review of Economics 3*(1), 363–394.

Arcidiacono, P., H. Sieg, and F. Sloan (2007). Living rationally under the volcano? an empirical analysis of heavy drinking and smoking. *International Economic Review 48*(1), 37–65.

Artac, M., A. R. Dalton, A. Majeed, J. Car, and C. Millett (2013). Effectiveness of a national cardiovascular disease risk assessment program (NHS health check): Results after one year. *Preventive Medicine 57*(2), 129–134.

Ashworth, M., J. Medina, and M. Morgan (2008). Effect of social deprivation on blood pressure monitoring and control in england: a survey of data from the quality and outcomes framework. *Bmj 337*, a2030.

Ashworth, M., P. Schofield, P. Seed, S. Durbaba, M. Kordowicz, and R. Jones (2011). Identifying poorly performing general practices in england: a longi-

tudinal study using data from the quality and outcomes framework. *Journal of Health Services Research & Policy 16*(1), 21–27.

Baker, G. P. (1992). Incentive contracts and performance measurement. *Journal of political Economy*, 598–614.

Banks, J., R. Blundell, Z. Oldfield, and J. P. Smith (2012). Housing mobility and downsizing at older ages in britain and the usa. *Economica 79*(313), 1–26.

Becker, G. S. (2007). Health as human capital: synthesis and extensions. *Oxford Economic Papers 59*(3), 379–410.

Blau, D. M. and D. B. Gilleskie (2008). The role of retiree health insurance in the employment behavior of older men. *International Economic Review 49*(2), 475–514.

Blundell, R., R. Crawford, E. French, and G. Tetlow (2016). Comparing retirement wealth trajectories on both sides of the pond. *Fiscal Studies 37*(1), 105–130.

BMA (2013). Focus on qof payments. `https://www.bma.org.uk/-/media/files/pdfs/practical%20advice%20at%20work/contracts/independent%20contractors/qof%20guidance/focusonqofpaymentsnov2013.pdf`. Accesed: 2016-08-09.

Bolton, P. and M. Dewatripont (2005). *Contract theory*. MIT press.

Borghi, J. (2008). Aggregation rules for cost–benefit analysis: a health economics perspective. *Health economics 17*(7), 863–875.

Bozio, A., R. Crawford, and G. Tetlow (2010). *The history of state pensions in the UK: 1948 to 2010*. Institute for Fiscal Studies.

Bradler, C., R. Dur, S. Neckermann, and A. Non (2013). Employee recognition and performance: A field experiment. *CESifo Working Paper 4164*.

Bridges, M. B. S., D. Hussey, and D. Mandalia (2015). The dynamics of ageing: The 2012 english longitudinal study of ageing (wave 6).

Brouwer, W. B. and M. A. Koopmanschap (2000). On the economic foundations of cea. ladies and gentlemen, take your positions! *Journal of health economics 19*(4), 439–459.

Cassell, M. M., D. T. Halperin, J. D. Shelton, and D. Stanton (2006). Risk compensation: the achilles' heel of innovations in hiv prevention. *Bmj 332*(7541), 605–607.

Chamberlain, G. (1992). Comment: Sequential moment restrictions in panel data. *Journal of Business & Economic Statistics 10*(1), 20–26.

Chernew, M., D. M. Cutler, K. Ghosh, and M. B. Landrum (2016). Understanding the improvement in disability free life expectancy in the us elderly population. Technical report, National Bureau of Economic Research.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2009). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. Technical report, National Bureau of Economic Research.

Chobanian, A. V., G. L. Bakris, H. R. Black, W. C. Cushman, L. A. Green, J. L. Izzo Jr, D. W. Jones, B. J. Materson, S. Oparil, J. T. Wright Jr, et al. (2003). The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure: the jnc 7 report. *Jama 289*(19), 2560–2571.

Cochrane, T., R. Davey, Z. Iqbal, C. Gidlow, J. Kumar, R. Chambers, and Y. Mawby (2012). NHS health checks through general practice: randomised trial of population cardiovascular risk reduction. *BMC Public Health 12*(1), 944.

CQC, C. Q. C. (2009). Closing the gap. tackling cardiovascular disease and health inequalities by prescribing statins and stop smoking services. *CQC: London*.

Culyer, A. J. (1989). The normative economics of health care finance and provision. *Oxford review of economic policy 5*(1), 34–58.

Cutler, D. M., S. Kadiyala, K. Murphy, and R. Topel (2003). The return to biomedical research: Treatment and behavioral effects. *Measuring the gains from medical research: An economic approach*, 110–62.

Czubek, M., S. Johal, G. Britain, and H. Revenue (2010). *Econometric analysis of cigarette consumption in the UK*. HM Revenue & Customs.

D'Agostino, R. B., R. S. Vasan, M. J. Pencina, P. A. Wolf, M. Cobain, J. M. Massaro, and W. B. Kannel (2008). General cardiovascular risk profile for use in primary care the framingham heart study. *Circulation 117*(6), 743–753.

De Nardi, M. (2004). Wealth inequality and intergenerational links. *The Review of Economic Studies 71*(3), 743–768.

De Nardi, M., E. French, and J. B. Jones (2010). Why do the elderly save? the role of medical expenses. *Journal of political economy 118*(1), 39–75.

Devlin, N. J. and P. F. Krabbe (2013). The development of new research methods for the valuation of eq-5d-5l. *The European Journal of Health Economics 14*(1), 1–3.

DfT, T. (2005). Road casualties great britain: 2004 annual report. *London: UK Department for Transport*.

Dixon, A., A. Khachatryan, A. Wallace, S. Peckham, T. Boyce, and S. Gillam (2010). The quality and outcomes framework (qof): does it reduce health inequalities. *Final report. NIHR SDO*.

Doran, T., S. Campbell, C. Fullwood, E. Kontopantelis, and M. Roland (2010). Performance of small general practices under the uk's quality and outcomes framework. *Br J Gen Pract 60*(578), e335–e344.

Doran, T., E. Kontopantelis, D. Reeves, M. Sutton, and A. M. Ryan (2014). Setting performance targets in pay for performance programmes: what can we learn from qof? *BMJ 348*.

Doran, T., E. Kontopantelis, J. M. Valderas, S. Campbell, M. Roland, C. Salisbury, and D. Reeves (2011). Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the uk quality and outcomes framework. *Bmj 342*, d3590.

Dow, W. H., T. J. Philipson, and X. Sala-i Martin (1999). Longevity complementarities under competing risks. *American Economic Review*, 1358–1371.

Duffie, D. and K. J. Singleton (1993). Simulated moments estimation of markov models of asset prices. *Econometrica 61*(4), 929–952.

Dumont, E., B. Fortin, N. Jacquemet, and B. Shearer (2008). Physicians' multitasking and incentives: Empirical evidence from a natural experiment. *Journal of Health Economics 27*(6), 1436 – 1450.

Ebrahim, S., F. C. Taylor, and P. Brindle (2014). Statins for the primary prevention of cardiovascular disease. *BMJ 348*, g280.

Eggleston, K. (2005). Multitasking and mixed systems for provider payment. *Journal of Health Economics 24*(1), 211–223.

Falaschetti, E., J. Mindell, C. Knott, and N. Poulter (2014). Hypertension management in england: a serial cross-sectional study from 1994 to 2011. *The Lancet 383*(9932), 1912–1919.

Feng, Y., A. Ma, S. Farrar, and M. Sutton (2015). The tougher the better: an economic analysis of increased payment thresholds on the performance of general practices. *Health economics 24*(3), 353–371.

Feng Lu, S. (2012). Multitasking, information disclosure, and product quality: Evidence from nursing homes. *Journal of Economics & Management Strategy 21*(3), 673–705.

Fichera, E., J. Banks, and M. Sutton (2014). Health behaviours and the patient-doctor interaction: The double moral hazard problem. Technical report, Economics, The University of Manchester.

Fichera, E., E. Gray, and M. Sutton (2016). How do individuals' health behaviours respond to an increase in the supply of health care? evidence from a natural experiment. *Social Science & Medicine 159*, 170–179.

Fichera, E. and M. Sutton (2011). State and self investments in health. *Journal of health economics 30*(6), 1164–1173.

French, E. (2005). The effects of health, wealth, and wages on labour supply and retirement behaviour. *The Review of Economic Studies 72*(2), 395–427.

French, E. and J. B. Jones (2011). The effects of health insurance and self-insurance on retirement behavior. *Econometrica 79*(3), 693–732.

Gillam, S. and N. Steel (2013). The quality and outcomes framework—where next? *BMJ 346*.

Glewwe, P., N. Illias, and M. Kremer (2010). Teacher incentives. *American Economic Journal: Applied Economics 2*(3), 205–227.

Goodwin, N., A. Dixon, T. Poole, and V. Raleigh (2011). Improving the quality of care in general practice. *London: The King's Fund*.

Gravelle, H., M. Sutton, and A. Ma (2010). Doctor behaviour under a pay for performance contract: Treating, cheating and case finding?*. *The Economic Journal 120*(542), F129–F156.

Gregory, S. (2009). *General practice in England: An overview*. King's Fund.

Grossman, M. (1972). On the concept of health capital and the demand for health. *The Journal of Political Economy 80*(2), 223–255.

Group, H. P. S. C. et al. (2002). Heart protection study collaborative group mrc/bhf heart protection study of cholesterol lowering with simvastatin in 20536 high-risk individuals: a randomised placebo-controlled trial. *Lancet 360*, 7–22.

Hai, R. and J. J. Heckman (2015). A dynamic model of health, education and wealth with credit constraints and rational addiction.

Hall, R. E. and C. I. Jones (2007). The value of life and the rise in health spending. *The Quarterly Journal of Economics*, 39–72.

Halliday, T. J., H. He, L. Ning, and H. Zhang (2015). Health investment over the life-cycle.

Harrell, F. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.

Heidi Guyer, Mary Beth Ofstedal, C. L. K. C. (2010). The feasibility of collecting physical measures and biomarkers in cross-national studies. In *Population Association of America 2010 Annual Meeting. Session 129: Demographic Studies Based on Biomarkers*.

Hippisley-Cox, J., C. Coupland, Y. Vinogradova, J. Robson, and P. Brindle (2008). Performance of the qrisk cardiovascular risk prediction algorithm in an independent uk sample of patients from general practice: a validation study. *Heart 94*(1), 34–39.

Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization 7*, 24–52.

HSCIC (2013). Prescriptions dispensed in the community: England - 2002-2012. Technical report, Prescribing and Primary Care Services, Health and Social Care Information Centre.

Imbens, G. and K. Kalyanaraman (2011). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies*, rdr043.

Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics 142*(2), 615–635.

ISER (2010, July). British household panel survey: Waves 1-18, 1991-2009. 7th edition. computer file 5151, UK Data Archive [distributor], Institute for Social and Economic Research, University of Essex, Colchester, Essex.

Kaarboe, O. and L. Siciliani (2011). Multi-tasking, quality and pay for performance. *Health Economics 20*(2), 225–238.

Kaestner, R., M. Darden, and D. Lakdawalla (2014). Are investments in disease prevention complements? the case of statins and health behaviors. *Journal of Health Economics 36*, 151 – 163.

Kahn, M. E. (1999). Diabetic risk taking: The role of information, education and medication. *Journal of Risk and Uncertainty 18*(2), 147–164.

Kelly, E. and G. Stoye (2014). Does gp practice size matter? gp practice size and the quality of primary care. IFS Report R101, Institute for Fiscal Studies.

Khwaja, A., F. Sloan, and S. Chung (2007). The relationship between individual expectations and behaviors: Mortality expectations and smoking decisions. *Journal of Risk and Uncertainty 35*(2), 179–201.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics 8*(1).

Kosfeld, M. and S. Neckermann (2011). Getting more work for nothing? symbolic awards and worker performance. *American Economic Journal: Micoeconomics 3*, 86–99.

Krogsbøll, L. T., K. J. Jørgensen, C. Grønhøj Larsen, and P. C. Gøtzsche (2012). General health checks in adults for reducing morbidity and mortality from disease: Cochrane systematic review and meta-analysis. *BMJ: British Medical Journal 345*.

Lader, D., S. Short, and J. Gershuny (2006). The time use survey, 2005. how we spend our time.

Lazear (2000). Performance pay and productivity. *American Economic Review 90*(5), 1346–61.

Leicester, A. and P. Levell (2012). Anti-smoking policies and smoker well-being: Evidence from the UK. *IFS Working Paper 13*.

Liu, J., N. Maniadakis, A. Gray, and M. Rayner (2002). The economic burden of coronary heart disease in the uk. *Heart 88*(6), 597–603.

MacAuley, D. (2012). The value of conducting periodic health checks. *BMJ 345*.

Malek, M. H. (2001). Implementing qalys. *What is series 2*(1).

Marmot, M., Z. Oldfield, S. Clemens, M. Blake, A. Phelps, J. Nazroo, A. Steptoe, N. Rogers, and J. Banks (2013, October). English longitudinal study of ageing: Waves 0-5, 1998-2011 20th edition.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698 – 714. The regression discontinuity design: Theory and applications.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica: Journal of the Econometric Society*, 995–1026.

McIntosh, E. (2006). Using discrete choice experiments within a cost-benefit analysis framework. *PharmacoEconomics 24*(9), 855–868.

Muralidharan, K. and V. Sundararaman (2011). Teacher performance pay: Experimental evidence from india. *Journal of Political Economy 119*, 39–77.

Murphy, K. M. and R. H. Topel (2003). The economic value of medical research. In K. M. Murphy and R. H. Topel (Eds.), *Measuring the gains from medical research: An economic approach*. Chicago: University of Chicago Press.

Murphy, K. M. and R. H. Topel (2006). The value of health and longevity. *Journal of Political Economy 114*(5).

Murphy, K. M. and R. H. Topel (2010). *Measuring the gains from medical research: an economic approach*. University of Chicago Press.

NatCen and UCL (2010, April). Health survey for England 5th edition.

Neal, D. (2011). The design of performance pay in education. In E. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of Economics of Education*, Volume 4, pp. 495–550. Oxford: Elsevier.

Ng, C. W. L. and K. P. Ng (2013). Does practice size matter? review of effects on quality of care in primary care. *Br J Gen Pract 63*(614), e604–e610.

NHS Employers (2010, May). Changes to qof 2010/11. URL: http://www.nhsemployers.org/your-workforce/primary-care-contacts/general-medical-services/quality-and-outcomes-framework/changes-to-qof-201011. Accessed: 2016-08-09.

NICE (2006, June). Cg 34: Hypertension. clinical management of primary hypertension in adults.

NICE (2006, August). TA94: statins for the prevention of cardiovascular events. NICE technology appraisal guidance.

NICE (2008). Lipid modification: Cardiovascular risk assessment and the modification of blood lipids for the primary and secondary prevention of cardiovascular disease. Clinical Guideline 67.

NICE (2011, August). Cg 127: Hypertension. clinical management of primary hypertension in adults.

NICE (2014). Cardiovascular disease: risk assessment and reduction, including lipid modification. Clinical Guideline 181.

Nichols, A. (2012). rd: Stata module for regression discontinuity estimation. *Statistical Software Components*.

Nordhaus, W. D. (2003). The health of nations: the contribution of improved health to living standards. In K. M. Murphy and R. H. Topel (Eds.), *Measuring the gains from medical research: An economic approach*. Chicago: University of Chicago Press.

Ozkan, S. (2014). Preventive vs. curative medicine: A macroeconomic analysis of health care over the life cycle. *Unpublished. https://sites.google.com/site/serdarozkan/Ozkan_2014.pdf*.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica: Journal of the Econometric Society*, 1027–1057.

Palumbo, M. G. (1999). Uncertain medical expenses and precautionary saving near the end of the life cycle. *The Review of Economic Studies 66*(2), 395–421.

Papageorge, N. W. (2015). Why medical innovation is valuable: Health, human capital and the labor market. *Johns Hopkins University, Mimeo*.

Phillips, C. and G. Thompson (2001). What is a qaly? *What is series*.

Picard, R. et al. (2012). Geodist: Stata module to compute geodetic distances. *Statistical Software Components*.

Pickering, T. G. (1996). White coat hypertension. *Current opinion in nephrology and hypertension 5*(2), 192–198.

Pischke, J.-S. (1995). Measurement error and earnings dynamics: Some estimates from the psid validation study. *Journal of Business & Economic Statistics 13*(3), 305–314.

Robson, J., I. Dostal, A. Sheikh, S. Eldridge, V. Madurasinghe, C. Griffiths, C. Coupland, and J. Hippisley-Cox (2016). The nhs health check in england: an evaluation of the first 4 years. *BMJ open 6*(1), e008840.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy 2*(3), 180–212.

Sandy, R., R. F. Elliott, W. S. Siebert, and X. Wei (2001). Measurement error and the effects of unions on the compensating differentials for fatal workplace risks. *Journal of Risk and Uncertainty 23*(1), 33–56.

Scholes, S., R. Taylor, H. Cheshire, K. Cox, and C. Lessof (2008). Retirement, health and relationships of the older population in england: the 2004 english longitudinal study of ageing. *London, United Kingdom: National Centre for Social Research*.

Shearer, B. (2004). Piece rates, fixed wages and incentives: Evidence from a field experiment. *Review of Economic Studies 71*(2), 513–34.

Shiroiwa, T., Y.-K. Sung, T. Fukuda, H.-C. Lang, S.-C. Bae, and K. Tsutani (2010). International survey on willingness-to-pay (wtp) for one additional qaly gained: what is the threshold of cost effectiveness? *Health economics 19*(4), 422–437.

Smith, J., H. Holder, N. Edwards, J. Maybin, H. Parker, R. Rosen, and N. Walsh (2013). Securing the future of general practice: new models of primary care. *London: Nuffield Trust*.

Smith, V. K., D. H. Taylor Jr, F. A. Sloan, F. R. Johnson, and W. H. Desvousges (2001). Do smokers respond to health shocks? *Review of Economics and Statistics 83*(4), 675–687.

Steptoe, A. and A. McMunn (2009). Health behaviour patterns in relation to hypertension: the english longitudinal study of ageing. *Journal of hypertension 27*(2), 224–230.

Strazzullo, P., S. M. Kerry, A. Barbato, M. Versiero, L. D'Elia, and F. P. Cappuccio (2007). Do statins reduce blood pressure? a meta-analysis of randomized, controlled trials. *Hypertension 49*(4), 792–798.

Sutton, M., R. Elder, B. Guthrie, and G. Watt (2010). Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers. *Health economics 19*(1), 1–13.

Taylor, F., M. D. Huffman, A. F. Macedo, T. Moore, M. Burke, G. Davey Smith, K. Ward, and S. Ebrahim (2013). Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev 1*(1).

Tobert, J. A. (2003). Lovastatin and beyond: the history of the hmg-coa reductase inhibitors. *Nature Reviews Drug Discovery 2*(7), 517–526.

Viscusi, W. K. (1993). The value of risks to life and health. *Journal of economic literature 31*(4), 1912–1946.

Viscusi, W. K. and J. E. Aldy (2003). The value of a statistical life: A critical review of market estimates throughout the world. *Journal of Risk and Uncertainty 27*(1), 5–76.

Ward, S., M. L. Jones, A. Pandor, M. Holmes, R. Ara, A. Ryan, W. Yeo, and N. Payne (2007). A systematic review and economic evaluation of statins for the prevention of coronary events.

Yogo, M. (2016). Portfolio choice in retirement: Health risk and the demand for annuities, housing, and risky assets. *Journal of Monetary Economics 80*, 17–34.