Essays on Preventive Care and Health Behaviors

ISBN: 978 90 5170 788 5

This book is no. **646** of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Essays on Preventive Care and Health Behaviors

Essays over preventieve zorg en gezondheidsgedrag

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof. Dr. H. A. P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
vrijdag 11 maart 2016 om 11:30 uur.

door
Hale Koç
geboren te Isparta, Turkey

Erasmus University Rotterdam

Promotiecommissie

Promotoren:	Prof. Dr. T. G. M. Van Ourti
	Prof. Dr. O. A. O'Donnell
Overige leden:	Prof. Dr. B. Donkers
	Dr. A. H. E. Koolman
	Dr. P. H. M. van Baal

Acknowledgements

Doing a PhD has been a long and challenging journey. Completing this journey, safe and sound, would not be possible without the accompaniment of some very special people in my life. I would like to express my gratitude to them in here.

I would like to start with my supervisors Tom and Owen. Working with you was a great privilege and thanks to you I have become a better researcher. Tom, you have been a wonderful mentor to me. You have always been accessible and I have benefited from your guidance not only academically but also personally. Owen, your respect and dedication for your work will be an inspiration for me for the rest of my career.

I am proud to be a member of the Health Economics group which has been a second home to me. Eddy, thank you for trusting and hiring me 4 years ago (despite my lack of experience in the field of health economics) and making all this happen. Hans, you are the best co-author one can ever ask for. I have enjoyed each and every one of our meetings over the Hong Kong-Rotterdam Skype line. Pilar, your door has always been open. I have always felt welcomed in your office no matter whether it is for a question or for a small break. Teresa, you have been the reminder of the nice, warm Mediterranean for me. Thank you for your Portugal tips, I hope to meet you there again! Darjusch, it has never been difficult to find something to talk about with you. I cherish our times inside and outside of the university together and looking forward to having many more of them in the future.

Writing a PhD thesis is not always a smooth process. Times of difficulty and despair have been short and bearable thanks to my fellow PhD students Arash, Kim, Max, Matthijs and Pourya. You made the 7th floor special and unforgettable. I will always remember our lunches and coffee breaks filled with laughter and intellect.

The research master years in the great city of Tilburg changed many lives forever including mine. I want to thank Haki, Gizem Hökelekli and Ayşegül for making "Number 9" as the most colorful house I have ever lived in. Sharing this house with you has contributed to who I am now and I am eternally grateful to have you in my life. Korhan, I still remember the warm smile on your face when we first met in the research master room. Thank you for being my friend, neighbor, and film festival buddy since then.

Gizem and Umut Karacadağlı, 11 years of friendship and still counting. Our weekends in Amsterdam have always helped me to remember that there is a life outside of the academia and it is fun!

My final words go to my family. I am blessed to have you as my family. You have put the love of learning in me and I thank you for your love and support at all times (Sevgili ailem, içime öğrenme sevgisini yerleştirdiğiniz ve beni her zaman desteklediğiniz için teşekkür ederim). Patrick, you are the biggest gift of the PhD to me. Without you, this thesis would not have been possible. I can not wait for our life together.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

One of the global health challenges of the 21st century is chronic diseases. As the leading cause of death, chronic diseases killed 38 million people worldwide in 2012, up from 31 million in 2000 (World Health Organization, 2014). Almost half of these deaths were premature, i.e. affect people younger than 70 years old (World Health Organization, 2014). Besides a heavy death toll, chronic diseases also hurt economies deeply. For example, chronic diseases costed the United States economy around 1.3 trillion dollar in 2003 in terms of treatment expenditures (20 percent) and reduced labor supply and productivity (80 percent; DeVol et al., 2007). Recognizing the rising burden of chronic diseases, in 2011 world leaders committed themselves to establish and strengthen national policies for prevention and control of chronic diseases in the United Nation Political Declaration on Noncommunicable Diseases (United Nations General Assembly, 2011).

Prevention methods in the industrialized world cover a wide range from health promotion activities that aim to encourage healthy behaviors to early diagnosis efforts with an objective to detect disease at a more treatable stage. The latter is commonly organized around population-based programs which extend disease screening (e.g. mammography for breast cancer, colonoscopy for colorectal cancer) to all members of a target population at little or no cost. The first part of this thesis raises two questions about organized screening or population-based prevention programs in high income countries in general: (i) are they effective in reducing chronic disease mortality (ii) can they address the widely documented socioeconomic disparities in preventive care use (Fletcher & Frisvold, 2009; Mullahy, 1999; Cutler & Lleras-Muney, 2010), and if yes, how?

Health behaviors in high income countries are under the spotlight in the second part of the thesis. According to Centers for Disease Control and Prevention (2009), four health behaviors – smoking, alcohol use, physical inactivity, unhealthy eating – are responsible for much of the disability, morbidity and mortality from chronic diseases in the United States. A similar picture emerges also for the remaining developed nations (World Health

Organization, 2009). Therefore it is justifiable to ask why people behave unhealthily, despite its harms, and how can we convince them to behave in a healthier way. When we look at time trends in health behaviors in the developed world, we see both positive and negative developments. For example there has been a considerable reduction in the prevalence of smoking during the second half of the 20th century in high income countries (Cawley & Ruhm, 2011). Alcohol consumption per capita has declined gradually in many European countries and in the United States over the same time period (Lakins et al., 2008; OECD, 2014a). In contrast, there is an unprecedented increase in unhealthy eating patterns and sedentary lifestyles that lead to obesity, especially in the United States and England (OECD, 2014b). This thesis puts forward lack of information/knowledge as a possible explanation for unhealthy behaviors and tests the ability of information provision in generating behavior change. Information argument is also used to explain why people from high socioeconomic class behave in a healthier way than people of low socioeconomic class.

Prevention is perceived by many as key to cut or avoid costs of treatment. When then the U.S. President Bill Clinton proposed a major health care reform in 1992, he argued that "it is just common sense ... [that] long-term costs to the health system will be lower if we have comprehensive preventive services". Not everyone agrees with this reasoning, at least not entirely. If incidence of the target disease is very small in the population, it is possible that prevention adds to health care costs rather than saves money (Russell, 1986). The reason is that prevention in this case is delivered to a large group of asymptomatic individuals only a small fraction of whom would develop the disease without the intervention and incur treatment costs. However, in health care, cutting costs should not be a goal in itself. What truly matters is whether we get enough health in return for our money – in other words whether prevention is cost-effective. If prevention produces more health with the same amount of resources compared to treatment, it should be implemented even if it doesn't save money. It has been shown that whether prevention is more cost-effective than treatment depends largely on type and frequency of the intervention, and target population of the intervention (Russell, 1993; Cohen et al., 2008; Russell, 2009). For example aspirin use to prevent heart disease and stroke is cost-effective among healthy, middle-aged men whose ten-year risk of disease is at least 5 percent. Screening for cervical cancer every three to five years is more cost-effective than no screening at all, but screening at a higher frequency costs more than it saves.

As well as its cost-effectiveness, health benefits of prevention are also open to discussion. It has been argued that the principle of "first do no harm" should be applied more strictly to the case of prevention because it is offered to healthy individuals, and without

proper assessment of harms and benefits, it carries the risk of turning them into patients, causing real harm (Gérvas et al., 2008; Welch et al., 2011). Screening for breast cancer is the most famous and heavily-debated example. According to Gøtzsche & Nielsen (2011) "for every 2000 women invited for screening for 10 years, one will avoid dying of breast cancer, while 10 healthy women, who would not have been diagnosed if there had not been screening, will be treated unnecessarily. Furthermore, more than 200 women will experience psychological distress including anxiety and uncertainty due to false positive findings." The International Agency for Research on Cancer (IARC), an independent organization within the framework of the World Health Organization, responds to this claim by arguing that the benefits of screening in terms of reduced breast cancer mortality outweigh its risks like false positives (Lauby-Secretan et al., 2015). Cutler (2008) provides evidence in line with the view of IARC by identifying screening as the main driver of the reduction in breast cancer mortality during the late twentieth century.

The second chapter of this thesis contributes to this debate by reporting evidence from the Dutch breast cancer screening program. The empirical strategy exploits the gradual spread of the program across municipalities between 1995 and 1997, which caused some women of a given age to be invited for screening for the first time later than others, to estimate the causal impact of the program on breast cancer and overall mortality. The results indicate that the program succeeded in reducing deaths from breast cancer. There is estimated to be around 170 more breast cancer deaths per 100,000 women at the end of a 17 year follow-up period among women who entered the program with a delay of at least 24 months and consequently received, on average, one less screening invitation than women who entered the program earlier. On the other hand, the program did not have any impact on overall mortality.

The third chapter looks at differences in the take-up of screening mammography across education groups in the Netherlands that has a national screening program and in the United States which does not. There are no educational disparities in mammography use among Dutch women invited through the national screening program for a free mammogram, while in the United States screening uptake differs by education level. Further analyses indicate that the education gradient in the United States is largely related to differences in income, insurance coverage and access to medical advice.

One of the most common explanations to the observed education gradient in preventive care use or health behaviors in general is health knowledge. In his seminal article, Grossman (1972) argues that educated individuals produce health more efficiently. One reason for this is allocative efficiency – higher educated individuals hold a higher level of health knowledge and on the basis of this choose healthier behaviors. Another reason is the productive efficiency – the more educated make use of a given amount of health

knowledge more efficiently. Chapter three considers whether the allocative efficiency hypothesis can explain why there is no education gradient in mammography screening in the Netherlands that operates an organized screening program while there is a gradient in the U.S. where there is no such program. The premise is that such a program may be expected to reduce the importance of information disparities by education. However, the analysis reveals that conditioning on subjective assessments of screening and breast cancer risk makes no difference or only slightly reduces the education gradient in the United States, suggesting that information disparities by education play at most a minor role in understanding education differences in screening behavior that exist in the absence of a universal screening program.

There are only a few studies which directly tested the allocative efficiency hypothesis as data on both health knowledge and health behaviors are scarce (Kenkel, 1991; Lange, 2011). Chapter four overcomes this issue by designing a Discrete Choice Experiment to answer the question why higher educated individuals follow healthier diets than the less educated. By randomly varying the information load that the respondents face, the experimental design allows testing between two explanations: health knowledge versus health valuation. The results indicate that the education disparities in diet derives mostly from superior health knowledge among the higher educated. When provided with information about adverse health consequences about unhealthy eating and daily recommended levels of important food components like calories, lower educated individuals switch from unhealthy meal choices to healthy ones, while hardly any change is observed in the choices of the higher educated. Nonetheless, even after fully equalizing health information across education groups, the lower educated tend to choose unhealthier diets, suggesting that they place a lower value on their health.

Chapter five evaluates the potential of information provision in inducing healthy behaviors among the chronically ill. Patients newly diagnosed with a chronic condition are exposed to a considerable amount of information. On one hand, they receive lifestyle advice from their doctor at the point of diagnosis which emphasizes the connection between health outcomes and health behaviors such as smoking and exercise. On the other hand, a new diagnosis informs the individual about his/her true underlying health status which can be taken as a wake-up call for change among some patients. Therefore, it is possible to see getting diagnosed with a new chronic condition as an equivalent of a strong information treatment at the individual level. The results show that, in line with the doctor's recommendation, individuals abstain from smoking and drinking, both in the short- and long-run, after being diagnosed with a chronic condition. This finding holds equally true for major conditions, such as cardiovascular disease, and minor conditions, such as diabetes, with the caveat that magnitude of the response in terms of smoking is

larger for major conditions. However, in terms of losing weight and getting physically active, there is hardly any response, if not a tendency towards the opposite of what is recommended.

Chapters four and five illustrate the role that information plays in modifying behavior. A well-studied example is the case of smoking. Demand for cigarettes dropped significantly in response to the new scientific evidence linking smoking to lung cancer in the 1950s. Cigarette consumption further dropped as a result of health warning labels on packages, mass media anti-smoking campaigns, and other public health efforts that disseminate information about health risks of smoking (Chaloupka & Warner, 2000). More recent but less conclusive evidence comes from the case of nutrition information. Some studies have shown that calorie consumption went down in response to mandatory menu-labeling regulations that mandate posting calorie information on menus in chain restaurants (Wisdom et al., 2010; Bollinger et al., 2011), yet some others find no impact on healthier food purchasing in fast-food chains (Elbel et al., 2009; Finkelstein et al., 2011).

This thesis contains valuable insights for policy discussions. First, the finding that the Dutch breast cancer screening program reduces breast cancer mortality is of considerable importance in the context of doubts about the effectiveness of such programs. This finding is based on a solid identification strategy, hence represents a valuable contribution to the current, extremely fragmented screening debate where not all evidence presented is of equal quality. But the lack of the impact of the program on all-cause mortality does leave some doubt over the effect on population health. The evidence presented in the thesis suggests that the role that health knowledge plays in generating educational disparities differs across health behaviors. Hence, the success of interventions that distribute health knowledge with the aim of eradicating educational disparities is likely to depend on the behavior that is targeted.

Chapter 2

Effect of Population-Based Screening Mammography on Breast Cancer Mortality in the Netherlands

Despite its power to detect tumors at an early stage while they are still treatable, the effectiveness of screening mammography in preventing breast cancer deaths has recently come into doubt. In this paper we evaluate the effect of the Dutch population-based screening mammography program, on both breast cancer and overall mortality. Identification relies on the geographic expansion of the program between 1995 and 1997 across municipalities. Using administrative and mortality register data, we find that a delay of at least 24 months in entry to the program – equivalent to receiving one less screen in a lifetime from the age at which a woman becomes eligible – raises breast cancer mortality by about 170 per 100,000 women, or by 10 percent relative terms.

This chapter is based upon:

Koç, H. and O'Donnell, O. and Van Ourti, T. and the National Evaluation Team for Breast Cancer Screening (2015). *Effect of Population-Based Screening Mammography on Breast Cancer Mortality in the Netherlands.* (Mimeo). Erasmus University Rotterdam.

2.1 Introduction

In Europe, around 460,000 women were newly diagnosed with breast cancer in 2012, and 131,000 died from the condition (Ferlay et al., 2013). By detecting tumors at an early stage while they are still treatable, screening with mammography is expected to reduce mortality from breast cancer. Starting with Health Insurance Plan trial in 1963 (Shapiro et al., 1982), the impact of screening on breast cancer mortality has been evaluated via several randomized controlled trials (RCTs). Based on favorable results emerging from these trials, several countries initiated nationwide organized screening programs in the late 1980s and 1990s. In 2002, the International Agency for Research on Cancer, an independent organization within the framework of the World Health Organization, evaluated the most recently published results from the RCTs, and estimated that screening reduced relative risk of breast cancer mortality by 25 percent among women aged 50-69 (International Agency for Research on Cancer, 2002, p. 98-99). The European Union has supported the implementation of screening programs as a part of its fight against cancer (Commission of the European Countries, 1990). Expert groups like the National Cancer Institute, the American Cancer Society, and the United States Preventive Services Task Force have also recommended use of screening mammography for early detection (Dodd, 1992; U.S. Preventive Services Task Force, 1989).

However, 40 years after the start of the first randomized trial, the medical benefits of screening with mammography have been questioned (Nelson et al., 2009; Independent UK Panel on Breast Cancer Screening, 2012; Gøtzsche & Jørgensen, 2013). Methodological concerns with the randomized trials, advances in breast cancer treatment and results from observational studies have fueled this discussion. This paper contributes to the ongoing debate by reporting evidence from the Dutch breast cancer screening program. This organized screening program invites all women aged 50-69 biennially to breast cancer screening.[1] The program spread gradually across municipalities between 1989 and 1997, which caused some women of a given age to be invited for screening for the first time substantially later than others, creating variation both in the age at first screen and in the total number of screens a woman could receive over the age range that determines eligibility.

We use administrative data on postal address, date of birth, date and cause of death on more than 250,000 women for the period 1995 to 2011. Since we do not have access to mortality data before 1995, we use the part of the roll-out that happened between 1995 and 1997 to find women who are otherwise comparable but differ in their time of first entry to the program. Our identification strategy rests on the assumption that geographical roll-

[1] Formally the target age range is 49-68. Women receive their first invitation in the year they turn 50.

out of the program occurred in a non-systematic manner, creating exogenous variation in the timing of invitation for a first screen. We show that there is no association between the spread of the program across municipalities and pre-program differences in breast cancer prevalence and incidence, as well as other population characteristics that may be related to breast cancer mortality.

Our results indicate that the program succeeded in reducing deaths from breast cancer. We estimate 172 more breast cancer deaths per 100,000 women at the end of a 17 year follow-up period among women who entered the program with a delay of at least 24 months and consequently received, on average, one less screening invitation than women who entered the program earlier. Magnitude of this effect is larger compared to other studies in the literature possibly due to differences in methodology, underlying assumptions and populations analyzed. On the other hand, we do not find any impact of the program on overall mortality. Our estimates are robust to a battery of checks on the validity of the identification. Allowing for provincial differences in breast cancer mortality which could arise from differences in the quality of breast cancer treatment and for women who move across municipalities has little or no impact on the estimates. A placebo test builds further confidence that the effect estimated is real by demonstrating that application of the identification strategy to older women, who are not eligible for the program, produces no significant impact on the breast cancer mortality of this group.

One of the central questions of the screening debate is how many breast cancer deaths can be averted with mammography. Miller et al. (2014) attempt to answer this question by using data from the Canadian National Breast Screening Study, which is an RCT that began in 1980s and involved 90,000 women. Half of the women between the ages of 50 and 59 years were randomly assigned to a treatment group that received an annual mammography screen for five years in addition to physical breast examinations. The control group received physical examinations only. After 25 years, breast cancer death rates were similar between treatment and the control groups (relative risk=1.02, no estimate of the absolute risk is available). The Swedish Two County Trial was conducted at the end of 1970s. It randomly invited women aged 40-49 years to receive mammography screening once in every 24 months, and those aged 50-74 once in every 33 months for seven years (Tabár et al., 2011). After 29 years, relative risk of death from breast cancer was reduced by 27-31 percent among women invited to screening, corresponding to an absolute risk reduction of 166-200 per 100,000 women.

The Canadian and Swedish trials provide contradictory evidence on the effectiveness screening in preventing deaths from breast cancer. A meta analysis of eight trials concluded that more than half of the trials were methodologically problematic (Gøtzsche & Jørgensen, 2013). In some studies, the randomization process failed to create similar

treatment and control groups and in others different study exclusion criteria were followed. Determination of cause of death was not blinded to treatment assignment in some experiments.[2] Trials judged to have made adequate randomization found that screening resulted in no significant reduction in breast cancer mortality, while trials with suboptimal randomization indicated a 25 percent decrease in relative risk due to screening.

Advances in breast cancer treatment – specifically, adjuvant systemic therapy – have contributed to the doubt about the effectiveness of screening in reducing breast cancer mortality (Jatoi & Miller, 2003).[3] Bleyer & Welch (2012) argue that screening mainly identifies small tumors that have no potential to become life-threatening disease. They claim that from 1976 to 2008, screening in the United States only marginally reduced the incidence of late-stage regional cancer, which often can be treated successfully anyway, and had no effect on that of life-threatening late-stage distant cancer. The observed reduction in breast cancer mortality in this period must, it is argued, be largely the result of improved treatment rather than screening.

In this paper we present evidence on the effect of a population-based screening program on mortality. Apart from the need for new sources of information in the face of doubts about the quality of the RCTs, there is a need for evidence on the life saving potential of screening when implemented in a large-scale program. Results from RCTs may not be relevant to establishing the effect of a nationwide program. Women volunteering for trials are not necessarily representative of the whole population, and RCT screening centers may not resemble the actual setting in which organized screening is offered with respect to the quality of staff and control of screening quality indicators like the number of false positives (Allcott, 2015).

The study closest in spirit to ours is Kalager et al. (2010) which evaluates the mortality effect of the Norwegian breast cancer screening program. Using the geographical roll-out of the program, as we do, to identify the effect of the program, a reduction of 2.4 breast cancer deaths per 100,000 person-years among screened women is estimated to be causally attributable to screening. This constitutes only one third of the reduction observed in breast cancer mortality over time between screened and unscreened women. The remaining two thirds can be due to advances in cancer treatment/diagnosis and heightened cancer awareness. At first sight, the magnitude of our estimate of the program

[2] The Canadian trial is one of the RCTs that are classified as adequately randomized. However, in the case of Swedish trial, treatment and control groups are argued to be incomparable and cause of death assignment was not blind. In response, the Swedish Cancer Society convened an independent overview committee to investigate these claims. The committee concluded that the original finding of lower breast cancer mortality among women invited to screening is robust.

[3] Adjuvant therapy is any treatment given after primary therapy to increase the chance of long-term disease free survival. It can include chemotherapy, hormonal therapy, radiation therapy, or a combination of treatments to kill any cancer cells that may have spread to the other parts of the body.

effect might seem huge in comparison to Kalager et al. (2010). However, when we express our estimate per 100,000 person-years, like Kalager et al. (2010) do (estimates per 100,000 women is not available), we get a reduction of 10 breast cancer deaths per 100,000 person-years at the end of a 17 year follow-up period among women who entered the program earlier.[4] Possible reasons for the difference between the two estimates are the difference in the follow-up periods, method of identification of the program effect and calculation of person-years.

Kalager et al. (2010) separate the impact of mammography screening from that of improved treatment and of greater awareness of the risks of breast cancer only under the assumption that the non-screening contributions to reduced mortality were the same in more and less recent periods. This is a highly questionable assumption given the steady advances in breast cancer treatment since the 1980s. By following a cohort of women who differ by at most 36 months in their time of entry into the screening program, we are able to more plausibly assume that differences in rates of mortality are not attributable to the treatment technology available. Robustness of our results to geographical differences in treatment quality supports this assumption. By following the same cohort of women, we also avoid changes in age composition, the primary determinant of breast cancer risk in a population, that could bias a study, such as that of Kalager et al. (2010), that makes comparisons across periods of screening and non-screening within a given region. Finally, RCTs have shown that it takes between 7 to 10 years to observe the impact of screening on breast cancer mortality statistics (Jatoi & Miller, 2003). Therefore, a 17 year follow-up period, in comparison with the 2.2 years that Kalager et al. (2010) follow an average female, allows us to estimate effects over a period of sufficient length for the full mortality effects of a screening program to emerge.

Otto et al. (2003) evaluate the impact of the Dutch breast cancer screening program on breast cancer mortality by analyzing the trend in breast cancer mortality between 1980-2001. They adjust the trend for gradual implementation of the program by denoting the year when the program was introduced in a municipality as 0. The authors estimate a 1.7 percent reduction in breast cancer mortality per year among women aged 55-74 after the introduction of the program compared to a pre-program period (no estimate of absolute or relative risk is provided in this study). The main disadvantage of this study is the absence of a real control group which would describe the trend in breast cancer mortality in the absence of the program. This deficiency has been tried to be compensated by use of a simulation model (MISCAN) to create a non-screened group, however this model cannot incorporate improvements in treatment over time. Our study uses geographical

[4] Person-years is a measure that combines the number of persons and their time contribution in a study. It is calculated by summing the number of years that each individual in the sample is observed.

roll-out of the program to find women who are yet to be invited and therefore can be used as controls for women who are already receiving invitations. Moreover, similar to Kalager et al. (2010), the follow-up period of this study is too short – only 4 years after the program reaches full-coverage of its target population – to observe the full mortality effect of the program.

Otten et al. (2008) look at the breast cancer incidence and mortality trends from 1975 till 2006 to assess the Dutch screening program. They show that breast cancer mortality declines by 2.3% per year for age group 55-64, and 2.8% for age group 65-74 after 1994, i.e. 5 years later than the start of the program (no estimate of absolute or relative risk is provided in this study). Compared to Otto et al. (2003), this study has a longer follow-up period of 9 years but the analysis solely relies on time-trends. Consequently, disentangling the mortality effects of screening, improved treatment and changing risk factors is not possible. The results are purely descriptive and it is not known whether the estimated reductions in breast cancer mortality can be causally linked to screening. In contrast, as we mentioned above, our study design allows us to separate the impact of screening from that of treatment technology on mortality.

Van Schoor et al. (2011) and Otto et al. (2012) conduct recent case-control studies to evaluate the mortality effect of the Dutch breast cancer screening program. Both studies assess the life saving potential of screening for the period when breast cancer is developing and potentially detectable by screening but not yet by a physical exam. It is estimated that screening mammography lower the relative risk of breast cancer mortality by 28 percent (1977-1991) and 65 percent (1992-2008) among women who attended either of the two screens preceding diagnosis compared to those who attended none (Van Schoor et al., 2011), and by 44 percent (1995-2003) among women who attended the screen just before the diagnosis compared to those who did not (Otto et al., 2012). These estimates are larger in comparison to our estimate of a reduction of 10 percent in relative risk of breast cancer mortality possibly due to the fact that we estimate the mortality effect of receiving one extra invitation, on average, in a lifetime from the program while Van Schoor et al. (2011) and Otto et al. (2012) estimate the effect of attending to the screening(s) prior to the diagnosis.

The main weakness of a case-control design is the inability to provide an estimate of the attributable risk (see Appendix 2.A for the derivation). From the relative risk, one cannot infer how many breast cancer deaths can actually be averted by screening. We estimate the attributable risk in addition to relative risk. Moreover, we do so without invoking the assumption that death from breast cancer is a rare event – rare disease assumption – which is required to identify the relative risk with a case-control design. Additionally, case-control studies are prone to bias since it is difficult to match cases

and controls on all relevant risk factors for breast cancer mortality. For example, Otto et al. (2012) require cases and controls to be born at the same year, receive the same number of invitations before case's diagnosis and get their first invitations at the same year. Van Schoor et al. (2011) correct only for differences in age at first invitation. In both studies, it is highly possible that cases are different from controls with respect to other observable (e.g., number of children, age of first birth) and unobservable (e.g., genetic susceptibility to breast cancer) risk factors. We compare across women who differ in timing of entry to the screening program only as a consequence of their municipality of residence. We confirm the non-systematic roll-out of the program by examination of municipality characteristics. Finally, case-control studies only look at the survival of women who did get screened. This is insufficient to judge a nationwide program because there are women who decide not to screen.

The rest of our paper is organized as follows. In section 2.2, we provide some institutional background on the Dutch breast cancer screening program. Section 2.3 describes the data used in the analysis. In section 2.4, we outline the empirical strategy. Section 2.5 documents and discusses the results, and section 2.6 checks for the internal validity of the results. Finally, we discuss the implications of our findings and conclude in section 2.7.

2.2 Institutional Setting

A population-based, organized screening program was initiated in the Netherlands in 1989. The target group of the program was women aged between 50 and 69. The *initial program* has gradually expanded across municipalities over a period of 9 years and in 1997 reached full coverage of its target group. During the 1990-1997 period, around 750,000 women were targeted by the program per year and on average 78.2 percent of the invited women participated (Appendix Table 2.B.1). In 1998, the upper age-limit of the program was extended to women aged 70-75. Within 3 years, *the extended program* also reached full coverage and since 2001, all women in the Netherlands between the ages 50 and 75 have been invited to participate in the screening program. In 2011, 80.1 percent of the invited women attended screening (Appendix Table 2.B.1).

The target women are identified via personal records provided by the municipal population registers. The records contain a unique identifying number, place of residence and the date of birth. All eligible women living in the same postcode area receive invitations at the same time. Every two years women get a personalized invitation letter with a fixed appointment to get screened.[5] A screening examination is usually carried out in a mobile

[5] Every year, two to seven percent of the target population, called definitive non-participants, indicate

screening unit. Women who do not respond are sent a reminder after 2-3 months.

Screening results are independently read by two radiologists and sent out to women in writing. Patients are recalled for further examinations only if both radiologists agree. In this case, the woman's general practitioner is informed in advance.

During the period we use to provide variation in screening exposure (1995-97), the primary test offered was X-ray mammography. Since the second half of 2004, digital screening began to be offered and in 2010, 94 percent of the screening examinations were digital.

2.3 Data

Via Statistics Netherlands, we have access to individual level administrative data that cover the period from January 1995 until December 2011. These data contain, for every female living in the Netherlands, date of birth, date and cause of death (if applicable), and residential address at postcode level.[6] From the Dutch Cancer Registry, we obtain data, at the municipality level, on the invasive and non-invasive breast cancer incidence rate for the years 1989-2011, and on the 5-year invasive breast cancer prevalence rate for the years 1994-1995. The National Evaluation Team for Breast Cancer Screening (NETB) provides us with data, at the municipality level, on the rate of participation in the program when it was first introduced in a municipality. Thanks to NETB, we also know the launch date (i.e. month and year) of the screening program in every municipality (see Otto et al., 2003 for more information on the geographic expansion of the program.). Based on this information we calculate, for every female, the date on which she receives her first invitation which happens on the earliest date she falls into the target age range of the program in her municipality of residence.

The fact that our mortality data start from 1995, while the screening program started already in 1989, has two important consequences. First, we cannot identify the full impact (if any) of the program on mortality since 1989. Second, even with respect to the impact on mortality since 1995, the estimate would suffer from survival bias if the program does indeed reduce breast cancer mortality and we were to use variation in entry to the screening program prior to 1995. The survivors to that year in the municipalities that have been in the program for a shorter period should be, on average, at lower risk of

that they do not want to receive invitations for screening (National Evaluation Team for Breast Cancer Screening, 2014). These women leave the program by their own choice because of medical reasons (already having checks/treatment for her breasts), or procedure related reasons (unpleasant experiences with the screening in the past) or personal reasons (disbelief in benefits of screening).

[6] Source of address data is municipality records. Every individual living in the Netherlands has to register his/her address with the municipality.

breast cancer mortality than the survivors to 1995 in the municipalities that have been exposed to the program for longer. If one compares the mortality of these survivors, then there will be a bias towards lower subsequent breast cancer mortality in the municipalities with shorter exposure. So, if there is an effect, we would obtain an underestimate of it if variation in screening over the full roll-out of the program from 1989-1997 were used.

Although we cannot do anything about the first problem, we can circumvent survival bias and get an unbiased estimate of the mortality impact of the roll-out during 1995-1997. To this end, we restrict the analysis to women who, in January 1995, were aged 49-68 and living in municipalities that started to offer screening for the first time in 1995 and afterwards. All women in this subsample are eligible to receive an invitation from the program as of January, 1995, and except the ones who moved from the municipalities which have already implemented the program, none of them has been invited by the program before. The assumption here is that moving is unrelated to the roll-out of the program. We consider it highly unlikely that high-risk women would move from one municipality to another simply to get access to mammography screening. According to Schouten et al. (1996), in the Netherlands, only less than 2 percent of the population above the age of 40 move annually. But if this occurs, then the estimated effect will have a downward bias. In a robustness check, we test the sensitivity of our estimates to dropping women who move between municipalities from the sample.

30 percent of the municipalities whose launch date of the program is known to us started organized screening between January 1995 and December 1997 (n=197). We are able to match addresses to 138 municipalities.[7] Of these, 67 implemented the program in 1995, 52 in 1996 and 19 in 1997 (see Figure 2.1).

We exclude women who were aged 64-68 in January 1995 since there is unobserved variation in this group in the time of entry to the *extended program* that raised the upper age limit for screening to 75 years in 1998-2000.

With these exclusions, we focus on Dutch females who, in January 1995, were aged 49-63 and living in the municipalities that started to offer organized screening in 1995 or afterwards. We follow this fixed sample of women regardless of mortality, or moving between municipalities, until December 31, 2011, giving us a balanced panel of 17 years. Women who move into the municipalities of analysis after January, 1995 are not included.

[7] We have access to data on residential addresses at the postcode level starting from 1995. The *key* to match addresses with municipalities is available only from 1999. Out of 197 municipalities which started organized screening in 1995 and afterwards, 54 of them had merged with bigger municipalities by 1999. We do not observe these municipalities in our dataset. As a result, we define the date of first invitation for a woman living in municipality A, which became municipality B sometime before 1999, with reference to the program launch date in B. This is unlikely to cause much bias because municipalities usually merged with adjacent municipalities which have similar launch dates. We dropped 5 municipalities for which breast cancer incidence/prevalence data are missing.

There are 263,777 women who satisfy our sampling criteria. For 1380 women who were alive in January 1995, but died before the program arrived to their municipality, we attribute a date of program entry according to the municipality of residence in January 1995. We drop 4803 women as they left the Netherlands and fell outside of the coverage of the screening program. We also drop 82 women who move to municipalities, before getting their first invitation, for which the start date of the program is unknown to us. We further exclude 1793 women who moved to municipalities which started organized screening before 1995, and got their first invitation there. We cannot calculate pre-program breast cancer incidence and prevalence rates or these municipalities, as explained in Section 2.4. We finally exclude 387 women whose date of first invitation is delayed for more than 35 months because of frequent movement between municipalities. In the end we follow 256,712 women and observe 3,604 breast cancer deaths out of a total number of 38,341 deaths during the course of 17 years. Mean age in the sample in January 1995 was 55 years, and average waiting time for the first invitation is 13 months (Table 2.1).

2.4 Estimation Strategy

We estimate the causal impact of the length of delay before entering the program on the cumulative mortality from breast cancer and all causes, at the end of a follow-up period of 17 years. Our identification strategy utilizes part of the geographical roll-out of the program that occurred between January, 1995 and December, 1997. We define length of delay as the number of months a woman has waited to receive her first invitation since January 1995.

We do not observe whether each woman responds to the invitation. However, we are able to establish that there is no statistically significant relationship between the introduction date of the program and participation in the program when it first arrived to a municipality (see Table 2.2, column II). Given this, while what we identify is the impact of a delay in the receipt of the first invitation for screening, it seems safe to assume that this corresponds to the impact of a delay in the receipt of first screening, after taking into account the fact that not all invited women get screened.

We estimate the following model

$$Y_i = \gamma_0 + \gamma_1 \text{Age}_i + \gamma_2 \text{Delay}_m + \varepsilon_i, \tag{2.1}$$

where Y_i is a binary variable which takes the value of 1 if woman i is dead from breast cancer by the end of 2011, and 0 otherwise, i.e. alive or dead from another cause.[8]

[8] Deaths whose primary cause is registered as breast cancer are counted as breast cancer deaths.

Age is a categorical variable that indicates whether woman i was of age 49-54, 55-59 or 60-63 years in January 1995.[9] *Delay* is equal to 1 for women who had to wait for 24 to 35 months before getting their first invitation, and 0 for those who waited for less than 24 months. We choose to dichotomize the program variable this way because, given age, in a biennial screening schedule, receiving the first invitation with a delay of 24 months or longer translates into the receipt of one less invitation/screening from the program in a lifetime. Note that since the program was rolled out at the municipality level, our program variable varies at the municipality level (m). ε_i is a random error term representing unobserved determinants of breast cancer mortality. We cluster standard errors over women living in the same municipality at the time of the first invitation.

If roll-out of the program has indeed created variation in the length of delay that is not correlated with breast cancer mortality risks, then, γ_2 is the causal impact on breast cancer mortality of waiting 2 years or longer for an invitation for screening.

We estimate Model 2.1 also for all-cause mortality. The effects may differ from those on breast cancer mortality because of a) screening influencing the attribution of deaths to breast cancer (Black et al., 2002), b) deaths arising from treatment of cancers that were not life-threatening (Early Breast Cancer Trialists' Collaborative Group, 2000), and c) competing risks (Honoré & Lleras-Muney, 2006).

In order to check whether the geographical roll-out of the program is indeed unrelated to breast cancer mortality risks, we regress the binary program delay variable on indicators of those risks at the municipality level prior to 1995, as follows[10]

$$\begin{aligned}\text{Delay}_m = \alpha_0 + \alpha_1 \text{Incidence}_m + \alpha_2 \text{Prevalence}_m \\ + \alpha_3 \text{Participation}_m + \alpha_4 \text{Population}_m + \alpha_5 \text{Target}_m + \epsilon_m\end{aligned} \tag{2.2}$$

We do not have data on pre-program breast cancer mortality rates, and use crude breast cancer incidence and prevalence rates as proxies. Given fluctuations in breast cancer incidence at the municipality level, we use the median rate over the 1989-1994 period ($Incidence_m$). For prevalence, we use the 5-year invasive breast cancer prevalence rate as of January 1994 ($Prevalence_m$). Since incidence counts the number of new breast cancer cases and prevalence counts the number of survivors, the difference between them provides an estimate of breast cancer mortality. *Participation* is the share of invited

These are deaths with ICD-9 code of 174 or ICD-10 code of C50.

[9] The findings are robust to controlling for age more flexibly via single year dummies.

[10] In addition, we regress a continuous variable which shows the implementation date (i.e. month and year) of the program on the same regressors. The results (not shown) are highly consistent with those presented in Table 2.2. The findings are also robust to using a categorical variable which shows the half-year period (i.e. first half of 1995, ..., second half of 1997) when the program was launched in an ordered probit model (see Appendix Table 2.B.2).

women in the municipality who responded positively to the invitation and got screened when the program was first introduced. *Population* is the average size of the female population and *Target* is the average share of the target age (i.e. 50-69) women in the female population between 1989 and 1994.

Table 2.2 presents estimates from four different variations of Model 2.2. The most important finding is that the introduction date of the program is not statistically related to the breast cancer incidence or prevalence rate before the start of the program (column I), and so is unlikely to be related to pre-program breast cancer mortality. This is in line with what is suggested by Figures 2.2 and 2.3, which respectively plot breast cancer incidence and prevalence rates against the program introduction date. Table 2.2 additionally reveals, as mentioned above, that there are no significant differences in the uptake of the screening invitation across municipalities with different implementation dates (column II).[11] Finally, the launch date of the program in a municipality is not correlated with demographics (column III). Column IV confirms that the program expanded across municipalities irrespective of pre-program breast cancer incidence and prevalence rates and demographics, and that uptake was not related to the launch date.[12]

2.5 Results

Estimates of Model 2.1 presented in the first column of Table 2.3 indicate 172 more breast cancer deaths per 100,000 women when implementation of the screening was delayed by at least 24 months in comparison with a delay between 0 and 23 months. This implies an increase of about 12 percent in the cumulative breast cancer death probability. This effect is equal to that of 5 years age difference (55-59 years versus 49-54 years). Women who were aged 55-59 years in January 1995 experienced 173 more breast cancer deaths per 100,000 women over the course of 17 years than women who were aged between 49 and 54.

We have demonstrated that the delay in program implementation is not significantly related to pre-program incidence/prevalence of breast cancer, nor to population size and the share of women within the target age group. This is not sufficient to exclude these potential confounders from the outcome model. Belloni et al. (2014) demonstrate that model

[11] We have no knowledge on the existence of differences in the share of definitive non-participants across municipalities. However, even if such differences exist, and even if they are related with the roll-out of the program, our estimates are unlikely to be affected from this since we explicitly account for differences in the share of women with breast cancer across municipalities by controlling for pre-program breast cancer incidence/prevalence.

[12] For sub-sets of municipalities with available data, we have confirmed that the program launch date is not correlated with average household income in 2005 (N=122) and with average female labor force participation rate during 1996-1998 (N=82).

selection solely on the basis of correlation of covariates with the treatment variable results in invalid inference for the treatment effect. We therefore test whether the same covariates are significantly correlated with breast cancer mortality, given treatment. The results presented in columns II-IV of Table 2.3 confirm that pre-program incidence/prevalence and demographics, entered separately or simultaneously are not significantly correlated with breast cancer mortality. According to Belloni et al. (2014)'s double selection criterion, the appropriate specification is the one given in the first column. This also gives the smallest (but not significantly different from the estimates in columns II-IV) estimate of the treatment effect and so relying on it is conservative with respect to the claimed impact of screening.

Entering the program with a delay can affect breast cancer mortality via two mechanisms. First, it can reduce the total number of invitations, and consequently the number of screenings a woman can get from the program in her lifetime. Second, for a given number of invitations, it can raise the age at which a woman receives her first screening. Breast cancer incidence rises with age, therefore being exposed to screening for the first time at a later age reduces a woman's chances of early diagnosis, and consequently survival. Using a binary program variable provides an estimate of the combined effect of these two mechanisms.

In Table 2.4, we show the results with using a categorical program variable which groups the number of months a woman waited before entering the program into six categories: 0-5 months, 6-11 months, ..., 30-35 months. We control only for age having confirmed that the pre-program incidence/prevalence and population size/share in target group covariates are not significantly correlated with either the categorical treatment variable or breast cancer mortality conditional on this.[13] The estimates provide suggestive evidence that the reduction in the total number of invitations/screenings is the main mechanism behind the program effect. There is no significant (statistically or substantively) effect of any delay between 6 and 23 months in comparison to waiting for less than 5 months. With biennial screening, variation in the delay to program entry of less than 24 months is likely to produce little variation in the number of screens received. Rather, it is variation in age at first screening that will be generated. The lack of any effect below this threshold suggests that this variation contributes less to the estimated treatment effect of waiting for more than 24 months to enter the program. The immediate increase in the magnitude and significance of the effect on crossing the 24 month threshold supports our baseline specification that uses the binary treatment variable. The point estimate rises further, but not significantly, for a delay of more than 29 months.

[13] The categorical treatment variable is regressed on the covariates using an ordered probit. Estimates from this and the mortality regression with this specification of treatment plus the controls are given in Appendix Tables 2.B.2 and 2.B.3.

The hypothesis that the effect of delayed program entry differs by age can be tested more directly by allowing the effect of a delay of at least 24 months to differ for women aged 49-54, 55-59 and 60-63 in January 1995. Doing so suggests that entering the program with a delay has the largest impact on breast cancer mortality of women aged 55-59 and on women aged 49-54 next (Appendix Table 2.B.4). However due to the small sample size within each age group, standard errors are very large. Pairwise Wald tests indicate no age differences in program effect in statistical sense. We believe that this is due to the loss of power of the test resulting from the small sample. In short, we see some indication of an age-specific program effect but a bigger sample is needed to draw reliable inference.

In Figure 2.5, we plot the evolution of the program effect over time (Appendix Table 2.B.5 contains the coefficient estimates). It is widely claimed in the medical literature that it takes 7 to 10 years for the impact of screening on breast cancer mortality to become apparent (Jatoi & Miller, 2003; Tabár et al., 2011). Our estimates suggest that there is indeed a lag before the effect emerges, but it is not quite as long as 7-10 years. Not until the end of 2001, when the program had been operating for up to seven years in the reference group and for no more than five years in the delayed treatment group, do we observe a statistically significant and sustained difference in the cumulative number of breast cancer deaths. The difference gets larger over time. In 2001, there were 97 excess breast cancer deaths per 100,000 women in the group that experienced a delay of at least 24 months before entering the program. This number increased by almost 80 percent within 10 years and reached to 172 in 2011. This is in alignment with other studies which report a positive correlation between the length of the follow-up period and the number of lives saved by screening (Tabár et al., 2011; Nyström et al., 2002). This correlation arises mainly because of the time lag between the detection of breast cancer and death from it.

Estimates of the impact of the delayed implementation of the screening program on all-cause mortality are presented in Table 2.5. In all specifications the standard error is large relative to the point estimate and there is no statistically significant effect of the program on all-cause mortality. The standard error is large also relative to the standard error of the point estimate for breast cancer mortality (see Table 2.3) due to lower explanatory power of the model and the resulting large error term. Unlike breast cancer mortality, the point estimate for all-cause mortality changes dramatically between columns II and III. Following the double-selection criterion for model specification Belloni et al. (2014), the preferred model is the one in the far right column, which suggests that all-cause mortality is correlated with pre-program incidence/prevalence of breast cancer and population size, but not with the delay in entry to the screening program.

These results suggest that while the program has successfully reduced the number of

deaths from breast cancer, it failed to make any impact on mortality from all causes. It has been argued that a discrepancy between the impact of screening on disease specific and all-cause mortality may arise due sensitivity of breast cancer mortality to correct identification of cause of death (Penston, 2011; Black et al., 2002). The high degree of reliability of cause-of-death statistics in the Netherlands (Harteloh et al., 2010) makes errors in the registration of deaths by cause an unlikely explanation.

Another explanation could be the existence of non-breast cancer deaths that are related to breast cancer screening or subsequent early treatment, but not counted in the breast cancer mortality figures. Fatal complications may arise from an invasive procedure carried out in response to a positive screening result. It has been shown that adjuvant tamoxifen therapy for early stage breast cancer is associated with an increased risk of developing endometrial cancer and, in the case of long-term use, dying from endometrial cancer (Van Leeuwen et al., 1994; Bergman et al., 2000). Use of radiation therapy on women with left-sided breast cancer can increase the probability of radiation induced heart disease, especially in young patients. Harris et al. (2006) document a higher rate of cardiac deaths, as well as chest pain, coronary artery disease, and myocardial infarction among left-sided breast cancer patients at 20 years after the receipt of radiation therapy as part of breast-conservation treatment for early-stage breast cancer. To the extent that these screening and treatment related deaths are attributed to causes other than breast cancer, the estimated impact of screening on breast cancer mortality will be of greater magnitude than the impact on all-cause mortality.

Finally, competing risks is a possible explanation for the lack of any significant impact on all-cause mortality. Thanks to the program more patients survive breast cancer and become at risk of dying from non-breast cancer-related causes. In the presence of dependent risks, a breast cancer survivor dies from another cause which shares the same risk factors with breast cancer. If risks are independent, the survivor dies from an unrelated cause "X" which does not have any common risk factor with breast cancer. In this case the person dies because she lived long enough to contract with cause "X". Looking at a cohort of women who are disease free after 5 years of adjuvant therapy, Chapman et al. (2008) find that 60 percent of the deaths observed during the median follow-up period of 4 years are due to causes other than breast cancer.

Our current analysis does not allow us to distinguish between these different explanations. Future research will look at the impact of the program on mortality from a variety of causes. Higher mortality rates from screening/treatment related causes, like endometrial cancer or cardiovascular disease, among women who entered the program with a shorter delay, and got one extra screening, will imply that screening kills as many women as it saves. In contrast, if existence of competing risks is the true explanation,

then higher mortality rates from a wider variety of causes should be observed among women with a shorter delay.

2.6 Internal Validity

In this section we assess the validity of our estimates and check their robustness to a number of assumptions we have made. Results are summarized in Table 2.6.

We are identifying the effect of the screening program by using geographic variation in its time of implementation. We have established that the start date of the program and breast cancer mortality are uncorrelated with pre-program breast cancer incidence/prevalence and the (target) population of municipalities. While this is reassuring, it remains possible that the estimates are biased by correlated unobservable determinants of breast cancer mortality that vary geographically with the start date of the program. One particular concern might be variation in the quality of treatment available. Comprehensive health insurance coverage with a sophisticated risk equalization system should limit the scope for such differences. But they cannot be ruled out. In years for which we observe variation in exposure to screening hospitals operated under a fixed budget determined by the government. All care delivered in a hospital including the drugs prescribed had to be financed from this budget. Niezen et al. (2006) argue that, under this system, there was considerable geographic variation in the availability of new, expensive cancer treatment drugs (e.g., Taxol-introduced in 1996, Herceptin-introduced in 1999). For example, small hospitals which find it more difficult, relative to large hospitals, to economize on other costs and make extra budget for the expensive drugs failed to offer them to their patients. Referring patients to larger or more specialized hospitals did not solve the problem as well, because these hospitals did not also have enough budget to treat such a large influx of expensive patients. The necessity to allocate patients to limited resources forced each hospital to give priority to the patients living in the regions they serve, turning where you live into an extremely important determinant of the quality of the care you get.[14] The Dutch Breast Cancer Foundation reports that, in 2004, which is not within the period over which we observe variation in screening, prescription of Herceptin was as low as 6 percent in the province of Friesland, while it was as high as 86 percent in the province of Zeeland (Borstkanker Vereniging Nederland, 2005). Apart

[14] The Dutch government tried to solve this issue by passing a law in 2002 which obliged health insurers to contribute to the costs of some government determined expensive drugs. Until 2005, the contribution rate had to be negotiated with the insurers, differed by drug and varied between 0 and 75 percent. In 2006, it became fixed at 80 percent. However, this also did not fully address the problem because, given the price level for expensive drugs, even the remaining 20 percent is a significant burden hospitals' budget. Furthermore, the criteria to be listed as an expensive drug is hard to meet. As of March 2006, there were only 16 drugs (12 of which are cancer drugs) approved as expensive drugs.

from financial capacity, hospitals could differ with respect to the available technology, knowledge, awareness and physician preferences (Savage & Widener, 2008; Hershman et al., 2008).

We cannot use fixed effects to allow for differences in treatment quality and other potentially unobservable time invariant confounders at the municipality level but we can allow for province fixed effects. To do this we must drop municipalities in the three provinces in which screening was already fully implemented by the beginning of 1997. No women experienced a delay of 24 months or longer in those municipalities creating a lack of variation in delay in program entry. Column I of Table 2.6 presents the estimated effect of a 24 month or more delay on breast cancer mortality for this restricted sample with control for age only. The estimate differs very little from that obtained from applying the same specification – control for age only – to the full sample (Table 2.3, column I). Allowing for province fixed effects causes a moderate but statistically insignificant rise in the estimated program effect.

We previously established that there is no correlation between launch date of the program in a municipality and the rate of response to the screening invitation at the start of the program. But we did not test whether breast cancer mortality is related to the response, which it should be if screening has an effect, because the initial screening participation rate is not available for five municipalities. To our surprise, dropping these municipalities has a large impact on the estimated effect. It drops from 0.172 (Table 2.3, column I) to 0.122 (Table 2.6, column III) but remains significant at the 5 percent level and within the 95 percent confidence interval of the original estimate. Further checks have shown that two municipalities, namely Westerveld and Middenveld, are responsible for the reduction in the estimated effect. These municipalities started screening late, i.e. in January-February in 1997 and have a larger pre-program breast cancer mortality rate than the other municipalities. Adding the initial participation rate has little or no impact (Table 2.6, column IV) which supports our baseline specification. It also supports interpretation of the estimated mortality effect of the delay in first invitation as the effect of the delay in the receipt of first screening after correcting for rate of attendance to screening.

Restricting the sample to women who did not move between municipalities, over the period 1995-97, that we use to identify the effect results in a (not significant) reduction in the estimated effect from 0.172 to a still highly significant 0.148 (Table 2.6, column V). The direction of this change is the opposite to what one would expect if high risk women were moving into the municipalities where they could access the program earlier.

Lastly, we run a placebo test to check whether the program effect we estimate is arising simply by chance or because of some systematic difference in breast cancer mortality risks

across municipalities that entered the program at different times. To this end, we restrict the sample to women aged 72-77 in January 1995 and living in municipalities that started to offer screening to women aged 50-69 at some time after that date. The older group are too old to qualify for the program and so its breast cancer mortality rate should not be related with the start date of the program. This is confirmed by the results in column VI of Table 2.6, lending further support to interpretation of the estimated effect for the younger age group as truly indicative of the impact of delayed exposure to the screening program.

2.7 Discussion and Conclusion

In this paper, we have assessed the impact of the Dutch breast cancer screening program on breast cancer and all-cause mortality over a 17 year period. We estimate that a delay of 24-35 months in entry to the program led to 170 excess breast cancer deaths per 100,000 women in comparison with a delay of 0-23 months. This corresponds to a 10 percent increase in relative risk. We do not find any statistically significant difference in the overall mortality rates of women with shorter or longer delay. Our findings are strengthened by our large sample size, nationwide design, and long follow-up period.

It should be noted that we report intention to treat estimates. Given that participation rate to the program is around 80 percent, the estimate of the mortality impact of a delay in the receipt of first screening mortality will be larger.

Women who waited for at least 24 months to enter the program can be expected to receive, on average, one less invitation for screening in a lifetime than women who did not have to wait for so long. Kalager et al. (2010), evaluating the Norwegian breast cancer screening program, estimate 2.4 fewer breast cancer deaths per 100,000 person-years due to screening, although the estimate is not statistically different from zero. The authors say that they observe a reduction in mortality mainly in the first four years of the follow-up period, which implies that the estimated screening effect is mainly driven by 2 screens. We estimate 24 fewer deaths per 100,000 women or 5.8 fewer deaths per 100,000 person-years due to 1 additional invitation at the end of 4 years (i.e. 1998) which is also not statistically significant. Given the hypothesized concave relationship between the number of screens and the size of the impact on breast cancer mortality in the medical literature (Fielder et al., 2004; Collette et al., 1984), our estimate seems large in comparison to the Norwegian study. Tabár et al. (2011) estimate 133-159 more deaths from breast cancer per 100,000 women among women aged 50-74 years who were not invited to screening relative to those who received 4-5 invitations, after 15 years from entry in the Swedish Two-County Trial. This effect is considerably smaller than our estimate of the effect of

1 invitation at the end of 17 years, i.e. about 170 more breast cancer deaths per 100,000 women. This could be related with the fact that the screening practice (e.g., single-view versus double-view mammography) and frequency (once in every 24-33 months versus 48 months) are different between the Swedish Two-County Trial and the Dutch program. Otto et al. (2012) find that relative risk of breast cancer mortality is 44 percent lower among women who responded to the invitation from the Dutch breast cancer screening program that precedes their diagnosis. This estimate is considerably larger than our relative risk estimate of 10 percent because it is based on the screen has the highest potential to make a difference. Unfortunately this study does not have an estimate of the mortality effect in absolute terms due to its case-control design.

Our findings lend support to the effectiveness of organized mammography screening in reducing breast cancer mortality at a time when this has been called into question. But they are less encouraging with respect to the impact on overall mortality. The correct implication to be drawn from the discrepancy in the estimated effects on the two outcomes hinges crucially on the reasons for it. One possible mechanism could be that all-cause mortality includes fatalities arising from screening or treatment related causes that are not counted as breast cancer deaths. In that case, it would indeed be time to reconsider effectiveness of mammography screening. But it could be that the absence of any effect on all-cause mortality is simply attributable to breast cancer survival due to screening inevitably being followed by death from something else over a reasonably long follow-up. In this case, mammography screening does raise the length of life.

Finally we would like to mention that the mortality impact of screening is not the only aspect of the ongoing mammography debate. With changes in screening technology, harmful effects of screening have also emerged as a growing concern. False positive results and over-diagnosis are claimed to be major problems with screening. Over-diagnosis refers to detection of small tumors which would not have caused illness and/or death. Some tumors grow so slowly that they either never reach the clinically detectable phase or the patient dies of something else before they do so. Patients in this condition are harmed by screening because there is no survival benefit of treating those tumors, and treatment has its own risks. Mammograms increase the rate of over-diagnosis because their working principle is to detect small tumors, and all detected tumors are treated as there is no way of knowing which ones will grow into invasive disease. Miller et al. (2014) report that, 22 percent of the screen detected cancers in the mammography group were over-diagnosed in Canadian National Breast Cancer Screening Study. In this paper we provide no estimate of the degree of over-diagnosis or false positives in the Dutch breast cancer screening program. Despite its significance, this topic is left for future research.

Figures

Figure 2.1: Geographical spread of the Dutch breast cancer screening program

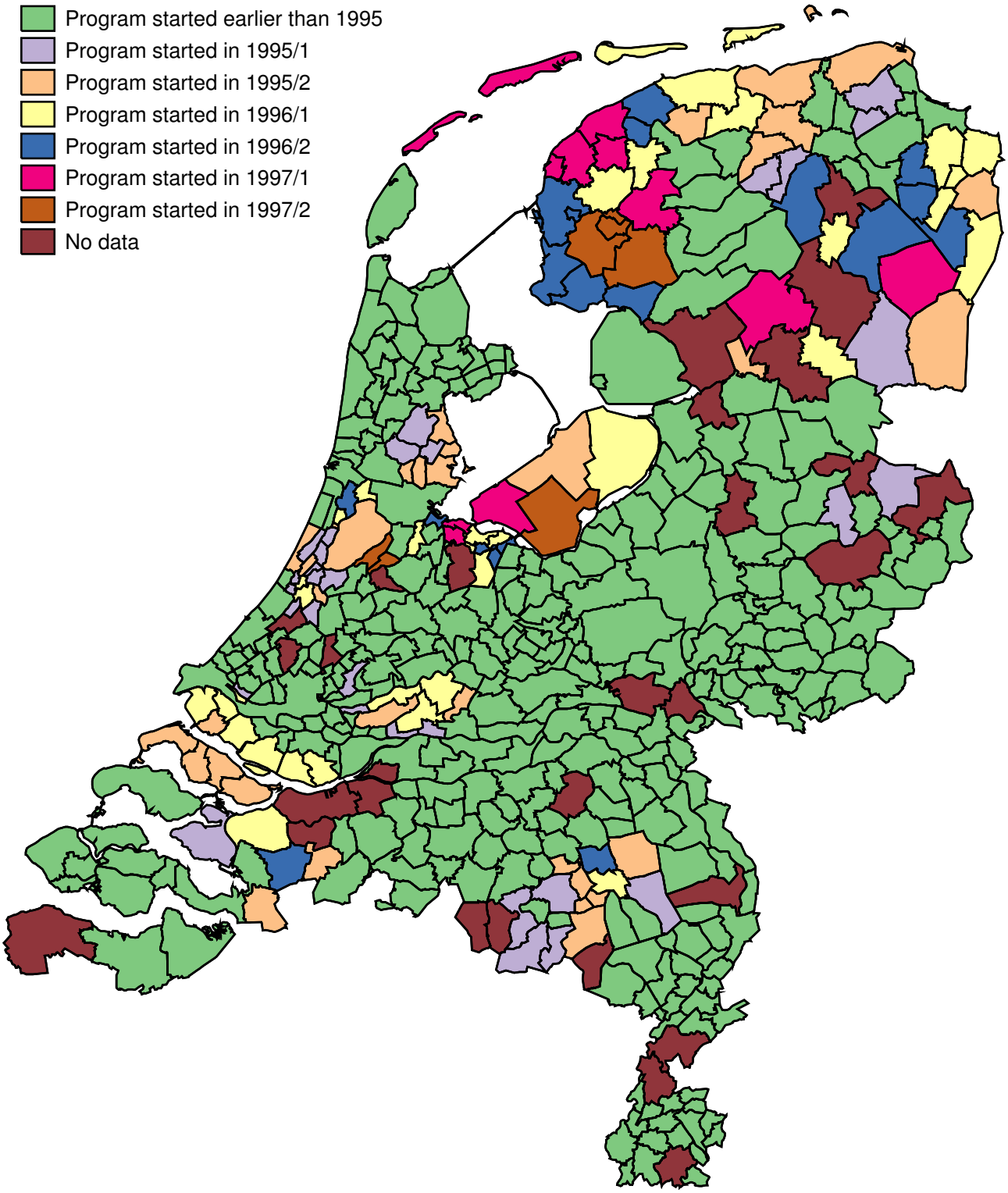


Figure 2.2: Launch of organized screening and pre-program breast cancer incidence

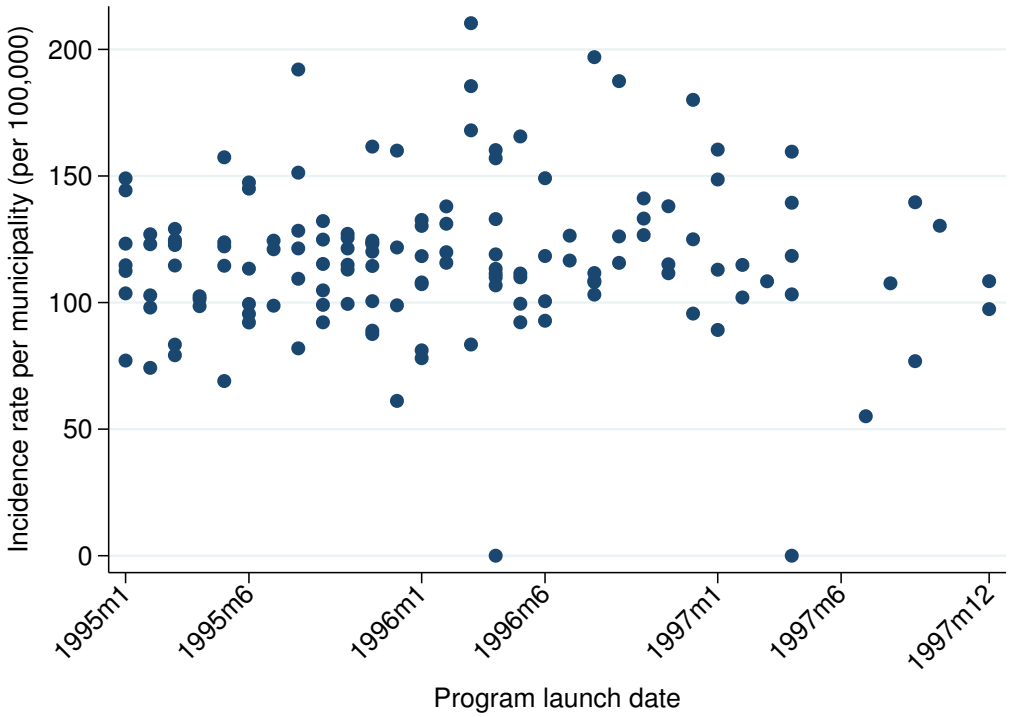


Figure 2.3: Launch of organized screening and pre-program breast cancer prevalence

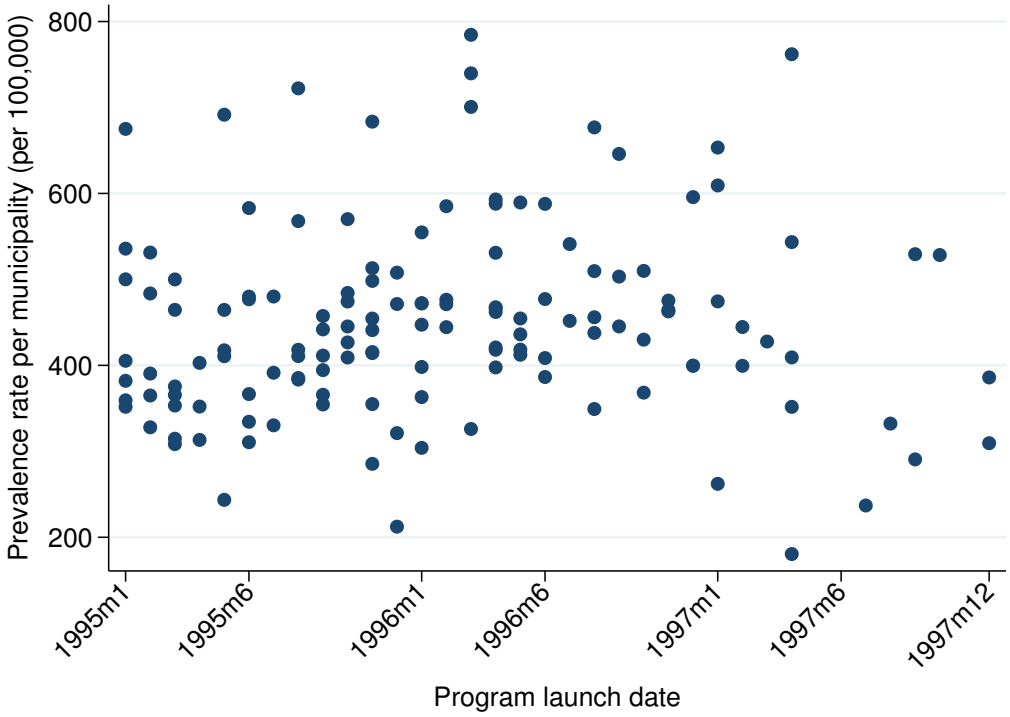


Figure 2.4: Launch of organized screening and participation to screening

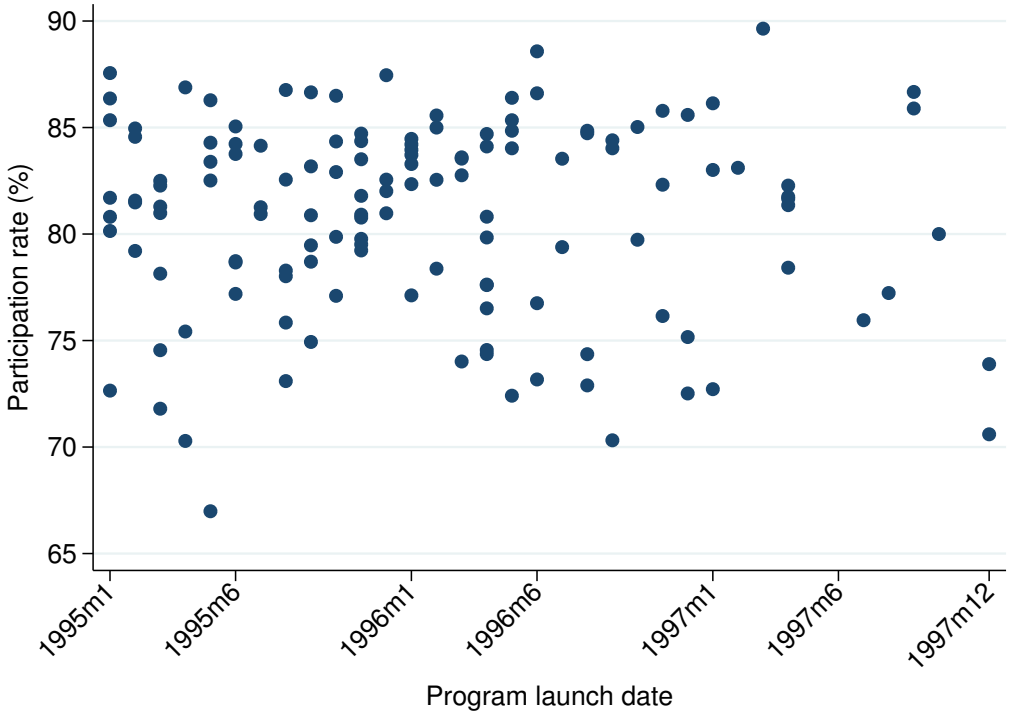
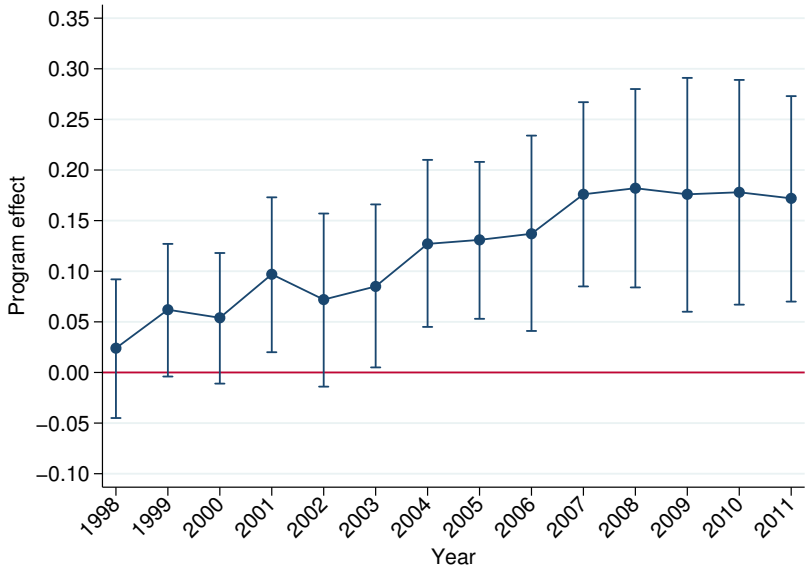


Figure 2.5: Evolution of the binary program effect over time



In each year, an indicator for having died from breast cancer between 1995 and the respective year is regressed on the binary program variable, i.e. entering the program with a delay of 24-35 months versus 0-23 months and on categorical age variable, i.e. 49-54, 55-59, 60-63 years in January 1995. Coefficient estimates for the program variable, expressed as percentage points, are plotted. Error bars show 95% confidence intervals.

Tables

Table 2.1: Sample characteristics

Individual characteristics[a]		
	Mean	Sample size
Age in January 1995	55.59	256,712
	(4.35)	
Months between Jan. 1995	12.62	256,712
and entry to the program	(8.15)	
Distribution of date of entry to the program		
	No. of municipalities	No. of individuals[b]
First half of 1995	33	61,350
Second half of 1995	34	60,384
First half of 1996	33	65,341
Second half of 1996	19	42,724
First half of 1997	12	17,591
Second half of 1997	7	9,322
Total	138	256,712

[a] Standard deviations are in parentheses.

[b] Living in those municipalities as of January 1995.

Table 2.2: Regression estimates of delay in launch of screening program of at least 24 months

	(I)	(II)	(III)	(IV)
Breast cancer statistics				
Breast cancer incidence (Incidence_m)	-0.002			-0.001
	(0.002)			(0.002)
Breast cancer prevalence (Prevalence_m)	0.000			0.000
	(0.001)			(0.001)
Program participation rate (Participation_m)		-0.002		-0.005
		(0.007)		(0.008)
Demographics				
Female pop.size (Population_m)			-0.002	-0.002
			(0.001)	(0.001)
Share of target pop. (Target_m)			-0.003	-0.003
			(0.014)	(0.014)
Constant	0.272*	0.270	0.211	0.756
	(0.155)	(0.576)	(0.265)	(0.781)
P-values on Joint Hypotheses				
$\text{Incidence}_m = \text{Prevalence}_m = 0$	0.472			0.553
$\text{Population}_m = \text{Target}_m = 0$			0.426	0.375
All coefficients				0.591
N	138	133	138	133
R-square	0.013	0.001	0.005	0.024

Notes: The dependent variable is a binary indicator of organized screening starting in a municipality in 1997 (1), as opposed to 1995-96 (0). Linear probability model estimates reported. Breast cancer incidence is the median breast cancer incidence during the period 1989-1994. Breast cancer prevalence is the 5-year invasive breast cancer prevalence as of January 1994. Program participation rate is the share of invited women who attended screening when the program was first introduced in a municipality. Female population size is expressed in 1000s. Share of target population is the percentage share of women aged 50-69 among the female population. Robust standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 2.3: Effect of the delayed program entry on cumulative breast cancer mortality to 2011 – Binary treatment

	(I)	(II)	(III)	(IV)
Delay in program entry				
0-23 months	base	base	base	base
24-35 months	0.172***	0.177***	0.187***	0.188***
	(0.051)	(0.053)	(0.052)	(0.054)
Age in January 1995				
49-54	base	base	base	base
55-59	0.173***	0.172***	0.172***	0.171***
	(0.055)	(0.055)	(0.055)	(0.055)
60-63	0.241***	0.240***	0.239***	0.239***
	(0.053)	(0.053)	(0.053)	(0.053)
Breast cancer statistics				
Breast cancer incidence (Incidence_m)		0.002		0.003
		(0.002)		(0.002)
Breast cancer prevalence (Prevalence_m)		0.000		0.000
		(0.000)		(0.000)
Demographics				
Female pop. size (Population_m)			0.001	0.001
			(0.001)	(0.001)
Share of target pop. (Target_m)			0.002	-0.004
			(0.008)	(0.010)
Constant	1.273***	1.136***	1.203***	1.166***
	(0.033)	(0.109)	(0.170)	(0.169)
P-values on Joint Hypotheses				
$\text{Incidence}_m = \text{Prevalence}_m = 0$		0.334		0.340
$\text{Population}_m = \text{Target}_m = 0$			0.582	0.460
No. of municipalities	138	138	138	138
No. of women	256,712	256,712	256,712	256,712

Notes: The dependent variable is an indicator for having died from breast cancer between 1995 and 2011. Coefficient estimates are expressed as percentage points, except the constant which is expressed as a percentage. See Table 2.2 for definitions of covariates. Standard errors are clustered over women who were living in the same municipality at the time of the first invitation, and reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 2.4: Effect of the delayed program entry on cumulative breast cancer mortality to 2011 – Categorical treatment

Delay in program entry	
0-5 months	base
6-11 months	0.034
	(0.070)
12-17 months	-0.044
	(0.059)
18-23 months	0.009
	(0.075)
24-29 months	0.162**
	(0.072)
30-35 months	0.185**
	(0.077)
Age in January 1995	
49-54	base
55-59	0.173***
	(0.055)
60-63	0.242***
	(0.053)
Constant	1.274***
	(0.050)
No. of municipalities	138
No. of women	256,712

Notes: The dependent variable is an indicator for having died from breast cancer between 1995 and 2011. Coefficient estimates are expressed as percentage points, except the constant which is expressed as a percentage. Standard errors are clustered over women who were living in the same municipality at the time of the first invitation, and reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 2.5: Effect of the delayed program entry on overall cumulative mortality to 2011 – Binary treatment

	(I)	(II)	(III)	(IV)
Delay in program entry				
0-23 months	base	base	base	base
24-35 months	0.006	-0.013	0.329	0.356
	(0.596)	(0.547)	(0.382)	(0.420)
Age in January 1995				
49-54	base	base	base	base
55-59	5.782***	5.778***	5.762***	5.752***
	(0.165)	(0.165)	(0.167)	(0.166)
60-63	14.211***	14.208***	14.165***	14.152***
	(0.240)	(0.240)	(0.243)	(0.240)
Breast cancer statistics				
Breast cancer incidence		0.017	0.024*	0.022*
		(0.014)	(0.013)	(0.012)
Breast cancer prevalence		-0.004	-0.006*	-0.006*
		(0.004)	(0.003)	(0.003)
Demographics				
Female pop. size			0.036***	0.035***
			(0.005)	(0.006)
Share of target pop.			-0.094	
			(0.084)	
Constant	9.678***	9.399***	10.143***	8.900***
	(0.205)	(0.851)	(1.360)	(0.676)
No. of municipalities	138	138	138	138
No. of women	256,712	256,712	256,712	256,712

Notes: The dependent variable is an indicator for having died from any cause between 1995 and 2011. Coefficient estimates are expressed as percentage points, except the constant which is expressed as a percentage. See Table 2.2 for definitions of covariates. Standard errors are clustered over women who were living in the same municipality at the time of the first invitation, and reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 2.6: Effect of delayed program entry (binary) on cumulate breast cancer mortality to 2011 – Robustness checks

	Province fixed effects[a]		Participation rate[b]		Non-movers[c]	Elderly women[d]
	(I)	(II)	(III)	(IV)	(V)	(VI)
Delay in program entry						
0-23 months	base	base	base	base	base	base
24-35 months	0.168***	0.191***	0.122**	0.128**	0.148***	-0.044
	(0.059)	(0.055)	(0.051)	(0.055)	(0.050)	(0.214)
Included controls	Age in January 1995	Age in January 1995 Province fixed effects	Age in January 1995	Age in January 1995 Participation rate	Age in January 1995	Age in January 1995
No. of municipalities	65	65	133	133	138	138
No. of women	133,626	133,626	246,253	246,253	247,797	62,176

[a] The sample includes women living in municipalities in Utrecht/Flevoland, Noord Holland, Drenthe/Overijssel and Friesland.
[b] The sample includes women living in municipalities for which we have data on the rate of participation to screening at the first arrival of the program in a municipality.
[c] The sample includes women who did not move during 1995-1997.
[d] The sample includes women who were aged 72-77 in January 1995.
The dependent variable is an indicator for having died from breast cancer between 1995 and 2011. Coefficient estimates are expressed as percentage points. Standard errors are clustered over women who were living in the same municipality at the time of the first invitation, and reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Appendix 2.A: Case-Control Sampling and Attributable Risk

The discussion in this appendix is based on Manski (2009).

Consider a population whose members are defined by a set of covariates x and an outcome y. The practice of case-control sampling divides the population into sub-populations on the basis of the possible values of the outcome. Consequently, for each $j \in Y$, $P(x|y=j)$ can be calculated. However, given the ultimate goal is to find $P(y|x)$, the inferential question arises: What does knowledge of $P(x|y=j)$ tell about $P(y|x)$?

To study this question, let's focus on the simple case of a binary outcome: $y=1$ indicates mortality from breast cancer, and 0 otherwise. Let $x=(w,r)$ where w contains some covariates like age, socioeconomic status, race; and r contains risk factors for breast cancer mortality. Assume r takes only two values: k and j (e.g., getting screened or not). A relevant research question is to find the probability of breast cancer death for the two different realizations of the risk factor: $P(y=1|w,r=k)$ versus $P(y=1|w,r=j)$.

This question can be answered via calculating the relative risk (RR)

$$RR = \frac{P(y=1|w,r=k)}{P(y=1|w,r=j)} \tag{2.3}$$

and/or attributable risk (AR)

$$AR = P(y=1|w,r=k) - P(y=1|w,r=j) \tag{2.4}$$

Relative risk is the ratio of the probability of breast cancer death conditional on getting screened to the probability of breast cancer death conditional on not getting screened. Attributable risk is the difference in the probability of breast cancer death conditional on getting screened and not.

Case-control sampling reveals the conditional distributions $P(w,r|y=1)$ and $P(w,r|y=0)$ but not the distribution $P(y=1|w,r)$. This can be seen by using Bayes' Theorem and the Law of Total Probability

$$\begin{aligned} P(y=1|w,r) &= \frac{P(w,r|y=1)P(y=1)}{P(w,r)} \\ &= \frac{P(w,r|y=1)P(y=1)}{P(w,r|y=1)P(y=1)+P(w,r|y=0)P(y=0)} \end{aligned} \tag{2.5}$$

While case-control sampling is informative about $P(w,r|y=1)$ and $P(w,r|y=0)$, it is silent about $P(y=1)$. To solve this problem, epidemiologists resort to the *rare-disease*

assumption which simply states that prevalence of the outcome (disease) in the population is low. Formally, this means that $P(y = 1|w)$ approaches zero. Under this assumption, both relative and attributable risk can be identified. To see this rewrite equation 2.5

$$\begin{aligned} P(y = 1|w, r) &= \frac{P(r|w, y = 1)P(y = 1|w)}{P(r|w)} \\ &= \frac{P(r|w, y = 1)P(y = 1|w)}{P(r|w, y = 1)P(y = 1|w) + P(r|w, y = 0)P(y = 0|w)} \end{aligned} \tag{2.6}$$

and insert it into the definitions of relative and attributable risk

$$\begin{aligned} RR = &\frac{P(r = k|w, y = 1)}{P(r = j|w, y = 1)} \\ &\times \frac{P(r = j|w, y = 1)P(y = 1|w) + P(r = j|w, y = 0)P(y = 0|w)}{P(r = k|w, y = 1)P(y = 1|w) + P(r = k|w, y = 0)P(y = 0|w)} \end{aligned} \tag{2.7}$$

$$\begin{aligned} AR = &\frac{P(r = k|w, y = 1)P(y = 1|w)}{P(r = k|w, y = 1)P(y = 1|w) + P(r = k|w, y = 0)P(y = 0|w)} \\ &- \frac{P(r = j|w, y = 1)P(y = 1|w)}{P(r = j|w, y = 1)P(y = 1|w) + P(r = j|w, y = 0)P(y = 0|w)} \end{aligned} \tag{2.8}$$

Letting $P(y = 1|w)$ approach zero gives

$$\lim_{P(y=1|w)\to 0} RR = \frac{P(r = k|w, y = 1)}{P(r = j|w, y = 1)} \times \frac{P(r = j|w, y = 0)}{P(r = k|w, y = 0)} \tag{2.9}$$

$$\lim_{P(y=1|w)\to 0} AR = 0 \tag{2.10}$$

Cornfield (1951) shows that equation 2.9 is the relative risk under the rare disease assumption. The right-hand side of the equation is called the odds ratio and thanks to equation 2.6 can be written as,

$$\begin{aligned} OR &= \frac{P(r = k|w, y = 1)}{P(r = j|w, y = 1)} \times \frac{P(r = j|w, y = 0)}{P(r = k|w, y = 0)} \\ &\equiv \frac{P(y = 1|w, r = k)}{P(y = 0|w, r = k)} \times \frac{P(y = 0|w, r = j)}{P(y = 1|w, r = j)} \end{aligned} \tag{2.11}$$

The above given derivation reveals two important points:

- Case-control sampling leads to calculation of the odds ratio. Under the rare-disease assumption, it is possible interpret the odds ratio as relative risk which is a function

of the unknown distributions $P(y = 1|w, r = k)$ and $P(y = 1|w, r = j)$.

- Case-control sampling is incapable of providing an estimate of the attributable risk.

Appendix 2.B: Additional Tables

Table 2.B.1: Dutch breast cancer screening program 1990-2011

	1990-1997	1998-2006	2011
Target population per year (x1000)	733-813	1,021-1,164	1,275
Overall attendance (%)	78.2	80.7	80.1
Breast cancer mortality[a]	91.6	76.7	64.6

[a] European standardized rate per 100,000 women for age category 50-74

Source: National Evaluation Team for Breast Cancer Screening (2014)

Table 2.B.2: Regression estimates of delay in launch of screening program measured in categories

	(I)	(II)	(III)	(IV)
Breast cancer statistics				
Breast cancer incidence (Incidence_m)	-0.004			-0.004
	(0.004)			(0.005)
Breast cancer prevalence (Prevalence_m)	0.002			0.002
	(0.001)			(0.001)
Program participation rate (Participation_m)		-0.008		-0.010
		(0.023)		(0.026)
Demographics				
Female pop. size (Population_m)			0.000	-0.000
			(0.007)	(0.007)
Share of target pop. (Target_m)			0.041	0.021
			(0.038)	(0.043)
P-values on Joint Hypotheses				
$\text{Incidence}_m = \text{Prevalence}_m = 0$	0.439			0.532
$\text{Population}_m = \text{Target}_m = 0$			0.564	0.872
All coefficients				0.855
N	138	133	138	133
Pseudo R-sq	0.004	0.000	0.003	0.004

Notes: Ordered probit coefficients are reported. Dependent variable is the start date of organized screening in a municipality, measured in half year intervals between 1995 and 1997. See Table 2.2 for definitions of covariates. Standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 2.B.3: Effect of the delayed program entry on cumulative breast cancer mortality to 2011 – Categorical treatment

	(I)	(II)	(III)
Delay in program entry			
0-5 months	base	base	base
6-11 months	0.031	0.052	0.042
	(0.069)	(0.074)	(0.074)
12-17 months	-0.059	-0.033	-0.051
	(0.063)	(0.066)	(0.068)
18-23 months	-0.015	0.010	-0.010
	(0.079)	(0.075)	(0.077)
24-29 months	0.162**	0.185**	0.175**
	(0.073)	(0.078)	(0.081)
30-35 months	0.184**	0.213**	0.210***
	(0.072)	(0.083)	(0.077)
Age in January 1995			
49-54	base	base	base
55-59	0.172***	0.172***	0.172***
	(0.055)	(0.055)	(0.055)
60-63	0.241***	0.240***	0.239***
	(0.053)	(0.053)	(0.053)
Breast cancer statistics			
Breast cancer incidence	0.002		0.002
	(0.002)		(0.002)
Breast cancer prevalence	0.000		0.000
	(0.000)		(0.000)
Demographics			
Female pop. size		0.001	0.001
		(0.001)	(0.001)
Share of target pop.		0.003	-0.004
		(0.009)	(0.010)
Constant	1.110***	1.182***	1.144***
	(0.110)	(0.178)	(0.179)
No. of municipalities	138	138	138
No. of women	256,712	256,712	256,712

Notes: The dependent variable is an indicator for having died from breast cancer between 1995 and 2011. Coefficient estimates are expressed as percentage points, except the constant which is expressed as a percentage. See Table 2.2 for definitions of covariates. Standard errors are clustered over women who were living in the same municipality at the time of the first invitation, and reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 2.B.4: Effect of the delayed program entry on cumulative breast cancer mortality to 2011 by age

24-35 months delay in program entry x Age:49-54	0.151
	(0.134)
24-35 months delay in program entry x Age:55-59	0.308*
	(0.185)
24-35 months delay in program entry x Age:60-63	0.031
	(0.157)
Age in January 1995	
49-54	base
55-59	0.156***
	(0.050)
60-63	0.254***
	(0.055)
Constant	1.275***
	(0.032)
No. of municipalities	138
No. of women	256,712

Notes: The dependent variable is an indicator for having died from breast cancer between 1995 and 2011. Coefficient estimates are expressed as percentage points, except the constant which is expressed as a percentage. Standard errors are clustered over women who were living in the same municipality at the time of the first invitation, and reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Table 2.B.5: Evolution of the binary program effect over time

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Delay in program entry														
0-23 months	base	base	base	base	base	base	base	base	base	base	base	base	base	base
24-35 months	0.024	0.062*	0.054	0.097**	0.072*	0.085**	0.127***	0.131***	0.137***	0.176***	0.182***	0.176***	0.178***	0.172***
	(0.035)	0.033	0.032	0.039	0.043	0.041	0.042	0.039	0.049	0.046	0.049	0.058	0.056	0.051
Constant	0.291***	0.368***	0.433***	0.518***	0.584***	0.660***	0.742***	0.815***	0.886***	0.963***	1.045***	1.121***	1.196***	1.273***
	(0.015)	(0.018)	(0.019)	(0.022)	(0.023)	(0.023)	(0.026)	(0.027)	(0.029)	(0.03)	(0.032)	(0.033)	(0.032)	(0.033)
No. of municipalities	138	138	138	138	138	138	138	138	138	138	138	138	138	138
No. of women	256,712	256,712	256,712	256,712	256,712	256,712	256,712	256,712	256,712	256,712	256,712	256,712	256,712	256,712

Notes: In each column, the dependent variable is an indicator for having died from breast cancer between 1995 and the respective year given in the column. All models control for age in January 1995 in three categories: 49-54, 55-59, 60-63. Coefficient estimates are expressed as percentage points, except the constant which is expressed as a percentage. Standard errors are clustered over women who were living in the same municipality at the time of the first invitation, and reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01

Chapter 3

What Explains the Education Disparity in Screening Mammography?

Less educated women in the United States are substantially less likely to receive screening mammography. It is not clear whether this is due to differences in access to screening or in perceptions of breast cancer risks and the effectiveness of screening. There is no education gradient in the Netherlands, where there is universal access to screening mammography. Using cross-sectional and cross-country comparable individual level data, we found that in the absence of a universal screening program in the U.S., determinants of access – income, insurance coverage and receipt of medical advice – appear to drive the education disparities in screening mammography. The Affordable Care Act requirement that mammography be covered by new health plans and Medicare may therefore be effective in reducing the gradient.

This chapter is based upon:

Koç, H. and O'Donnell, O. and Van Ourti, T. (2013). *What Explains the Education Disparity in Screening Mammography? A Comparison of the United States with the Netherlands.* (Mimeo). Erasmus University Rotterdam.

3.1 Introduction

Breast cancer is the most commonly diagnosed cancer among women worldwide, accounting for 23 percent of all new cancer cases (American Cancer Society, 2011). It is also the leading cause of cancer-attributable mortality of females around the globe being responsible for 14 percent of such deaths. Breast cancer is curable if detected sufficiently early and treated appropriately (World Health Organization, 2011). Screening mammography – a mammogram taken when there is no sign of breast cancer – is the recommended early detection tool for breast cancer due to its ability to detect pre-cancerous cells while they are still treatable (U.S. Preventive Services Task Force, 2009).

Utilization of screening mammography increased steeply in the last two decades of the twentieth century (Bleyer & Welch, 2012). But there remain large differences in uptake by education: In the United States (e.g. Lange, 2011; Picone et al., 2004) and a number of European countries (e.g. Avitabile et al., 2011; Jusot et al., 2012; Palència et al., 2010), lower educated women tend to receive mammograms less often than their higher educated counterparts. This may reflect higher barriers to access faced by less educated women (Palència et al., 2010; Walsh et al., 2011). But it could also arise from differences in perception of the breast cancer base rate, the incorporation of information available from objective risk factors and beliefs about the effectiveness of screening (Lange, 2011; Palència et al., 2010; Walsh et al., 2011; Carman & Kooreman, 2014).

We use comparable data that include information on subjective perceptions of breast cancer risk and screening effectiveness to estimate and explain the education gradient in screening mammography in the United States and the Netherlands. Comparison of these two countries is instructive because they differ in access to screening. The Netherlands operates a universal, fully publicly subsidized screening program covering all women aged 50-75. Mammograms are delivered in mobile screening units to maximize accessibility (Holland et al., 2007; Rijksinstituut voor Volksgezondheid en Milieu, 2015). There is no such program in the U.S, but the U.S. Preventive Services Task Force (USPSTF) recommends biennial screening of women aged 50-74 (U.S. Preventive Services Task Force, 2009). Prior to the compulsory coverage of mammography by new health plans and Medicare following the 2010 Affordable Care Act (ACA), insured, as well as uninsured, American women were likely to pay out-of-pocket for a mammogram (U.S. Centers for Medicare and Medicaid Services, 2011).

The absence of organized, subsidized screening in the United States leaves scope for the observed socioeconomic difference in mammography uptake (Lange, 2011; Picone et al., 2004). If this is mainly driven by financial barriers, then the coverage extension brought about by the ACA would be expected to reduce the gradient. If, on the other

hand, the education gradient is attributable mainly to perceptions of breast cancer risk and of the effectiveness of mammography, then health promotion programs that raise awareness of risks among less educated women will be required to even out the gradient. We aim to weigh the plausibility of these two scenarios by estimating the dependence of mammography on access and perceptions, as well as education, in both the U.S. and the Netherlands.

3.2 Methods

3.2.1 Data sources

Data from the American Life Panel (ALP) and the Dutch Longitudinal Internet Studies for the Social Sciences (LISS) were analyzed (CentER Data, 2011; RAND Corporation, 2011). Both surveys were administered over the internet following a similar protocol, and contained modules that include detailed individual-level information on breast cancer screening. Female respondents aged 40+, the earliest age recommended for screening (U.S. Preventive Services Task Force, 2009; American Cancer Society, 2012b), were extracted from the breast cancer module of the ALP (n = 646), which was fielded from mid-December 2011 until early January 2012, and from the disease prevention module of LISS (n = 1490), which was fielded in September 2008.

3.2.2 Variables

Dependent variables

The outcome of interest was receipt of a mammogram in the past 2 years. In LISS, the question was: "Have you had a mammogram in the last 2 years?". In ALP, respondents were asked "Have you ever had a screening mammogram", and if the answer was yes, "When did you have your most recent screening mammogram?". Based on these questions, the binary variable indicating mammogram usage in the last 2 years was constructed. The ALP explicitly asked about receipt of a screening mammogram, while no differentiation was made between receipt of a screening and a diagnostic mammogram in LISS, where the latter is conventionally defined as a test provoked by some indication of the possible presence of breast cancer. Therefore, Dutch women who reported previously having been diagnosed with breast cancer were dropped from the study sample.

Independent variables

Educational attainment was measured by the highest level completed with a diploma. To ensure comparability between the U.S. and the Netherlands, categories were constructed based on ISCED mappings of qualifications (Unesco Institute for Statistics, 2011), distinguishing between low (upper secondary, i.e. high school, education or less, ISCED<4), middle (post-secondary non-tertiary education, ISCED=4), and high education (higher education, ISCED>4).

Some additional independent variables were common for the U.S. and the Netherlands, while others were country-specific. The set of common covariates included age – controlled with 5-year intervals in the Netherlands and as 40-50, 50-65, 65-75 & 75+ in the U.S. to avoid small cell sizes (e.g. Avitabile et al., 2011; Palència et al., 2010), household income (e.g. Avitabile et al., 2011; Lairson et al., 2005), race/ethnic origin (e.g. Sambamoorthi & McAlpine, 2003; Selvin & Brett, 2003), objective risk factors for breast cancer – the number of first degree relatives with breast cancer for both the U.S. and the Netherlands, and whether a women has given birth only for the Netherlands (e.g. American Cancer Society, 2012a), and perceptions of breast cancer risk (Lange, 2011; Carman & Kooreman, 2014) and of mammogram effectiveness in reducing mortality from breast cancer. Risk perception was measured by the reported probability of getting breast cancer within 5 years. Perceived effectiveness of mammography in reducing the risk of death from breast cancer was measured on a four and five point categorical scale for the U.S. and the Netherlands, respectively. Household income was measured before taxes for both countries and entered into the models as quartile group indicators.

Covariates included in the models estimated with U.S. data only were: i) whether the respondent had full insurance (versus partial or no) coverage for a mammogram (e.g. Meissner et al., 2007; Sambamoorthi & McAlpine, 2003; Selvin & Brett, 2003), ii) whether a mammogram had been recommended by a health care provider (e.g. Meissner et al., 2007; O'Malley et al., 2001), and iii) perceived probabilities of a mammogram giving a) a false positive, b) a false negative, and c) a true positive. For the Netherlands, an indicator of having received an invitation for a mammogram from the national breast cancer screening program (Carman & Kooreman, 2014) was included, as was having been invited in some other manner, which should mainly correspond to referrals given that Dutch women only very rarely receive a mammogram in a private clinic with no referral. Reporting having had a friend who died from breast cancer, which may influence risk perceptions, was also included in some models estimated for the Netherlands.

3.2.3 Statistical analysis

The education gradient in mammography in the U.S. and the Netherlands was estimated by several logit models with education entered as a categorical variable. The first model estimated included only age, in addition to education, to obtain the age-standardized education gradient. Additional covariates were cumulatively added to reveal the extent to which the association between education and screening was related to these variables. The second model added the objective risk factors, race/ethnic origin, and subjective perceptions of breast cancer risk and mammogram effectiveness at detecting breast cancer and preventing death from it. The third model added income, insurance coverage and whether a health care provider had recommended getting a screening mammography in the case of the United States. For the Netherlands, whether an invitation for mammography had been received from the national screening program or through referral were added at this stage.

From the logit model estimates, the difference between the probability of receiving a mammogram at a given education level and at the reference level was calculated for each observation and averaged across the sample. These average marginal effects (or adjusted risk differences) on the probability of getting a mammogram were also calculated for the other covariates. Sample weights were applied throughout.

3.3 Results

3.3.1 Descriptive statistics

Table 3.1 reports the rate of mammogram receipt in each country for women within the recommended age range for screening (50-75) and those outside of this range. At target ages, mammography prevalence in the previous two years was high (>80 percent) in both countries. At other ages, screening was still rather high in the United States, but relatively uncommon in the Netherlands. In the United States, three-quarters of women aged 40+ were recommended by a health care provider to get a mammogram and the same fraction had insurance that fully covered the costs of regular mammograms. Neither rate varied between women within and outside the recommended age range for screening. Less than 10 percent of Dutch women in the eligible age range reported not having been invited for screening by the national program.

Table 3.2 presents relative frequencies of the control variables by country. Around half of the American and Dutch women had no more than upper secondary (high school) education. American women had a higher expectation of getting breast cancer in the coming 5 years as well as a stronger belief in the effectiveness of a mammogram in

preventing death from breast cancer.

3.3.2 Education disparities in screening mammography

Table 3.3 presents the age-adjusted differences by education in the probability of receiving a mammogram. There was a pronounced, significant gradient in the United States: Women with low education were 11.5 percentage points less likely to be screened for breast cancer than their counterparts with higher education (p-value=0.040). The middle education group is 10.4 points less likely to get a screening mammogram than the most educated (p-value=0.056). There is no significant age-standardized education gradient in mammography use in the Netherlands (Wald test joint significance p-value=0.397).

3.3.3 Accounting for education disparities

The first two rows of Table 3.4 give the estimated education gradient in screening uptake after controlling for covariates in addition to age. Controlling for objective risk factors, race and perceptions of the risk of breast cancer and the effectiveness of mammography in averting death from this condition had little effect on the estimated education gradient in the United States (column 1). The difference in the screening probability between the most and least educated women narrows slightly from 11.5 (Table 3.3) to 9.7 percentage points but remains significant (p-value=0.061). The estimated difference between the top two education groups actually widens from 10.5 to 12.5 points (p-value=0.020). This persistence of the education gradient is not the result of differences in objective risk factors offsetting differences in risk perceptions. Controlling for age and objective risk factors only, the education gradient remains similar: Low and middle educated women are estimated to have a 11.1 (p-value=0.043) and 11.3 (p-value=0.034) percentage point lower probability to be screened compared to higher educated women (not shown in Table 3.4).

Adding income, insurance coverage and medical advice to get screened (column 2) has a much larger impact on the education gradient in the U.S., which decreases in magnitude and becomes statistically insignificant (Wald test joint significance p-value=0.653). Insurance coverage and medical advice are the strongest correlates of screening uptake. American women with insurance that fully covered mammography costs were 22.5 percentage points more likely to be screened than women with partial or no coverage; and those recommended by a medic to get screened were 38.4 percentage points more likely to do so.

Controlling for objective risk factors, ethnicity and perceptions of risk and the effectiveness of screening did not change the conclusion that there is no education gradient in

the age-standardized rate of mammography in the Netherlands (Wald test joint significance p-value=0.580) (Table 3.4, column 3). After controlling for income, being invited for screening from the national program or by referral, the least educated Dutch women had a statistically significant 3.8 percentage point lower probability of getting a mammogram than the most educated women (p-value=0.044). A program invitation itself raised the probability of getting a mammogram by 69 percentage points, while referral raised it by 42 points.

3.4 Discussion

There is a clear education disparity in age-standardized rates of mammography in the United States. Controlling for differences in perceptions of breast cancer risk and of screening effectiveness did not affect this gradient, while conditioning on income, insurance coverage and receipt of medical advice markedly weakened its magnitude and rendered it insignificant. This suggests that lower educated American women were less likely to be screened because they were poorer, had less comprehensive insurance cover and, perhaps because of that, were less likely to come into contact with a physician who recommended mammography, and not because they perceived less benefit from screening. The importance of financial barriers to mammography is evident from the fact that the education gradient becomes insignificant (Wald test joint significance p-value=0.268) after controlling only for income, in addition to age (estimates not shown in Table 3.4). Less educated, poorer women do not have the insurance cover and access to medical advice that are the main determinants of mammography screening in the United States.

In the Netherlands, a moderate education gradient in mammography emerges only after controlling for differences in the propensity to be invited for screening through the national program or by referral. Further analysis revealed that this education disparity existed among women who were not invited for screening by the national program. Among such women, who are uninvited mostly because they are outside the 50-75 age range, the least educated were 5.6 percentage points (p-value<0.027) less likely than the most educated to get a mammogram. Given the lower incidence of life-threatening breast cancer in this age range and risks arising from false positive screens, it is not obvious that this disparity is to the advantage of the more educated.

3.4.1 Role of health insurance and access to screening

In the U.S., the strongest correlates of screening uptake were medical advice and insurance coverage (Table 3.4, column 2). In the Netherlands, it was an invitation for screening

(Table 3.4, column 4). These determinants are largely responsible for the differences between the two countries in the distribution of breast cancer screening. Table 3.5, column 1 shows that, in the U.S., the probability of receiving medical advice to get a mammogram did not vary by age (Wald test joint significance p-value=0.981) and by income (Wald test joint significance p-value=0.949). It varied (only marginally) with objective risk factors (Wald test joint significance p-value=0.218) and by education (Wald test joint significance p-value=0.161) and, most strongly, by insurance coverage (p-value=0.001). In turn, comprehensive insurance coverage of mammogram was determined by age (Wald test joint significance p-value=0.006), due to Medicare qualification at 65, but also by income (Wald test joint significance p-value=0.030) and education (Wald test joint significance p-value=0.138), but not by objective risk factors (joint significance p-value=0.289) (column 2).

Higher educated and better-off American women are more likely to be insured, which is associated with a greater likelihood of being advised to undertake a mammogram, perhaps because doctors are more likely to recommend screening to patients they believe can afford it (O'Malley et al., 2001; Urban et al., 1994). Better-off women may also be more likely to consult with doctors known to recommend screening, or may be more successful in asking for a referral.

In the Netherlands, invitation for screening, which is the dominant determinant of mammography receipt, was almost exclusively determined by age and objective risk factors, and not by education or income (Table 3.5, columns 3 and 4). Differences in the role of financial barriers and in access to medical advice seem important in explaining the strong socioeconomic gradient in breast cancer screening in the U.S. and its absence in the Netherlands.

3.4.2 Strengths and weaknesses

This study is strengthened by the availability of comparable data on perceptions of breast cancer risk and mammogram effectiveness allowing consideration of the hypothesis that education disparities in screening behavior reflect differences in ability to accumulate and process information.

A weakness of the study is that while the sample sizes are sufficient for the main analyses, the U.S. sample does limit the extent of disaggregated analysis that is feasible and, together with the categorical nature of the income information, is partly responsible for the insignificance of income despite its importance in reducing the education gradient.

While insurance coverage is identified as an important correlate of breast cancer screening in the U.S., it is not possible to conclude from the analysis that insurance raises the likelihood of receiving a mammogram. It could be that women who want to be

screened are more likely to purchase an insurance plan that covers it.

Control for some objective risks increases the likelihood that the education gradient in insurance coverage in the U.S. implies differential financial barriers to screening for equal risk. But this is not guaranteed because of lack of control for risk factors, such as nulliparity and later age at first birth, which are more common among higher educated women (Heck & Pamuk, 1997; Menvielle et al., 2011). Nevertheless, it seems unlikely this is driving the education gradient in insurance since age and the number of relatives dying of breast cancer are much more important risk factors.

3.5 Conclusions

There is a clear socioeconomic gradient in breast cancer screening in the U.S. that appears to reflect differences in financial barriers and insurance coverage. Differences in perceptions of breast cancer risks and screening effectiveness appear to be less important in explaining education disparities. This, together with absence of a gradient in the Netherlands, where there is a universal breast cancer screening program and risk perceptions are also unimportant in explaining screening uptake, suggests that access, rather than information may be the more important determinant of the distribution of mammography.

Tables

Table 3.1: Rate of screening mammograms by country and age

	United States		Netherlands	
	50–75 (recommended screening age)	40–50, 75+ (not recommended screening age)	50–75 (recommended screening age)	40–50, 75+ (not recommended screening age)
Mammogram:				
in the last 2 years	80.82	69.38	84.90	16.39
recommended by doctor	76.00	74.60	n.a.	n.a.
covered by insurance	75.95	74.80	n.a.	n.a.
invite from screening program	n.a.	n.a.	91.41[a]	5.03[b]
invite in other manner	n.a.	n.a.	3.60	8.57
N	501	145	861	537

[a] The remaining 8.59 percent includes women who left the program by their own choice and refused to receive further invitations (National Evaluation Team for Breast Cancer Screening, 2009), and women who had recently entered the eligible age range at the time of interview but had not yet received an invitation for screening.

[b] Women aged 75-77 might have been screened during the last two years while they were still below 75 and received an invitation from the national program.

All numbers are stated as percentages.

Table 3.2: Means of control variables by country, women aged 40+

	United States	Netherlands
Education		
Low	47.00	53.72
Middle	20.54	19.60
High	32.46	26.68
Age group		
40-44 (U.S. 40-49)[a]	30.07	16.74
45-49	–	18.60
50-54 (U.S. 50-64)	42.00	16.52
55-59	–	16.24
60-64	–	13.52
65-69 (U.S. 65-75)	20.24	9.80
70-75	–	5.51
75+	7.69	3.08
Gross household income[b]		
Low	30.06	25.75
Low-middle	22.87	24.25
High-middle	27.98	25.04
High	19.10	24.96
Race/Ethnic origin		
White	77.11	n.a.
Non-white	22.89	n.a.
Dutch	n.a.	90.70
Breast cancer in the family		
No	85.03	87.98
Yes – 1 family member	11.46	10.80
Yes – 1+ family member	3.51	1.22
Has given birth	n.a.	84.26
Has a friend who died of breast cancer	n.a.	48.71
Perceived risk of getting breast cancer in the next 5 years	24.64	20.28
Perceived effectiveness of mammogram at preventing death		
Negative	n.a.	12.45
Low	20.66	24.96
Medium-low	22.34	23.03
Medium-high	32.05	21.67
High	24.94	17.88
Perceived probability of mammogram:		
giving a false-positive result	32.78	n.a.
giving a false-negative result	13.40	n.a.
detecting breast cancer	62.22	n.a.
N	646	1,398

[a] Wider age intervals are used for the U.S. because of the smaller sample size.

[b] The income quartiles constructed from the U.S. data did not divide the sample into four equally sized groups since they were derived from a categorical income variable.

All numbers are stated as percentages.

Table 3.3: Age-adjusted education differences in the probability of screening by country

	United States	Netherlands
Education		
Low	-0.115**	-0.027
	(0.056)	(0.023)
Middle	-0.104*	0.001
	(0.054)	(0.028)
High	base	base

Average marginal effects from logit models are reported. Both models control for age. For each individual, we calculate the marginal effect of an education category by taking the difference between the predicted probability of screening in that category and in the reference category. Average marginal effect is the average of individual marginal effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses.

Table 3.4: Differences in probability of screening mammogram by country

	United States		Netherlands	
	(1)	(2)	(3)	(4)
Education				
Low	-0.097*	0.024	-0.023	-0.038**
	(0.052)	(0.046)	(0.023)	(0.019)
Middle	-0.125**	-0.011	-0.004	-0.025
	(0.053)	(0.043)	(0.028)	(0.023)
High	base	base	base	base
Breast cancer in the family				
No	base	base	base	base
Yes - 1 family member	0.057	0.051	0.054*	0.045*
	(0.052)	(0.047)	(0.032)	(0.025)
Yes - 1+ family member	-0.006	-0.010	0.104	-0.002
	(0.145)	(0.080)	(0.104)	(0.079)
Has given birth	n.a.	n.a.	0.009	0.028
			(0.027)	(0.023)
Race/Ethnic origin				
White	-0.135**	-0.075	n.a.	n.a.
	(0.060)	(0.054)		
Dutch	n.a.	n.a.	-0.008	-0.010
			(0.033)	(0.026)
Perceived risk of getting breast cancer in the next 5 years	0.001	0.000	0.001**	0.001**
	(0.001)	(0.001)	(0.001)	(0.000)
Perceived effectiveness of mammogram at preventing death				
Negative	n.a.[a]	n.a.[a]	-0.018	0.020
			(0.034)	(0.028)
Low	base	base	base	base
Medium-low	0.088	-0.001	0.019	0.039*
	(0.095)	(0.062)	(0.028)	(0.023)
Medium-high	0.300***	0.166***	0.077***	0.054**
	(0.079)	(0.054)	(0.028)	(0.023)
High	0.235***	0.081	0.048	0.040*
	(0.086)	(0.057)	(0.030)	(0.024)

Table 3.4 – *continued from previous page*

	United States		Netherlands	
	(1)	(2)	(3)	(4)
Has a friend who died of breast cancer	n.a.	n.a.	0.042**	0.021
			(0.020)	(0.016)
Perceived probability of mammogram:				
giving a false-positive result	0.001	0.000	n.a.	n.a.
	(0.001)	(0.001)		
giving a false-negative result	-0.001	0.000	n.a.	n.a.
	(0.002)	(0.001)		
detecting breast cancer	0.000	0.001*	n.a.	n.a.
	(0.001)	(0.001)		
Income				
Low		base		base
Low-middle		-0.028		0.013
		(0.055)		(0.022)
Middle-high		0.006		0.009
		(0.049)		(0.023)
High		0.072		0.029
		(0.058)		(0.023)
Mammogram				
covered by insurance		0.225***		n.a.
		(0.050)		
recommended by doctor		0.384***		n.a.
		(0.063)		
invitation by program		n.a.		0.693***
				(0.041)
invitation in other manner		n.a.		0.415***
				(0.019)

[a] No coefficient since ALP respondents are presented with options 0%, 10%, 25%, 50%, 75%, 90% and 100%. Average marginal effects from logit models are reported. All models control for age. Reference category for white is non-white. For each individual, we calculate the marginal effect of a covariate category by taking the difference between the predicted probability of screening in that category and in the reference category. Average marginal effect is the average of individual marginal effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses.

Table 3.5: Differences in probability of medical advice, insurance coverage and screening invitation for mammogram

	United States		Netherlands	
	Medical advice to get mammogram	Insurance coverage for mammogram	Invited for mammogram by national program	Invited for mammogram in other manner
Age group				
40-49	base	base	base	base
50-64	0.022	-0.023	0.864***	-0.069***
	(0.059)	(0.064)	(0.014)	(0.017)
65-75	0.016	0.179**	0.859***	-0.082***
	(0.071)	(0.070)	(0.021)	(0.018)
75+	0.030	0.243***	0.129**	n.a.[a]
	(0.098)	(0.070)	(0.060)	
Education				
Low	-0.038	-0.107*	0.019	0.004
	(0.055)	(0.057)	(0.017)	(0.016)
Middle	-0.115*	-0.016	0.015	-0.006
	(0.061)	(0.045)	(0.021)	(0.017)
High	base	base	base	base
Income				
Low	base	base	base	base
Low-middle	0.016	-0.085	0.015	-0.036*
	(0.065)	(0.080)	(0.018)	(0.020)
High-middle	0.031	0.065	0.035**	-0.019
	(0.059)	(0.069)	(0.018)	(0.020)
High	0.002	0.130*	0.003	-0.033
	(0.081)	(0.067)	(0.020)	(0.020)
Breast cancer in the family				
No	base	base	base	base
Yes - 1 family member	0.090*	0.087	-0.028	0.088***
	(0.047)	(0.055)	(0.021)	(0.030)
Yes - 1+ family member	-0.011	0.102	-0.066	0.342***
	(0.127)	(0.109)	(0.065)	(0.121)
Has children	n.a.	n.a.	-0.006	-0.002
			(0.018)	(0.17)
Race/ethnic origin				
White	-0.129**	-0.016	n.a.	n.a.
	(0.061)	(0.068)		
Dutch	n.a.	n.a.	-0.009	0.006
			(0.022)	(0.020)
Invited in other manner	n.a.	n.a.	-0.405***	n.a.
			(0.053)	
Insurance coverage	0.204***	n.a.	n.a.	n.a.
	(0.061)			

[a] No coefficient since there are no women in the 75+ age range that have been invited in another manner.
Average marginal effects from logit models are reported. Reference category for white is non-white. For each individual, we calculate the marginal effect of a given category by taking the difference between the probability of getting screened in this category and in the reference category. Average marginal effect is simply the average of individual marginal effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses.

Chapter 4

Thought for Food: Understanding Education Disparities in Food Consumption

Higher educated individuals are healthier and live longer than their lower educated peers. One reason is that lower educated individuals engage more often in unhealthy behaviors, including consumption of a poor diet, but it is not clear why they do so. In this paper, we design a Discrete Choice Experiment, based on an economic model of food consumption, to understand the relationship between education and diet. Our results show that differences in dietary knowledge are responsible for the greatest part of the education disparity in diet. However, even when faced with the most explicit information regarding components of a healthy diet, lower educated individuals still state choices that imply a lower concern for negative health consequences of unhealthy eating patterns. This is consistent with the model's prediction that part of the education differences across health behaviors is driven by a higher "value of health" among the higher educated.

This chapter is based upon:

Koç, H. and van Kippersluis, H. (2015). *Thought for Food: Understanding Educational Disparities in Food Consumption.* (Tinbergen Institute Discussion Paper No. 15-034/V)

4.1 Introduction

The question "How can we induce people to look after their health?" has recently been chosen as one of the 10 most pressing questions in the social sciences (Giles, 2011). Answering this question requires a solid understanding of why people behave unhealthily. Given the strong disparities in the prevalence of healthy behaviors across education groups (e.g. Cutler & Lleras-Muney, 2010; Cawley & Ruhm, 2011), the reasons why lower educated individuals behave unhealthily is of particular interest. In a recent review, however, Cutler et al. (2011) note that *why* education affects health behaviors remains largely unclear.

In this paper our aim is to understand educational differences in one important type of health behavior, namely diet. With obesity rapidly approaching smoking as the leading preventable cause-of-death, dietary behavior gains unprecedented significance. We use an economic model to guide our design of a Discrete Choice Experiment (DCE), in which respondents make repeated choices between hypothetical meals that differ in taste, monetary price, preparation time, and health consequences. By randomly varying the information load that respondents face, our experimental design enables us to identify the parameters of the model, and helps us understand why lower educated individuals follow unhealthier diets: Is it because they know less about the health consequences of eating unhealthily or because they are very well aware of the consequences but simply care less about them?

Health behaviors have attracted a considerable amount of attention in the economics literature. One reason is the widely documented contribution of the education gradient in health behaviors to the education gradient in health (Cutler & Lleras-Muney, 2010). One can initially be inclined to argue that the education gradient in health behaviors is simply due to higher educated individuals earning a higher income. Drewnowski & Specter (2004) argue that unhealthy diets composed of energy dense foods (such as refined grains, added fats and sugars) are more affordable than healthy diets, and that the low cost of energy dense foods may partially explain the high prevalence of obesity among people with low levels of education. While an appealing argument, it is possible to find many unhealthy habits which are costly and more prevalent among the lower educated, such as smoking and binge drinking. Therefore it is unlikely that income is the sole explanation.

A further explanation can be the positive association between education and the personality traits that are needed to initiate and maintain healthy lifestyles, such as self-regulation (ability to defer an immediate reward for a future reward), internal locus of control (perceived control over one's life), and self-efficacy (Leganger & Kraft, 2003; Saffer, 2014). According to Conti & Hansman (2013), personality traits contribute to the

association between education and health behaviors (see also Barsky et al., 1997; Picone et al., 2004), although the direction of causality is not well-established.

Differences in health knowledge across educational groups are another explanation that is often stressed to explain the relationship between education and health behaviors. Grossman (1972) and Meara (2001) emphasize the "productive efficiency" hypothesis, which states that higher educated individuals make more efficient use of information, partly due to differences in cognitive ability (Bijwaard & Van Kippersluis, 2015). According to the "allocative efficiency" hypothesis, higher educated individuals choose more efficient inputs into health investment (healthier lifestyles), typically thought to be caused by better health knowledge and a more receptive attitude towards new information. Kenkel (1991), Meara (2001) and Cutler & Lleras-Muney (2010) provide support for the allocative efficiency hypothesis by showing that higher educated individuals have superior knowledge on the health effects of smoking, drinking and exercise, although this observation explains only a limited portion of the education disparities in health behaviors.

In sum, the literature established that education is closely related to a large battery of health behaviors, but there is no consensus about the underlying mechanisms and their relative contributions to the association. Gaining a better understanding of the mechanisms is important from a policy perspective. If educational differences in the possession of health knowledge are the key to the observed disparities, then policy efforts should be directed more towards equalizing health knowledge among people from different educational backgrounds. On the other hand, if educational differences in cognitive skills or personality traits are the main drivers, then, depending on the direction of causality, more structural changes that aim to reduce differences in education, cognitive ability, and personality traits would be needed.[1]

We argue that the progress and consensus on why higher educated individuals engage in healthier behaviors are at least partially hampered by the exclusive reliance on revealed preference data. This paper is, to the best of our knowledge, the first in the economics literature to use a Discrete Choice Experiment (DCE) to investigate educational disparities in dietary behavior.[2]

Compared to revealed preference approach, our DCE defines, not assumes, the choice

[1] From a more liberal perspective, any kind of intervention that goes beyond equalizing opportunity would be deemed unnecessary as differences across educational groups are the result of free, albeit constrained, choices.

[2] The food demand literature that employs discrete choice experiments has mainly focused on estimating the demand for novel food attributes (Adamowicz & Swait, 2011). We are aware of only one study by Kamphuis et al. (2015) in public health literature, which aims to document socioeconomic differences in dietary choice in a similar fashion to us. However, unlike us, they do not discriminate between the role of health knowledge and the value of health in driving these differences.

sets and food attributes that individuals face. In revealed preference data, one typically assumes artificial choice sets on basis of other individuals' choices, but it is unclear whether these are the actual choice sets, and whether the objective attributes in the artificial choice sets match the perceived attributes of the individual. Moreover, the DCE allows us to separate preferences for tastiness from preferences for health aspects, by explicitly controlling for tastiness as a product attribute. In revealed preference data, even if we were to observe all objectively measured food attributes like price or fat content, when an individual is observed to choose the option with a high amount of fat over low, it is difficult to tell whether this is due to a dismissal of negative health consequences of high fat intake or due to the association of fat content and taste. Finally, in contrast to revealed preference data, options in the DCE are not restricted to products that currently exist in the market.

On top of the novel experimental design, which we view as complementary to more conventional revealed preference approach, our study further contributes to the existing work on the effect of information provision on diet. Downs et al. (2009) find that calorie consumption went down in hamburger restaurants in Brooklyn but not in Manhattan after posting calorie information became mandatory, suggesting that socioeconomically disadvantaged groups benefit more from provision of health information. Wisdom et al. (2010) and Bollinger et al. (2011) find that providing information on calorie content led to significantly lower calorie consumption, yet Elbel et al. (2009) and Finkelstein et al. (2011) find no effect of the menu labeling law on healthier food purchasing in fast-food chains.

We contribute to the menu labelling literature by choosing a non-fast-food restaurant setting and by investigating preferences for more health attributes like fat and sodium, apart from calories. While menu labelling studies have almost exclusively focused on calories (for exceptions see Mathios, 2000; Wansink & Chandon, 2006; Variyam, 2008), overconsumption of sodium and saturated fat may be equally, or even more, harmful to health. Furthermore, customers of fast-food chains may have a different profile, or mindset, than customers of normal restaurants or home cooks. Conditional on the choice to enter a fast-food restaurant, convenience and taste may be the priority rather than calories, or health consequences in general.

Our results indicate that the education disparity in diet derives mostly from superior health knowledge among the higher educated. When faced with health information, better educated respondents do not change their valuation of health related product attributes, while lower educated respondents start to put a higher value on these attributes. This finding suggests that the lower educated are the main beneficiaries of health information, and the education gradient in unhealthy food choice becomes smaller upon provision

of such information. Nonetheless, even after fully equalizing health information across education groups, the better educated tend to choose healthier diets. This suggests that higher educated respondents place a higher marginal value on their health – i.e. they care more about the health consequences of dietary patterns. Auxiliary analyses suggest that at least part of these differences in the value of health derives from higher incomes among the higher educated.

The remainder of this paper is organized as follows. In section 4.2, we present our economic model of food consumption and discuss its insights and predictions. Section 4.3 explains in detail the design of the DCE, and section 4.4 describes its implementation in the LISS internet panel. Section 4.5 contains the empirical models that we estimate whose results are presented in section 4.6. In section 4.7, we present several robustness checks, before discussing the results in section 4.8.

4.2 A Model of Food Consumption

4.2.1 Model Formulation

The model presented here is a static formulation of the model developed by Galama & Van Kippersluis (2010), and builds on the human capital theory of the demand for health investment by Grossman (1972). Individuals maximize the utility function $U(C_h, C_u, H)$, which is concave in healthy consumption C_h, unhealthy consumption C_u, and subjective health H. We differentiate between subjective health H, and objective health H^*, since in practice individuals rely on a subjective perception of their health rather than their objective health status (Ippolito, 1981; Johansson-Stenman, 2011). Subjective health is equal to objective health multiplied by the parameter $\lambda(E)$:

$$H = \lambda(E)H^*(C_h, C_u) \tag{4.1}$$

$\lambda(E)$ takes the value of 1 when individuals have perfect knowledge about their objective health; while it deviates from 1 when individuals over- or underestimate their objective health. We assume $\lambda(E)$ is a function of education since the lower educated generally have worse health knowledge than the higher educated (Kenkel, 1991; Cutler & Lleras-Muney, 2010).

Objective health is a weakly increasing function of healthy consumption, i.e. $\partial H^*/\partial C_h \geq 0$, and a strictly decreasing function of unhealthy consumption, i.e. $\partial H^*/\partial C_u < 0$. Healthy and unhealthy consumption are produced by combining goods and services pur-

chased in the market (X_h and X_u), and own time inputs (τ_{C_h} and τ_{C_u}):

$$C_h \equiv C_h(X_h, \tau_{C_h}) \tag{4.2}$$
$$C_u \equiv C_u(X_u, \tau_{C_u}) \tag{4.3}$$

Individuals maximize utility under two constraints. The first constraint is time (Equation 4.4). The total amount of time available to an individual, Ω, is fixed, and it is divided between work τ_w, time inputs into healthy consumption, time inputs into unhealthy consumption, and some time is lost due to sickness $s(H)$:

$$\Omega = \tau_w + \tau_{C_h} + \tau_{C_u} + s(H) \tag{4.4}$$

The second constraint that individuals face is the budget constraint (Equation 4.5). The amount of money available is determined by the product of the wage rate per hour $w(E)$, which is a function of education, and the number of hours worked. Expenditures include purchases of healthy and unhealthy consumption goods and services, at their respective prices p_{C_h} and p_{C_u}.

$$w(E)\tau_w = p_{C_h} X_h + p_{C_u} X_u \tag{4.5}$$

The Lagrangian of the above described optimization problem can be written as:

$$\Im = U(C_h, C_u, H) + \mu\Big\{w(E)\big[\Omega - \tau_{C_h} - \tau_{C_u} - s(H)\big] - p_{C_h} X_h - p_{C_u} X_u\Big\} \tag{4.6}$$

where μ is the Lagrange multiplier for the budget constraint, or the marginal value of income.

4.2.2 First-order conditions

The first-order condition for healthy consumption is (see Appendix 4.A for derivations):

$$\frac{\partial U}{\partial C_h} + q_H\, \lambda(E)\, \frac{\partial H^*}{\partial C_h} \;=\; \mu\, \pi_{C_h} \tag{4.7}$$

The left-hand side of (4.7) is the marginal benefit of healthy consumption, which is the sum of the marginal utility from healthy consumption and the perceived health benefit of healthy consumption. The perceived health benefit is given by the product of the marginal value of health q_H and the subjective assessment of the amount of health gained by healthy consumption, $\lambda(E)\,[\partial H^*/\partial C_h]$. The marginal value of health is described in (4.8) and equals to the sum of the marginal utility of health (consumption benefit) and

the marginal effect of health on productive time (production benefit).

$$q_H = \frac{\partial U}{\partial H} + \mu\, w(E) \left[-\frac{\partial s}{\partial H} \right] \tag{4.8}$$

The right-hand side of (4.7) multiplies the marginal value of income μ with the marginal cost of healthy consumption π_{C_h}, which is a function of the monetary price p_{X_h}, and the opportunity cost of time $w(E)$:

$$\pi_{C_h} \equiv \frac{p_{X_h}}{\partial C_h / \partial X_h} = \frac{w(E)}{\partial C_h / \partial \tau_{C_h}} \tag{4.9}$$

Similarly, the first-order condition for unhealthy consumption is

$$\frac{\partial U}{\partial C_u} \quad = \quad \mu\, \pi_{C_u} + q_H\, \lambda(E) \left[-\frac{\partial H^*}{\partial C_u} \right] \tag{4.10}$$

where

$$\pi_{C_u} \equiv \frac{p_{X_u}}{\partial C_u / \partial X_u} = \frac{w(E)}{\partial C_u / \partial \tau_{C_u}} \tag{4.11}$$

Like in the case of healthy consumption, the left-hand side of (4.10) captures the marginal utility that an individual derives from unhealthy consumption. The first term on the right-hand side multiplies the marginal value of income μ with the marginal cost of unhealthy consumption (see Equation 4.11). The second term on the right-hand side of (4.10) is the perceived health cost of unhealthy consumption. It is the product of the marginal value of health q_H and the subjective assessment of the "unhealthiness" of the good $\lambda(E)\,[-\partial H^*/\partial C_u]$.

4.2.3 Insights from the model

The model provides a framework for understanding the trade-offs while making dietary choices. On the one hand, individuals enjoy consuming unhealthy food, but this affects their health negatively, and in turn leads to a reduction in both utility and productive time in the labor market. Healthy food, on the other hand, improves health but may provide less utility and may be more costly and time-intensive to prepare. Therefore, depending on the relative prices, required time inputs, preferences, and subjective perceptions of (un)healthiness, individuals choose an optimal bundle of healthy and unhealthy food items.

The first order conditions for healthy and unhealthy consumption, given in (4.7) and (4.10) suggest that dietary choices are mainly governed by four product attributes: (i) taste, as reflected in the marginal utilities of consumption, $\partial U/\partial C_h$ and $\partial U/\partial C_u$, (ii)

the monetary price, p_{C_h} and p_{C_u}, (iii) the opportunity cost of time, $w(E)$, and (iv) the health consequences, $\partial H^*/\partial C_h$ and $\partial H^*/\partial C_u$. How much value individuals attach to these attributes is determined by parameters of the utility function (for taste), and by the marginal value of income, μ (for the monetary price and opportunity cost of time). Due to differences in budget and time constraints (e.g. Cutler et al., 2003; Drewnowski & Specter, 2004), differences in the efficiency of using market inputs and own time in production (Michael & Becker, 1973), and differences in preferences (Drewnowski, 1997), individuals from different educational backgrounds may have a different valuation of each attribute.

The first order conditions also illustrate that the valuation of health consequences depends on the product of q_H and $\lambda(E)$. By assumption $\lambda(E)$ is a function of education. Equation (4.8) illustrates that the marginal value of health is also a function of education. Higher educated individuals earn higher wages, and since better health enables them to generate more earnings by increasing the time spent working, they value health more (i.e. the higher educated have higher production benefits of health). As a result, $\partial q_H/\partial E > 0$.

In the empirical analysis, we seek to discriminate between the model parameters $\lambda(E)$ and q_H that are potentially causing education disparities in unhealthy diets. Our main question is: Are education disparities in dietary choice mainly caused by disparities in health knowledge, i.e. driven by $\lambda(E)$, or do they simply reflect disparities in the marginal value of health, i.e. q_H? A secondary question is to what extent any potential differences in the marginal value of health are driven by education, health, the marginal value of income, and preferences.[3]

Answering these questions is not straightforward and puts huge requirements on the data. In particular, q_H and $\lambda(E)$ are not separately identified. Only in the special case that $\lambda(E) = 1$, we can identify the marginal value of health. Since consumer demand theory does not provide any credible instrumental variables (Etilé, 2011), we have to rely on experimental variation in the amount of health information in order to separately identify q_H.

[3] Equation (4.8) suggests that the marginal value of health additionally may depend on health H, the marginal value of income μ, and preferences. Since all of these variables are likely to be correlated to education, educational disparities in the marginal value of health may partially derive from differences in the marginal value of income, health status, and preferences. It makes intuitive sense that preferences include time preferences as many of the health consequences of (un)healthy consumption are realized in the future. However, the model is unable to accommodate that due to its static nature.

4.3 Discrete Choice Experiment

A Discrete Choice Experiment (DCE) is a stated preference technique that aims at eliciting individual preferences for attributes of a certain (consumption) good. DCEs are strongly grounded in Random Utility Theory (Louviere et al., 2010). In a DCE, multiple choice sets are presented to respondents, and in each choice set respondents make a choice between two or more alternatives. An alternative is described by a number of attributes, each of which can take several levels. The fundamental idea is that utility is derived from the bundles of attributes that make up the consumption goods, and not the consumption goods per se (Lancaster, 1966).

4.3.1 Setting of the Design

The setting of our DCE is the choice for a dinner meal. In terms of dietary behavior, dinner seems to be the most relevant setting as it contains the largest fraction of calories (36 percent), fat (42 percent), salt (36 percent), and fiber (36 percent) in Dutch diet (Van Rossum et al., 2011), with similar findings for the U.S. (Cutler et al., 2003, p. 101). Moreover, the largest disparities in healthy diets across education groups seem to derive from regular meals rather than snacks, candies and other refreshments. In our sample, 38 percent of the higher educated eat candies and snacks at least once a week, compared with only 32 percent of the lower educated.

The question we present to respondents is "Which of the following two meals would you eat regularly, i.e. at least twice a week?"[4] By asking a general question about which of the two meals they would prefer, we intend to avoid the dependence of the choice on the respondent's current appetite as well as his/her earlier food choices during the day or the week.

An example choice set is shown in Table 4.1. The design is unlabeled with a forced choice between the two meals. That is, we make respondents choose between "Meal A" and "Meal B", and do not allow them to choose neither of the two. In a labeled design (e.g. pizza vs. mashed potatoes), individuals might have intrinsic preferences for, and associations with, the specific alternatives. This will contaminate the estimation of the attribute importance (Hensher et al., 2005, p. 113). The reason why we force respondents to make a choice is that we are mainly interested in the trade-offs between the attributes, and failure to make a choice does not convey any information on attribute importance (Hensher et al., 2005, p. 176).

[4] See Appendix 4.B for the full introductory text.

4.3.2 Attributes and Levels

The economic model developed in Section 4.2 determines our selection of attributes. The model postulates that taste, monetary cost, time (or opportunity) cost, and health consequences are the four attributes that are influential on choice. These four attributes reassuringly coincide with the results from Food Choice Questionnaire developed by Steptoe et al. (1995) who demonstrate that, out of the nine factors that emerge to be important for food choice, sensory appeal, health, convenience and price are the most important ones. Moreover, in a pilot study that we conducted among 87 respondents, no other attribute was consistently mentioned to be important by more than 5 percent of the respondents, establishing confidence in our selection of attributes. Nonetheless, to err on the side of prudence, we added the sentence "Assume all other characteristics of the meals are the same, e.g. they are equally filling, biological, fair-trade, contain an equal amount of carbohydrates and proteins etc." to the introductory text, in order to prevent respondents from making assumptions about possibly omitted attributes.

While the attributes taste, monetary cost, and time cost all seem relatively easily interpretable, the attribute health consequences is more difficult to operationalize. We choose to operationalize this attribute by dividing it into three separate attributes which we call *health attributes*: calories, saturated fat, and sodium. We restricted health consequences to three attributes to reduce the cognitive burden on the respondents, and consequently avoid choices made based on only a subset of the attributes (Mangham et al., 2009).

There are three reasons why we chose calories, saturated fat and sodium as the health attributes. First of all, the most recent Dietary Guidelines for Americans state that all three are associated with health consequences (U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2010). Overconsumption of calories is associated with overweight, obesity, and diabetes; overconsumption of sodium is associated with high blood pressure and stroke; and overconsumption of saturated fat is associated with high cholesterol and cardiovascular disease.[5] Secondly, listing the amounts for these three attributes on the Nutrition Facts label is compulsory (European Union, 2006). Third, official guidelines prescribe a daily recommended intake for these three attributes (U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2010). This in contrast with for example sugar, for which the daily recommended intake

[5] Admittedly, there is a certain amount of uncertainty regarding the relationship between dietary components and health. While negative health consequences of overconsumption of calories or sodium is quite well established, we are aware of recent studies that challenge the association between saturated fat and cardiovascular disease (e.g. Malhotra, 2013; Chowdhury et al., 2014). Nonetheless, we decided to follow the most recent Dietary Guidelines for Americans which is prepared by two federal government bodies with the official goal of protecting American citizens' health.

is not agreed upon.

The proposed levels for the attributes are:

- **Price** - 2 Euro, 6 Euro, 10 Euro
- **Time** - 10 minutes, 30 minutes, 50 minutes
- **Taste** - OK, Good, Very Good [6]
- **Calories** - 800 calories, 1100 calories, 1400 calories
- **Saturated Fat** - 10 gram, 20 gram, 30 gram
- **Sodium** - 900 milligram, 1200 milligram, 1500 milligram [7]

4.3.3 Experimental Design

Design size We present each individual with 18 choice sets which is seen as a practical limit before boredom sets in (Hanson et al., 2005). Using 5 blocks of 18 choice sets ensures that the total number of choice sets generated is 90, which gives a comfortable buffer to identify all main effects and two-way interactions of the attribute levels.[8] The blocking of the design is performed such that the levels of every attribute are evenly divided over the blocks.

Generating an efficient design As the example choice set in Table 4.1 illustrates, there is an awful amount of possible combinations to generate 90 choice sets with two alternatives. We have opted for an "efficient design" that chooses the 90 most informative choice sets, for a given set of prior values. Efficiency is achieved by minimizing the median "D-Error", which is the determinant of the asymptotic variance-covariance (AVC) matrix of the parameters (Huber & Zwerina, 1996; Hensher et al., 2005, p. 153). As a part of statistical efficiency, efficient designs rule out the so-called dominant alternatives – uninformative choice sets where one of the alternatives is superior to the other in all aspects (e.g. a meal that is tastier, cheaper, quicker, and healthier than the other meal).

[6] Kamphuis et al. (2015) have run pilots where they experiment with levels such as tasteless or bad for the attribute "taste". They have observed that respondents never choose a meal that tastes bad. Therefore, we set "OK – neither good nor bad" as the minimum level of tastiness a meal can reach in our experiment.

[7] Refer to Appendix 4.B for a detailed motivation of the levels.

[8] The number of choice sets directly determines the number of parameters that are identified. With 6 attributes, each with 3 levels, the full factorial including main effects and all interactions amounts to $3^6 = 729$ parameters. Since three-way and higher-order interactions are unlikely to be of importance (e.g. Hensher et al., 2005; Lancsar & Louviere, 2008), 90 choice sets allow us to identify all main effects (12) and two-way interactions (60).

When generating an efficient design, prior values have to be assigned to the parameters. Given that all our attributes have a clear ordinal structure (e.g. a lower price always yields a higher utility than a higher price; fewer saturated fat yields more utility than more saturated fat for a given taste), the sign of the parameters is easily determined. The magnitude of the parameters is less well-established, therefore we use Bayesian priors with 1000 Halton draws from a normal distribution, to ensure robustness against misspecification (see Appendix 4.B for details).

Gradually adding health information Scenario I (or the baseline scenario or no health information scenario) identifies the product of the model parameters $q_H(t)$ and $\lambda(E)$. This suggests that, for example, if the higher educated care more about calories, we do not know whether this is because they know more about the possible dangers of overconsuming calories, or because they care more about the health consequences of overconsuming calories. In order to separate out the effect of these two parameters, we generate two additional scenarios where we gradually add health information until the point where $\lambda(E)$ is equal to 1.

Scenario II (or health information scenario) is identical to scenario I, except for the fact that descriptions of the health attributes in the introductory text are supplemented with information about the adverse health effects of overconsumption, and the recommended daily allowances for dinner. Additionally, respondents are reminded in every choice set of the recommended intake via the sentence "The recommended intake for a dinner is 800 calories, 10 gram saturated fat, and 900 milligram sodium". An example choice set is given in Table 4.2.

The objective of scenario II is to see how education disparities in the value individuals attach to the health attributes change when health information is provided. However, while reducing potential differences in possession of health information across education groups, scenario II still requires cognitive capabilities to process and internalize the given health information. Therefore, in scenario III (or explicit health information scenario) we make the health information even more explicit with the aim of fully equalizing health information across groups, i.e. fixing $\lambda(E)$ at 1. For that purpose, we replace the three health attributes – calories, saturated fat, sodium – with the single attribute "health consequences":

- **Health consequences** - Healthy, i.e. associated with reduced risk of disease, Health Neutral, Unhealthy, i.e. associated with increased risk of disease

An example choice set is given in Table 4.3. Despite compromising of the realism of the choice that respondents face, scenario III makes sure that all respondents are on the same

page in terms of health information. Hence, we ascribe any potential differences in the valuation of the attribute "health consequences" to differences in the marginal value of health $q_H(t)$, rather than differences in health knowledge $\lambda(E)$.

4.4 Data and Descriptive Statistics

4.4.1 Data

Our DCE is implemented in the LISS (Longitudinal Internet Studies for the Social Sciences) panel administered by CentERdata (Tilburg University, The Netherlands). The LISS panel is a monthly internet panel that runs since October 2007, covering 5000 households or 8000 individuals who are paid upon completing a questionnaire. The panel is based on a true probability sample of households drawn from the population register by Statistics Netherlands. Households that can not otherwise participate are provided with a computer and internet connection.

A representative sample of 4,377 panel members have been randomly selected for the DCE in the first wave, which took place in April 2014. We have restricted the sample to 18+ individuals as younger ones often live with their parents, typically do not cook, and as a consequence do not have much choice over what they eat for dinner. Each respondent is first randomly assigned to one of the three scenarios, and then to one of the five blocks within a scenario (see Section 4.3.3 for more detail on the scenarios and blocks). Within a block, each respondent is presented with 18 randomly ordered choice sets, with randomly ordered attributes (Kjær et al., 2006). Table 4.4 provides descriptive statistics for the sample. Our respondents are between 18 and 91, with an average age of 51, and roughly equally divided between men and women. Randomization of individuals to the different scenarios worked properly, as the means of the majority of the variables do not differ across scenarios.[9]

In a second wave, in May 2014, we have collected information about the respondents' time preferences, health knowledge (both objective and self-assessed), health valuation, and dietary habits.[10] Wave II is administered only among respondents who participated

[9] Due to chance, a couple of exceptions occurred. The respondents in scenario II eat more often in a restaurant than the respondents in scenario I (p-value=0.049), and the respondents in scenarios II and III consume vegetables more frequently than the respondents of scenario I (p-values are 0.047 and 0.091 respectively). The respondents of scenario III have a lower level of education than the respondents of scenario I (p-value=0.001). This does not pose any problems to our analysis of the education disparities in dietary behavior because we either compare respondents with different levels of education *within* scenarios, or respondents with the same level of education *across* scenarios.

[10] We ask the additional questions in a separate wave, one month later than the DCE, in order to avoid any priming and/or learning effects. It is likely that asking questions about health knowledge before the start of the DCE will make health attribute(s) salient compared to other product attributes in the

in wave I. After accounting for non-response (18.2 percent in the first wave and 10.5 percent in the second wave), we have 3,157 individuals who responded to both waves. After dropping respondents with missing values for variables used in the analysis, we end up with a final sample size of 2,869.[11]

4.4.2 Variables

The main variables used in the analysis are defined below. Note that all these variables are fixed for a given respondent across choice sets, and hence cannot be controlled for in the empirical estimation (see Section 4.5 for details). The variables do differ across respondents, however, and we will exploit this variation by estimating the models for different subgroups. To limit the number of subgroups, we choose to dichotomize all respondent characteristics defined below.

Education We measure education by the highest level completed with a diploma.[12] We dichotomize level of education as "lower education" referring to primary/secondary school or lower vocational education, and "higher education" referring to higher vocational education or university. 67 percent of our respondents are lower educated.

Self-reported health Self-reported health is measured by the question "How would you describe your health, generally speaking?", with excellent, very good, good, moderate, and poor as possible answers. We group the first three categories as "good health", and the latter two as "poor health".

Health knowledge Respondents' health knowledge with respect to diet is measured via 12 yes/no questions, including a "I do not know" option (see Table 4.6 for the questions). Half of the questions is about calories, saturated fat, and sodium as these are the dietary components that we focus on in our DCE. The other half of the questions asks about other

DCE, resulting in an overestimation of the relative importance of the health attribute(s). Asking health knowledge questions immediately after the DCE may result in an overestimation of health knowledge among scenario II respondents, as they were provided with health information. Interestingly, when we compare the number of correct answers given to health knowledge questions among respondents from scenario I and scenario II, we observe no difference, suggesting that answering questions in scenario II did not result in a lasting accumulation of health knowledge.

[11] According to Lancsar & Louviere (2008), to estimate reliable models, one rarely needs more than 20 respondents per parameter. We have 12 main effects in scenarios I and II, and 8 main effects in scenario III. Therefore our sample size gives us a very comfortable buffer to identify all main effects reliably, and even permits estimating two-way interactions if deemed necessary.

[12] Respondents between age 18 and 25 may still be attending school or university. However, when we restrict the analyses to individuals aged 25 and above, our results remain unchanged.

dietary components with the aim of getting an idea about the level of general knowledge the respondent has.

We construct a dietary knowledge index by counting the number of correct answers for 12 questions. We follow a strict scoring procedure in the sense that we count "do not know" responses also as incorrect, as both incorrect and "do not know" responses indicate a lack of knowledge. Then we construct a binary knowledge variable where a person is considered to have "high health knowledge" if she has answered more questions correctly than the median respondent, and "low health knowledge" otherwise.

Income Our economic model shows that the optimal levels of both healthy and unhealthy consumption are influenced by μ, i.e. the marginal value of income. We use current income to measure how tight the budget constraint is. Income is measured continuously as net monthly household income in Euros. We recode this variable into a binary one, where "high income" corresponds to income levels above the median level of income in the sample, and "low income" otherwise.

Future orientation Following Oreopoulos & Salvanes (2011), we have asked respondents to rate their agreement with the statement "Nowadays, a person has to live pretty much for today and let tomorrow take care of itself", with strongly disagree, disagree, neutral, agree and strongly agree as possible answers. We treat responses to this statement as a proxy for future orientation (or time preference) where a higher degree of agreement implies a higher level of orientation to the future. A person is defined to have "high future orientation" if she (strongly) disagrees with the statement, and "low future orientation" otherwise. While simplistic, we still prefer to use this measure over more conventional ones, as it is simple for respondents to understand and respond to. Indeed, the non-response and irrational response rate for a more classical way of measuring time preference via *money now versus money later* type of comparisons is very large in our sample (around 30 percent).[13]

Diet We have asked respondents whether they follow any diet, with possible dietary restrictions for salt, cholesterol, calories, or other diets. The binary variable "diet" takes the value of 1 if the respondent follows any kind of diet, and 0 otherwise.

[13] Our future orientation measure is strongly correlated to the classic time preference question "If offered 100 euros now or X euros in 6 months, what would be the smallest amount of money you would accept rather than the immediately available 100 euros?". After omitting irrational responses, i.e. X<=100 (leaving N=1778), those who are future oriented have on average discount rates that are 30 percentage points lower (p-value < 0.01).

Dietary habits We have asked respondents how often they consume the following goods, on a scale of 1 (Never) to 6 (Every day): fruit, vegetables, candy, sodas and energy drinks, and general snacks. Using the same scale, we have also asked them to rate how often they (i) cook at home on basis of raw ingredients, (ii) cook processed meals at home, (iii) have a take-away or home-delivered meal, and (iv) eat out in a restaurant.

4.4.3 Descriptive Evidence of Disparities across Education Groups

Table 4.5 documents average dietary habits among the lower and the higher educated respondents. In line with existing evidence (e.g. De Irala-Estevez et al., 2000; Cutler & Lleras-Muney, 2010), higher educated individuals are found to engage in healthier eating behaviors than the lower educated: They consume fruit and vegetables significantly more frequently (p-values 0.000 and 0.004, respectively), and drink sodas significantly less often (p-value=0.028) than the lower educated. No significant difference is found in snacking behavior, but, interestingly, the higher educated seem to eat candies slightly more often (p-value=0.003).

The table additionally documents educational disparities in dietary knowledge. As indicated by a significantly larger share of respondents with high health knowledge, overall dietary information is more widespread among the higher educated (p-value=0.000). Further analysis of the individual dietary knowledge questions (not shown) reveals that for 9 out of 12 questions, the share of respondents giving a correct answer is higher among the higher educated at 1 percent significance level. These respondents are more knowledgeable about the recommended amounts of calories and saturated fat; about the health consequences of overconsuming calories, saturated fat and sodium; and more generally about what constitutes a healthy diet and diet-disease connections.

In the next section we will describe our empirical approach to estimate the contribution of health knowledge to the reported disparities in healthy diets across education groups.

4.5 Empirical Estimation

4.5.1 From Economic Model to Empirical Estimation

Optimal levels of healthy and unhealthy consumption, given in equations (4.7) and (4.10) suggest that the utility individuals derive from consumption goods is a function of taste, price, time cost, and health consequences. Translated into a Random Utility Framework

this implies that the utility that individual i derives from meal j can be written as

$$U_{ij} = \mathbf{x'_{ij}}\beta_{\mathbf{i}} + \varepsilon_{ij} \tag{4.12}$$

where $\mathbf{x_{ij}}$ is the matrix of product attributes taste, price, time, health consequences, β_i is the vector of individual specific coefficients/valuations of the attributes, and ε_{ij} is an error term.

The β_i coefficients are the empirical analogues of the model parameters. In particular, the coefficients for the health attributes in scenario I identify the product of the marginal value of health q_H and the health knowledge parameter $\lambda(E)$. On the other hand, Scenario III fixes $\lambda(E)$ to 1, allowing us to interpret the coefficient on the attribute "health consequences" as the empirical translation of q_H, the marginal value of health. Estimating the exact model parameters $\lambda(E)$ and q_H would however require rather stringent functional form assumptions that we do not want to impose.

The economic model suggests that the coefficients, β_i, on the product attributes are heterogeneous. For example, according to Equation (4.8), the marginal value of health, which is empirically translated to the coefficient on the attribute "health consequences" in Scenario III, depends on education, the health stock, the marginal value of income, and time preference. Therefore, we specify a mixed logit (also known as random parameters logit) model, which allows for taste heterogeneity by letting each individual have an individual-specific coefficient on every attribute.

4.5.2 Panel Mixed Logit

Model specification Generalizing the standard logit model, mixed logit specifies an *individual specific* vector of coefficients β_i as in Equation (4.12). β_i follows the density $f(\beta_i|\theta)$, which describes the variation in tastes in the population with θ representing the parameter set.

As in the regular logit model, the respondent compares the utility of choosing alternative $j = 0$ with the utility of choosing alternative $j = 1$, and chooses the alternative with greater utility.[14] What we observe is the outcome, $y_i = \{0, 1\}$, of these latent utility comparisons. The mixed logit model defines the unconditional probability of individual

[14] Psychologists heavily criticize the assumption of rational economic agents who have complete and transitive preferences, and based on these preferences, choose the option that gives them the highest amount of utility (e.g. Simon, 1956; Krantz, 1991). They argue that preferences are constructed in the process of elicitation rather than being fixed. They are sensitive to the way a choice problem is presented or the mode of response (Slovic, 1995; Kahneman, 2011). Therefore, choice behavior can be better be described by "bounded rationality" where agents aim to attain a satisfactory level of utility (but not necessarily maximal).

i choosing alternative j in a given choice set as

$$P_{ij}^{u}(\theta) = \int \left(\frac{e^{x_{ij}\beta_i}}{1 + e^{x_{ij}\beta_i}} \right) f(\beta_i|\theta) d\beta_i \tag{4.13}$$

Hence, the unconditional probability in the mixed logit model is simply a weighted average of standard logit probabilities for different values of β_i. Respective weights for each β_i value are provided by the density $f(\beta_i|\theta)$. Thus, the regular logit model is a special case of the mixed logit model where β_i takes a single value b for everyone and $f(\beta) = 1$ for $\beta = b$.

The unconditional likelihood of individual i making the observed series of choices, $(y_{i1}, y_{i2}, ..., y_{i18})$, can be derived from his/her unconditional probabilities for every choice situation t, and is given by

$$L_{i}^{u}(\theta^*) = \int \prod_{t=1}^{18} \left(\frac{e^{x_{ijt}\beta_i}}{1 + e^{x_{ijt}\beta_i}} \right) f(\beta_i|\theta) d\beta_i \tag{4.14}$$

where the inner part is the conditional likelihood given the parameter values

$$L_{i}^{c}(\beta_i) = \prod_{t=1}^{18} \frac{e^{x_{ijt}\beta_i}}{1 + e^{x_{ijt}\beta_i}} \tag{4.15}$$

and the log-likelihood function for the model is $LL(\theta) = \sum_i \ln L_i^u(\theta)$.

Since the loglikelihood function depends on unknown parameters, and involves an integral (see Equation 4.14) that cannot be solved analytically, exact maximum likelihood estimation is not possible. Instead a distribution is specified for the density $f(\beta_i|\theta)$ with given values of the parameter set θ. In our case, we choose to specify a normal distribution. A value of β_i is drawn from the normal density and in turn standard logit probabilities are calculated for each choice set. The product of the standard logit probabilities are used to calculate the conditional likelihood given in (4.15). This process is repeated for many draws and the average of the resulting conditional likelihoods is used to approximate the unconditional likelihood:

$$SL_{i}^{u}(\theta) = (1/R) \sum_{r=1,...,R} L_{i}^{c}(\beta_i^r|\theta) \tag{4.16}$$

where R is the number of draws, $\beta_i^r|\theta$ is the r^{th} draw from $f(\beta_i|\theta)$, and $SL_i^u(\theta)$ is the simulated likelihood of individual i's sequence of choices.

The simulated log-likelihood function is constructed as $SLL(\theta) = \sum_i \ln SL_i^u(\theta)$ and is maximized to find a consistent estimator of the true parameter vector θ (see Train,

2009 for a complete discussion).

Extracting individual-level coefficients Mixed logit model allows for random coefficients whose distribution in the population, $f(\beta_i|\theta)$, is estimated. Revelt & Train (2000) show that, by using a respondent's choices, it is possible to determine the position of her preferences in the overall distribution of preferences. To that end, one should distinguish between $f(\beta_i|\theta)$ and $h(\beta_i|\tilde{y}_{i1}, \tilde{y}_{i2}, ..., \tilde{y}_{i18}; \theta)$ which is the distribution of coefficients among the subpopulation of people who made a particular sequence of choices $\tilde{y}_{i1}, \tilde{y}_{i2}, ..., \tilde{y}_{i18}$. By Bayes' rule:

$$h\left(\beta|\tilde{y}_{i1}, \tilde{y}_{i2}, ..., \tilde{y}_{i18}; \theta\right) = \frac{P\left(\tilde{y}_{i1}, \tilde{y}_{i2}, ..., \tilde{y}_{i18}|\beta\right) f\left(\beta_i|\theta\right)}{P\left(\tilde{y}_{i1}, \tilde{y}_{i2}, ..., \tilde{y}_{i18}|\theta\right)} \tag{4.17}$$

where $P\left(\tilde{y}_{i1}, \tilde{y}_{i2}, ..., \tilde{y}_{i18}|\theta\right) = \int P\left(\tilde{y}_{i1}, \tilde{y}_{i2}, ..., \tilde{y}_{i18}|\beta_i\right) f(\beta_i|\theta)\, d\beta_i$ is the choice probability integrated over all possible value of β_i. Since, this integral does not have an analytical solution, again we resort to simulation, and for each individual compute the mean of the distribution $h\left(\beta_i|\tilde{y}_{i1}, \tilde{y}_{i2}, ..., \tilde{y}_{i18}; \theta\right)$.

We are particularly interested in the individual specific preferences for the product attribute "health consequences" in scenario III, which can be interpreted as the empirical analogue of the marginal value of health. Looking at individual specific preferences will give us the opportunity to explore heterogeneity in the marginal value of health, i.e. to what extent do education disparities in the marginal value of health reflect differences in income, time preference, and health. It will additionally allow us to gauge the validity of the stated preference data, by comparing the individual specific stated preferences for health with the reported food choices.

4.6 Results

4.6.1 Results on the Full Sample

Table 4.7 presents the parameter estimates, estimated separately for each scenario. All attributes have a statistically significant impact on food choice with the expected sign. A higher price and longer preparation time make an alternative less likely to be chosen. On the other hand, tastier and healthier alternatives are more likely to be chosen.[15] The estimated standard deviations are highly significant for the majority of attributes, indicating that there is a considerable amount of heterogeneity in the valuation of these

[15] A robustness check has shown that replacing the attribute name "taste" with "sensory appeal", which additionally connotes smell and visual appearance, does not make a difference for our results, and if anything it makes individuals care less about the attribute.

attributes across respondents. In the following sections, we will have a closer look at the sources of this heterogeneity, educational attainment in particular.

Because coefficients from a mixed logit model are not directly interpretable, we report the average marginal effect of each attribute level on the choice probability for all scenarios in Table 4.8. For example, when the price of an alternative increases from 2 Euros to 10 Euros, the probability of choosing that alternative is reduced by 17 to 24 percentage points, ceteris paribus. Likewise, when the calorie content of a meal increases from 800 calories to 1400 calories, the probability of that meal being chosen goes down by 16 to 18 percentage points. The price, taste, and calories seem to be the most important product attributes in the baseline scenario (scenario I), yet one should still be careful with comparing the relative importance of different attributes as the variation in attribute levels is hard to compare (e.g. "taste" varies from OK to Very Good, while "price" varies from 2 to 10 Euros).[16]

Scenario II is identical to scenario I except for the fact that it informs respondents about diet-disease connections and daily recommended amounts of calories, saturated fat and sodium. Comparing results from the two scenarios reveal that valuation of the health attributes is higher among respondents of scenario II. In other words, people place a higher value on the health attributes when faced with health information.[17] Interestingly, respondents in scenario II place a relatively lower value on the price and taste compared to their peers in scenario I. This suggests that when faced with new health information, individuals are willing to trade off part of the taste and price, but not time, for a healthier meal.

4.6.2 Disparities across Education Groups and the Role of Health Knowledge

In this section we explore heterogeneity in the valuation of health attributes by level of education. Table 4.9 presents average marginal effects, estimated separately for lower and higher educated respondents (first and second columns for each scenario, respectively).[18] The third column shows the difference between the two marginal effects.

In the baseline scenario, i.e. without any externally provided health information,

[16] Willingness to pay estimates reported in Appendix Table 4.D.1 give similar findings. In the baseline scenario, respondents are willing to pay the highest price to switch from a meal that contains 1400 calories to one that has 800 calories, ceteris paribus, or from a meal that tastes just OK to one that is very tasty.

[17] One should be careful when comparing the absolute size of the coefficients/marginal effects across scenarios as different scenarios can have different error scales. Against this possibility, willingness to pay estimates in Appendix Table 4.D.1, which are insensitive to scaling adds confidence to our finding by showing that willingness to pay for health attributes is higher under scenario II than scenario I.

[18] Refer to Appendix Table 4.D.2 for the full set of marginal effects.

higher educated individuals put a higher emphasis on the health attributes while making food choices. Differences are statistically significant at 5 percent level for calorie level 1400, for both levels of saturated fat, and for sodium level 1200 mg.[19]

Having shown the presence of educational disparities in valuation of health attributes, the next question to ask is whether higher educated make healthier food choices because they know more about the consequences of unhealthy consumption. The experimental design permits answering this question by comparing the education disparity among respondents randomly assigned to scenarios I and II. Our results indicate that all differences observed in the baseline scenario in the valuation of health attributes between the lower and the higher educated respondents disappear upon provision of information on recommended daily intake levels and health consequences under scenario II. We also observe that, while the higher educated hardly change their valuation (column 2 versus column 5), the lower educated start to care significantly more about calories, saturated fat and sodium when exposed to health information (column 1 versus column 4).[20] For example, among lower educated respondents, increasing the sodium content from 900 mg. to 1200 mg. reduces the choice probability by 8.3 percentage points under scenario II, in comparison to only 4.6 percentage points under scenario I. This implies that lower educated respondents are the main beneficiaries of health information, and supply of health information have rendered the educational disparities statistically insignificant.

4.6.3 Disparities across Education Groups and the Role of Value of Health

In scenario III, we give respondents the most explicit health information available, that is, we tell them whether a meal is "healthy", "health neutral", or "unhealthy". This equalizes all health information across respondents, enabling us to attribute any potential education disparity in the valuation of the attribute "health consequences" entirely to a difference in marginal value of health between the lower and the higher educated. The final three columns of Table 4.9 show that, under the most explicit health information available, there is a small and statistically insignificant difference in how much the lower and higher educated value "health consequences". This is consistent with our earlier finding that a large part of the education disparity in diet is driven by health knowledge.

[19] It should be noted that one's level of caloric intake is proportional to her energy need. Since lower educated individuals engage more often in physically demanding jobs (60 percent of the lower educated in contrast to 30 percent of the higher educated in our sample), difference in energy need may partially be responsible for the difference in preference for calories between the two groups.

[20] The fact that it is the lower, not the higher educated, who have changed behavior rules out any potential differences in the ability to process or internalize information favoring the higher educated as a possible explanation for our findings.

Despite lack of statistical significance, Table 4.9 shows that even with the most explicit health information, the lower educated concern themselves slightly less with the health consequences of their choices than the higher educated, suggesting the existence of an educational difference in the valuation of health. In order to gain more insight into this difference, we extract the individual level coefficients for the attribute level "Health consequences = Healthy" as a proxy for the value of health, and use them as the dependent variable in a cross-sectional regression on education, income, health, and preferences, among respondents of scenario III.[21]

Table 4.10 shows that, conditional on a standard set of demographic variables – age, age-squared and gender – and diet status, higher educated individuals put a significantly higher value on their health (p-value=0.069). Given that standard deviation of the dependent variable is estimated as 0.93 in Table 4.7 (last column), moving from lower to higher education is associated with a 0.13 standard deviation increase in the marginal value of health. Table 4.10 further shows that women place a higher value on health, and there is positive relationship between age and value of health up to the age of 60.

Moving across columns shows that, while adding a proxy for time preference or self-reported health does not significantly change the coefficient on education, the addition of income renders it insignificant. This suggests that at least part of the education disparities in the value of health derives from higher income among the higher educated. This is consistent with Galama & Van Kippersluis (2010) who argue that, at higher levels of income, due to diminishing marginal utility of consumption, people start to care more about other goods, in particular health leading to a higher marginal value of health among the rich.

4.7 Robustness Checks

Assessing the predictive validity of stated preferences As we have also mentioned in Section 4.4, wave II has presented respondents with questions about their dietary habits, i.e. how often they consume certain dietary components, how often they cook at home etc. Such questions are not only common ways to measure dietary intake (Thompson & Subar, 2008, p. 11), but also established to be predictive for actual food intake (e.g. Willett et al., 1985). For every respondent in scenario III, we observe (i) their reported actual food choices, and (ii) their stated preferences with respect to healthy meals as measured by the individual-level coefficient on the attribute level "Health con-

[21] In this analysis we ignore the uncertainty around the individual level coefficients and treat them as deterministic variables. Without adjustment for this fact, it is likely that we get an underestimate of the standard errors.

sequences=Healthy". By using these two pieces of information it is possible to assess the predictive validity of the stated preferences for the reported food choices.

Table 4.11 shows that the stated preferences for hypothetical healthy food options are highly correlated with the actual choice of healthy food options. Respondents with a higher valuation or stated preference for healthy alternatives tend to consume more fruit and vegetables, and less soda, candies and snacks. They also cook more at home, and eat processed or take-away meals less often. To have an idea about the magnitude of the relationship, compare the 47 percent probability of eating fruit everyday for someone with an average valuation of healthy alternatives with the 56 percent probability for someone whose valuation of healthy alternatives is one standard deviation above the average. Results are similar for vegetables and other options building confidence that the stated preference data have external validity for actual choices, i.e. revealed preferences.

The second panel of Table 4.11 shows that stated preferences are equally predictive for actual choices irrespective of educational attainment.[22] While we are wary of many potential biases that remain in terms of the exact size of the coefficients from stated preference data (e.g. WTP is notoriously overstated in SP data due to hypothetical response bias) we argue that these biases are unlikely to differ systematically by educational attainment. Therefore, we conclude that the sign, and the relative magnitude across education groups, of stated preference coefficients contain useful and reliable information.

Priming effects Despite the evidence presented above which is in line with the importance of health knowledge for educational disparities in dietary behavior, one can argue that when faced with health information individuals start to care more about health attributes only because provision of such information makes health attributes salient, i.e. draws attention to these attributes. In order to investigate this possibility, we have generated an additional three-scenario DCE and implemented this among 892 respondents residing in the U.S. via Amazon Mechanical Turk (MTurk).[23] The first two scenarios are identical to the original DCE where respondents are given no external health information in the first scenario, and provided with information about diet-disease relationships and recommended daily allowances in the second. The third scenario is constructed identically to the first two scenarios but the attribute "time" is made more salient via addition

[22] Only for fruit, vegetables, and ready-cooked meals, the magnitude of the relationship between stated and actual choices differ significantly between the higher and the lower educated. However, for fruit and vegetables the relationship is stronger among the higher educated, while for ready-cooked meals it is weaker. Hence, no systematic patterns are discernible.

[23] MTurk is an online survey instrument. The randomization of respondents across scenarios has worked properly, with no statistically significant differences at 5 percent level with respect to age, gender, race, household size, education, and income across scenarios. See Appendix 4.C for details of the exact implementation.

of the uninformative sentence "If you spend t minutes on preparing food, you cannot do anything else in those t minutes". Since individuals are randomly assigned to the scenarios, and the information provided is completely useless, any difference in the value that scenario I (no information) and scenario III (time information) respondents attach to the time attribute would be due to salience.

Table 4.12 presents average marginal effects from mixed logit models estimated for each scenario. Column III shows the difference in the valuation of health attributes under the no health information scenario (column I) and health information scenario (column II). In line with our previous results, respondents attach a higher value to the health attributes when faced with health information. The differences are statistically significant for levels 1100 cals. and 1400 cals. for attribute "calories", and for level 30 gr. for attribute "saturated fat". Furthermore, comparing no time information scenario (column IV) with the time information scenario (column V) reveals that the importance of attribute "time" does not differ significantly (column VI). These results corroborate our main finding from Section 4.6.2 that individuals start to care more about health attributes when presented with health information, and that the information content, not salience, is the driving force behind behavior change.

Other robustness checks While the mixed logit model provides a flexible way of accounting for unobserved heterogeneity, one may be worried that the results are driven by the assumption on the distribution of individual specific coefficients in the population, $f(\beta_i|\theta)$. Additional analyses have shown that all average marginal effects are very similar when using a log-normal instead of the current normal distribution or when estimating a fixed effects logit model where coefficients are assumed to be fixed across respondents.

Another issue is possible differences in error scale between models estimated for the low and high educated within the same scenario. It might be the case that higher educated individuals make fewer errors while filling out the survey resulting in smaller error variance and larger parameter estimates for the higher educated. To account for this possibility, we have estimated the generalized multinomial logit model of Fiebig et al. (2010) and got similar results.

The randomization of the order in which the choice sets and the attributes are presented to the respondents makes it very unlikely that there is any systematic bias in the answers of respondents. Still, the results may suffer from what is known as "left-right bias" meaning that some respondents systematically prefer the left option, i.e. meal A. We have included an intercept in our mixed logit models to check whether this is the case. In all scenarios the constant term has turned out to be insignificant (results available upon request) suggesting that left-right bias is not an issue for our results.

Finally, the results may be different across demographic variables such as gender, age, and the size of the household. We find that men, younger individuals, and those living on their own generally care less about the health consequences, and that educational disparities are larger within those population groups. Nonetheless, also among women, individuals above 50, and among those living in larger households we observe educational disparities that become considerably smaller upon the provision of health information. This suggests that while demographic characteristics are important determinants of food choice, our main results hold irrespective of the population subgroup (results available upon request).

4.8 Discussion

While it is established that health behaviors such as diet contribute to the gap in health and life expectancy across education groups, *why* the higher educated eat healthier diets is unclear (Cutler & Lleras-Muney, 2010). To the best of our knowledge, this is the first attempt to adopt a stated preference Discrete Choice Experiment to understand education disparities in (un)healthy food choices. While somewhat unconventional, and potentially subject to hypothetical response bias, we argue that stated preference data can be complementary to revealed preference data in revealing mechanisms, and is even essential to separate individual preferences for health and taste.

We have two main messages. First, a large of part of education disparities in diet derives from differences in health knowledge. Higher educated individuals have superior knowledge on the adverse health consequences of overconsuming salt, fat and calories, and therefore value the health consequences of food more than lower educated individuals. When confronted with health information regarding the health consequences of the health attributes and the recommended allowance, higher educated individuals hardly changed their valuation of the attributes, while lower educated individuals strongly responded. Providing health information substantially reduced the disparities in the valuation of the health attributes, and suggests that health information/health knowledge is a key contributor to education disparities in diet.

The second message that derives from our results is that, even conditional on the most explicit health information (that is, individuals know exactly which meal is healthy and which one is unhealthy), and conditional on the price and time inputs in meal preparation, higher educated individual still have a higher valuation of the health consequences. In other words, conditional on health knowledge, higher educated individuals simply care more about their health. Auxiliary analyses suggest that at least part of these differences in the marginal value of health derive from higher incomes among the higher educated.

This result is consistent with theory, where higher educated individuals earn higher wages, accumulate more income, and because of diminishing marginal utility of consumption, start putting a higher marginal value on their health.

One major concern of stated preference data is hypothetical response bias (Loomis, 2011; Hausman, 2012). While this is an unavoidable limitation of any stated preference approach, we argue that the specific criticism applies mostly to the exact size of the estimated coefficients, and does not interfere with our aim of understanding education disparities in diet. Firstly, the mixed logit model we apply in our estimations allows extracting the individual-level stated preferences for healthy food attributes. Correlating these individual stated preferences to actual healthy food choices shows strong predictive power: stated preferences are predictive for revealed preferences. Second, we are careful not to interpret the actual size of the coefficients and implied WTP values, as these are likely to be biased. We do compare the relative magnitudes of the coefficients, which is less problematic since there is no evidence for systematic differences in hypothetical biases across education groups.

The food choice setting we investigate is the choice for a dinner meal. Arguably, the time lapse between the choice and actual consumption of dinner is sufficiently long for our rational decision framework to be applicable. In contrast, for snacks and in fast food contexts, it is established that impulses, self-control, and craving play an important role (e.g. Shiv & Fedorikhin, 1999). In these contexts, our rational model could serve as the 'long-run self' or 'cool state', while an additional 'short-run self' or 'hot state' (Bernheim & Rangel, 2004; Fudenberg & Levine, 2006) would have to be added in order to accommodate the temptations associated with the consumption of snacks and other fast food.

We do not attempt to estimate the causal effect of education on diet. Hence, the education disparities in diet, and variations in health knowledge and the value of health, are likely to reflect variables correlated to educational attainment, such as cognitive and non-cognitive abilities (e.g. Conti et al., 2010). Instead, the analysis in this paper reveals mechanisms through which education, and its correlates, impact on food choices, which is an essential input into any policy discussion on encouraging healthy diets.

The implications of our results are threefold. First, with the caveat that the point estimates should be interpreted with caution, the results allow for some counterfactual simulations to gauge what it requires to make the lower educated eat healthier diets. There exists a widespread notion that healthy meals are expensive, inconvenient, and usually not very tasty. In contrast, unhealthy meals are generally cheap, tasty and convenient. One interesting counterfactual is what the tax should be on an unhealthy, tasty, and convenient meal in order to make lower educated individuals switch to a healthy

and inconvenient meal that is not so tasty. Our results suggest that the price of the unhealthy alternative should be at least 6 Euros, representing an enormous 200 percent additional tax compared with an assumed baseline price of 2 Euros. If healthy meals were more convenient (that is, reducing the preparation time from 50 to 10 minutes), the indifference price would still have to be 4 Euros, which implies a 100 percent tax. Therefore, our results suggest that while taxes on unhealthy food and the availability of convenient healthy food would make healthier options more attractive for lower educated individuals, the required taxes and time gains would have to be implausibly large to make these options equally attractive as a tasty, cheap, quick and unhealthy meal.

Second, the provision of health information significantly increased the marginal value attached to the health consequences of food, especially among lower educated individuals. The assumption of perfect information is clearly rejected by the data. This is in line with the results of Downs et al. (2009), Roberto et al. (2010), Wisdom et al. (2010) and Bollinger et al. (2011), who all find that calorie labeling reduced the amount of calories ordered and consumed. Arguably, disparities in diet that result from lack of information, jointly with the large external medical costs of obesity (Cawley & Meyerhoefer, 2012), and potential self-control issues (Hoch & Loewenstein, 1991), could give a justification for policy intervention. The first movements in this regard are already made (e.g. Guidingstars.com). The findings of this study suggest that health warnings may be a more promising alternative than the introduction of taxes to promote healthy diets among the lower educated, not just in terms of calories, but also in terms of salt and fat, and outside of fast-food settings.

A third implication is that, even with the most explicit health information possible, education disparities in food choices will remain. The empirical finding that conditional on health information, and a standard set of demographic characteristics, higher educated individuals value health consequences more is consistent with the theoretical prediction that the marginal value of health is positively influenced by educational attainment. The disparities in food choice, as well as more general health behavior, that derive from disparities in the marginal value of health are the result of free, albeit constrained, choices. For the part of disparities that results from disparities in the marginal value of health, policy intervention is unlikely to be successful. In fact, it may even be welfare reducing, unless the deeper causes of variation in the marginal value of health are tackled, which seems a heroic task.

Tables

Table 4.1: Example choice set – Scenario I

	Meal A	Meal B
Price	2 Euro	6 Euro
Time	10 min.	30 min.
Taste	OK	Very good
Calories	1400 calories	800 calories
Saturated Fat	10 gr.	30 gr.
Sodium	1200 mg.	900 mg.

Table 4.2: Example choice set – Scenario II

	Meal A	Meal B
Price	2 Euro	6 Euro
Time	10 min.	30 min.
Taste	OK	Very Good
Calories	1400 calories	800 calories
Saturated Fat	10 gr.	30 gr.
Sodium	1200 mg.	900 mg.

The recommended intake for a dinner is 800 calories, 10 gram saturated fat, and 900 milligram sodium.

Table 4.3: Example choice set – Scenario III

	Meal A	Meal B
Price	6 Euro	10 Euro
Time	50 min.	30 min.
Taste	Very Good	Good
Health consequences	Unhealthy	Health Neutral

Table 4.4: Distribution of respondent characteristics across scenarios

	Scenario I	Scenario II	Scenario III
Age	51.25	50.94	51.89
Male	0.47	0.48	0.47
Education=High	0.37	0.33	0.30
Health=Good	0.81	0.81	0.80
Health knowledge=High	0.45	0.46	0.44
Income=High	0.50	0.52	0.47
Future orientation=High	0.41	0.44	0.43
Currently on (any kind of) diet	0.15	0.16	0.14
Frequently go to a restaurant	0.00	0.01	0.00
Frequently eat fruits	0.82	0.81	0.81
Frequently eat vegetables	0.94	0.96	0.96
Frequently eat candy	0.35	0.33	0.34
Frequently drink soda	0.31	0.28	0.29
Frequently snack	0.16	0.15	0.13
No. of respondents	974	917	978

Notes: Mean values are reported. High education refers to having completed higher vocational education or university with a diploma. Good health is self reported health as very good or excellent (measured in a separate wave, number of respondents is 736, 683 and 725 in scenarios I, II and III, respectively). High health knowledge is having answered a higher number of health knowledge questions correctly than the median respondent. High income refers to an income level above that of the median respondent in the sample. High future orientation indicates (strong) disagreement with the statement "Nowadays, a person has to live pretty much for today and let tomorrow take care of itself". Frequently is defined as at least 3-4 times per week.

Table 4.5: Distribution of dietary habits and health knowledge across education groups

	Low educated	High educated
Dietary habits		
Frequently eat fruits	0.79	0.86
Frequently eat vegetables	0.95	0.97
Frequently eat candy	0.32	0.38
Frequently drink soda	0.31	0.27
Frequently snack	0.14	0.16
Health knowledge		
Health knowledge=High	0.39	0.58
No. of correct answers to health questions	7.54	8.61
No. of respondents	1,916	953

Notes: Mean values are reported. We define frequently as at least 3-4 times per week. High health knowledge is having answered more than the median number of questions correctly.

Table 4.6: Health and dietary knowledge items - Percentage of respondents reporting the correct answer

	Low educated	High educated
Even in the absence of overweight, poor diet is associated with cardiovascular disease, hypertension, and type 2 diabetes. (True)	89.35%	91.50%
There are health benefits of limiting those foods which contain high levels of added sugar such as soft drinks, cordial and biscuits. (True)	87.42%	92.24%
Overconsumption of sodium can lead to hypertension and heart diseases. (True)	77.09%	87.09%
Depending on age and physical activity level, experts recommend that an adult male should consume around 2500 calories, and an adult female should consume around 2000 calories, per day. (True)	68.89%	78.80%
Consumption of fruits and vegetables is associated with reduced risk of many chronic diseases. (True)	68.68%	79.01%
Sodium is a form of sugar. (False)	64.41%	81.43%
Meat, chicken, fish and eggs should make up the largest part of our diet. (False)	64.87%	79.01%
Experts advise to eat a variety of vegetables, especially dark green, red and orange vegetables. (True)	66.49%	65.06%
Choosing wholemeal bread provides no health benefits. (False)	62.00%	71.14%
A high intake of saturated fat can protect against heart diseases. (False)	49.32%	64.74%
According to experts around 30 percent of the calories in a day should come from saturated fat. (False)	36.01%	50.05%
For a healthy adult it is recommended to limit sodium intake at dinner to at most 1500 mg. (False)	19.26%	20.99%
No. of respondents	1,916	953

Table 4.7: Coefficient estimates from mixed logit models with all normally distributed coefficients

		No health info.	Health info.	Explicit health info.
Price – Baseline is 2 Euro				
6 euro	Mean coefficient	-0.516***	-0.483***	-0.682***
		(0.031)	(0.032)	(0.033)
	Std. dev. of coefficient	0.383***	0.365***	0.189**
		(0.052)	(0.051)	(0.088)
10 euro	Mean coefficient	-1.341***	-1.121***	-1.797***
		(0.068)	(0.068)	(0.074)
	Std. dev. of coefficient	1.190***	1.167***	1.511***
		(0.059)	(0.059)	(0.056)
Time – Baseline is 10 minutes				
30 min.	Mean coefficient	-0.087***	-0.134***	-0.306***
		(0.029)	(0.031)	(0.031)
	Std. dev. of coefficient	0.133*	0.187**	0.064
		(0.069)	(0.076)	(0.069)
50 min.	Mean coefficient	-0.537***	-0.535***	-0.987***
		(0.051)	(0.054)	(0.055)
	Std. dev. of coefficient	0.935***	0.967***	1.152***
		(0.048)	(0.049)	(0.046)
Taste – Baseline is "OK"				
Good	Mean coefficient	0.661***	0.524***	0.404***
		(0.032)	(0.033)	(0.032)
	Std. dev. of coefficient	0.129**	0.057	0.004
		(0.057)	(0.053)	(0.057)
Very good	Mean coefficient	1.066***	0.848***	0.908***
		(0.043)	(0.043)	(0.041)
	Std. dev. of coefficient	0.744***	0.675***	0.554***
		(0.039)	(0.043)	(0.048)
Calories – Baseline is 800 cals.				
1100 calories	Mean coefficient	-0.447***	-0.632***	
		(0.028)	(0.030)	
	Std. dev. of coefficient	0.026	0.013	
		(0.049)	(0.050)	
1400 calories	Mean coefficient	-0.967***	-1.179***	
		(0.041)	(0.044)	
	Std. dev. of coefficient	0.753***	0.784***	

		(0.040)	(0.042)	
Saturated Fat – Baseline is 10 gram				
20 gram	Mean coefficient	-0.210***	-0.239***	
		(0.030)	(0.031)	
	Std. dev. of coefficient	0.043	0.054	
		(0.043)	(0.048)	
30 gram	Mean coefficient	-0.488***	-0.629***	
		(0.031)	(0.034)	
	Std. dev. of coefficient	0.243***	0.323***	
		(0.057)	(0.052)	
Sodium – Baseline is 900 milligram				
1200 mg.	Mean coefficient	-0.330***	-0.508***	
		(0.028)	(0.030)	
	Std. dev. of coefficient	0.016	0.002	
		(0.046)	(0.045)	
1500 mg.	Mean coefficient	-0.695***	-0.971***	
		(0.035)	(0.037)	
	Std. dev. of coefficient	0.561***	0.609***	
		(0.044)	(0.042)	
Health consequences – Baseline is "Unhealthy"				
Health neutral	Mean coefficient			2.611***
				(0.061)
	Std. dev. of coefficient			0.168**
				(0.067)
Healthy	Mean coefficient			3.771***
				(0.093)
	Std. dev. of coefficient			1.182***
				(0.056)
No. of observations		35,064	33,012	35,208
No. of respondents		974	917	978

Table 4.8: Average marginal effects calculated from mixed logit models with all normally distributed coefficients

	No health info.	Health info.	Explicit health info.
Price=6 euro	-0.084***	-0.076***	-0.095***
	(0.007)	(0.007)	(0.006)
Price=10 euro	-0.206***	-0.170***	-0.236***
	(0.014)	(0.014)	(0.015)
Time=30 min.	-0.014**	-0.021***	-0.042***
	(0.006)	(0.005)	(0.005)
Time=50 min.	-0.090***	-0.086***	-0.140***
	(0.013)	(0.012)	(0.010)
Taste=good	0.106***	0.082***	0.056***
	(0.006)	(0.007)	(0.004)
Taste=very good	0.172***	0.132***	0.125***
	(0.009)	(0.009)	(0.007)
Calories=1100 cals.	-0.072***	-0.098***	
	(0.004)	(0.005)	
Calories=1400 cals.	-0.156***	-0.185***	
	(0.009)	(0.011)	
Saturated fat=20 gr.	-0.034***	-0.037***	
	(0.006)	(0.006)	
Saturated fat=30 gr.	-0.079***	-0.098***	
	(0.006)	(0.007)	
Sodium=1200 mg.	-0.053***	-0.079***	
	(0.004)	(0.006)	
Sodium=1500 mg.	-0.113***	-0.152***	
	(0.007)	(0.009)	
Health conseq.=neutral			0.334***
			(0.009)
Health conseq.=healthy			0.454***
			(0.013)
No. of observations	35,064	33,012	35,208
No. of respondents	974	917	978

Notes: Omitted categories: Price: 2 euro, Time: 10 min., Taste: OK, Calories: 800 cals., Saturated fat: 10 gr., Sodium: 900 mg., Health: Unhealthy. Standard errors are obtained by using 100 bootstrap iterations and reported in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 4.9: Average marginal effects calculated from mixed logit models by education groups

	No health info.			Health info.			Explicit health info.		
	Low educated	High educated	Δ	Low educated	High educated	Δ	Low educated	High educated	Δ
Calories=1100 cals.	-0.064***	-0.081***	-0.017	-0.091***	-0.108***	-0.017			
	(0.007)	(0.008)	(0.011)	(0.007)	(0.009)	(0.011)			
Calories=1400 cals.	-0.138***	-0.181***	-0.043**	-0.172***	-0.203***	-0.031			
	(0.012)	(0.016)	(0.020)	(0.013)	(0.017)	(0.020)			
Saturated fat=20 gr.	-0.025***	-0.051***	-0.026**	-0.037***	-0.043***	-0.006			
	(0.007)	(0.009)	(0.012)	(0.008)	(0.011)	(0.014)			
Saturated fat=30 gr.	-0.067***	-0.098***	-0.032**	-0.101***	-0.101***	0.000			
	(0.008)	(0.012)	(0.014)	(0.011)	(0.011)	(0.015)			
Sodium=1200 mg.	-0.046***	-0.068***	-0.022**	-0.083***	-0.071***	0.012			
	(0.006)	(0.008)	(0.010)	(0.007)	(0.009)	(0.011)			
Sodium=1500 mg.	-0.106***	-0.125***	-0.018	-0.157***	-0.132***	0.025			
	(0.010)	(0.013)	(0.016)	(0.011)	(0.013)	(0.017)			
Health=neutral							0.329***	0.338***	0.010
							(0.022)	(0.015)	(0.027)
Health=healthy							0.443***	0.479***	0.035
							(0.023)	(0.019)	(0.029)
No. of observations	22,248	12,816		21,960	11,052		24,768	10,440	
No. of respondents	618	356		610	307		688	290	

Notes: Omitted categories: *Calories*: 800 cals. *Saturated fat*: 10 gr. *Sodium*: 900 mg. *Health*: Unhealthy. Standard errors for the difference of the marginal effects are obtained using 500 bootstrap iterations. * p<0.1, ** p<0.05, *** p<0.01. Standard errors are in parentheses.

Table 4.10: Determinants of the value of health

	Base	W/ Income	W/ Future	W/ Health	Full	Base r/s
Education	0.118*	0.083	0.108*	0.129*	0.090	0.133*
	(0.065)	(0.066)	(0.066)	(0.076)	(0.076)	(0.076)
Age	0.020**	0.017*	0.020*	0.012	0.010	0.013
	(0.010)	(0.010)	(0.010)	(0.012)	(0.012)	(0.012)
Age-squared	-0.000*	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.0001)	(0.0001)	(0.000)	(0.000)	(0.000)	(0.000)
Male	-0.167***	-0.178***	-0.165***	-0.226***	-0.234***	-0.221***
	(0.060)	(0.060)	(0.060)	(0.068)	(0.068)	(0.068)
Diet	0.072	0.061	0.065	0.123	0.106	0.113
	(0.086)	(0.086)	(0.086)	(0.100)	(0.100)	(0.100)
Income		0.209***			0.184***	
		(0.060)			(0.069)	
Future			0.080		0.057	
			(0.060)		(0.069)	
Health				0.119	0.116	
				(0.086)	(0.086)	
No. of observations	978	978	978	725	725	725

Notes: Dependent variable is the vector of individual-specific coefficients on the attribute level "Healthy" from the mixed logit model, estimated for scenario III. Column "Base" refers to the baseline specification which includes demographic variables, education, age and gender, and diet status. Columns "W/ Income", "W/Future", "W/Health" add income, measure of future orientation and self-reported health status to the base specification, respectively. Column "Full" is the full specification which includes all controls at the same time. Since self-reported health is not available for everyone, column "Base r/s" estimates the base model for the sub-sample of respondents whose health status is available. See Section 4.4.2 for all variable descriptions. * p<0.1, ** p<0.05, *** p<0.01. Standard errors are in parentheses.

Table 4.11: Predictive validity of stated preferences for actual choices

	Fruit	Vegetables	Soda	Snacks	Candy	Cook at home	Ready-made	Take-away
Model I:								
SP Health	0.243***	0.245***	-0.140***	-0.246***	-0.106***	0.166***	-0.123***	-0.130***
	(0.039)	(0.039)	(0.036)	(0.037)	(0.036)	(0.037)	(0.037)	(0.040)
Pseudo R-sq	0.014	0.018	0.005	0.017	0.003	0.008	0.005	0.006
Model II:								
SP Health	0.228***	0.234***	-0.129***	-0.246***	-0.116***	0.167***	-0.143***	-0.136***
	(0.039)	(0.040)	(0.037)	(0.038)	(0.036)	(0.038)	(0.038)	(0.040)
SP Health*Education	0.041**	0.045**	-0.030	0.001	0.026	-0.002	0.049**	0.016
	(0.020)	(0.020)	(0.018)	(0.019)	(0.018)	(0.019)	(0.019)	(0.020)
Pseudo R-sq	0.016	0.020	0.005	0.017	0.003	0.008	0.007	0.006
No. of observations	978	978	978	978	978	978	978	978

Note: Dependent variables are the consumption frequency of fruits, vegetables, sodas and energy drinks, snacks, candies, home-cooked meals, ready-made meals, and take-away meals, respectively, measured in 6 categories – never, less than once a week, 1-2 times a week, 3-4 times a week, 5-6 times a week, everyday. The independent variable "SP Health" is the individual-specific coefficient on the attribute level "Health consequences=Healthy" estimated via mixed logit model among scenario III respondents. Model II additionally includes the independent variable "SP Health*Education" which is the interaction term between "SP Health" and "Education". Ordered probit coefficients are reported. Standard errors are presented in parentheses. * p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Table 4.12: Average marginal effects for the MTurk sample

	No health info.	Health info.	Δ	No time info.	Time info.	Δ
Price=6 euro	-0.142***	-0.129***	0.013	-0.142***	-0.118***	0.024*
	(0.010)	(0.017)	(0.019)	(0.010)	(0.011)	(0.013)
Price=10 euro	-0.279***	-0.238***	0.040	-0.279***	-0.244***	0.035
	(0.023)	(0.040)	(0.046)	(0.023)	(0.023)	(0.033)
Time=30 min.	-0.035***	-0.039***	-0.004	-0.035***	-0.042***	-0.007
	(0.007)	(0.009)	(0.012)	(0.007)	(0.007)	(0.010)
Time=50 min.	-0.164***	-0.134***	0.030	-0.164***	-0.184***	-0.020
	(0.022)	(0.019)	(0.029)	(0.022)	(0.022)	(0.032)
Taste=good	0.099***	0.083***	-0.016	0.099***	0.098***	-0.001
	(0.008)	(0.010)	(0.013)	(0.008)	(0.008)	(0.012)
Taste=very good	0.205***	0.209***	0.004	0.205***	0.224***	0.019
	(0.016)	(0.024)	(0.027)	(0.016)	(0.016)	(0.023)
Calories=1100 cals.	-0.040***	-0.069***	-0.029**	-0.040***	-0.050***	-0.010
	(0.007)	(0.011)	(0.013)	(0.007)	(0.008)	(0.010)
Calories=1400 cals.	-0.117***	-0.156***	-0.039**	-0.117***	-0.140***	-0.023
	(0.012)	(0.018)	(0.019)	(0.012)	(0.014)	(0.017)
Saturated fat=20 gr.	-0.024***	-0.029***	-0.005	-0.024***	-0.023***	0.001
	(0.008)	(0.011)	(0.014)	(0.008)	(0.007)	(0.012)
Saturated fat=30 gr.	-0.047***	-0.076***	-0.029**	-0.047***	-0.051***	-0.004
	(0.010)	(0.011)	(0.014)	(0.010)	(0.012)	(0.016)
Sodium=1200 mg.	-0.047***	-0.036***	0.011	-0.047***	-0.032***	0.015
	(0.007)	(0.006)	(0.010)	(0.007)	(0.006)	(0.009)
Sodium=1500 mg.	-0.088***	-0.090***	-0.002	-0.088***	-0.085***	0.002
	(0.009)	(0.011)	(0.014)	(0.009)	(0.008)	(0.013)
No. of observations	10,512	10,872		10,512	10,872	
No. of respondents	292	302		292	302	

Notes: Omitted categories: Price: 2 euro, Time: 10 min., Taste: OK, Calories: 800 cals., Saturated fat: 10 gr., Sodium: 900 mg. Estimates based upon mixed logit models with all normally distributed coefficients. Standard errors (in parentheses) are obtained by using 100 bootstrap iterations. * p<0.1, ** p<0.05, *** p<0.01.

Appendix 4.A First-order conditions

Associated with the Lagrangian (equation 4.6) we have the following conditions:

$$\begin{aligned}
\frac{\partial \Im}{\partial X_h} &= 0 \Rightarrow \\
\frac{\partial U}{\partial C_h} &= \mu \frac{p_{X_h}}{\partial C_h / \partial X_h} - \lambda(E) \frac{\partial H^*}{\partial C_h} \left[\frac{\partial U}{\partial H} - \mu\, w(E) \frac{\partial s}{\partial H} \right], \qquad (4.18)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \Im}{\partial \tau_{C_h}} &= 0 \Rightarrow \\
\frac{\partial U}{\partial C_h} &= \mu \frac{w(E)}{\partial C_h / \partial \tau_{C_h}} - \lambda(E) \frac{\partial H^*}{\partial C_h} \left[\frac{\partial U}{\partial H} - \mu\, w(E) \frac{\partial s}{\partial H} \right] \qquad (4.19)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \Im}{\partial X_u} &= 0 \Rightarrow \\
\frac{\partial U}{\partial C_u} &= \mu \frac{p_{X_u}}{\partial C_u / \partial X_u} - \lambda(E) \frac{\partial H^*}{\partial C_u} \left[\frac{\partial U}{\partial H} - \mu\, w(E) \frac{\partial s}{\partial H} \right], \qquad (4.20)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \Im}{\partial \tau_{C_u}} &= 0 \Rightarrow \\
\frac{\partial U}{\partial C_u} &= \mu \frac{w(E)}{\partial C_u / \partial \tau_{C_u}} - \lambda(E) \frac{\partial H^*}{\partial C_u} \left[\frac{\partial U}{\partial H} - \mu\, w(E) \frac{\partial s}{\partial H} \right] \qquad (4.21)
\end{aligned}$$

Equations (4.18) and (4.19) provide the first-order condition for healthy consumption (4.7). Similarly, equations (4.20) and (4.21) provide the first-order condition for unhealthy consumption (4.10).

Appendix 4.B Experimental design

Selection of attribute levels

All attributes have three evenly-spaced levels, with the objective of drawing a balance between cognitive burden on the respondents and flexibility while estimating attribute importance (Mangham et al., 2009). Selected levels should not only be realistic, but also their range should be as wide as possible in order to prevent respondents from ignoring certain attributes due to insufficient variation in their levels (Hensher, 2006).

For attribute "price", the lowest level, 2 Euros, is consistent with a cheap but still realistic home-cooked dinner. The average price per person for a home-cooked meal in a 2-person household is calculated as 2.37 Euro in the Netherlands (NIBUD, 2014). The middle category, 6 Euros, reflects either a more luxurious home-cooked meal or a ready-to-serve processed meal from the supermarket. According to the authors' calculations on the basis of the website of the largest Dutch supermarket Albert Heijn (www.ah.nl), processed meals at Dutch supermarkets typically cost between 3 and 6 Euros. The top-level, 10 Euros, would correspond to a home-cooked meal prepared with very luxurious ingredients or more typically to a standard take-away meal. Authors' calculations from the main Dutch food delivery website (www.thuisbezorgd.nl) reveal the average price of take-away or home-delivery meals in the Netherlands to be around 10 Euros.

We follow Kamphuis et al. (2015) while determining the levels for attributes "taste" and "time". Kamphuis et al. (2015) have experimented with different levels for the attribute "taste" in pilot studies and concluded that setting the level of taste for an option to "non-tasty" causes all other attributes of this option to be ignored. This can be because individuals require a meal that is at least "OK" in terms of taste before starting to consider other attributes. Therefore, we choose "OK‘" as the minimum level of tastiness a meal can have in our experiment.

Kamphuis et al. (2015) divide attribute "time" into two components as travel time (to the grocery store) and preparation time. Their analyses have proven this distinction to be relatively unimportant. Therefore we decide to combine travel time and preparation time into one attribute and determine the levels on basis of the sum of travel and preparation time.

Regarding the health attributes we equate the lowest level for each attribute to its respective recommended intake level since the Dutch population is known to overconsume calories, saturated fat and sodium, on average (Van Rossum et al., 2011). The middle levels coincides roughly with the average intake levels, while the highest levels fall within the 95th percentile of the regular intake distribution. For "calories", we choose the levels 800 calories, 1100 calories, and 1400 calories. The daily recommended intake is

2000 calories for women, and 2500 calories for men. Given that dinner, on average, makes up for 36 percent of total calorie consumption (Van Rossum et al., 2011), the recommended caloric intake for dinner is about 720 calories for women and 900 calories for men. Taking an average implies an average recommended caloric intake of 800 for dinner, which comprises our lowest level. 1100 calories corresponds to an average dinner, while 1400 calories represents a high-calorie dinner, but is still realistic and corresponds roughly to the 95th percentile (total calories for men in the 95th percentile are around 3,700, of which 36 percent is between 1,300 and 1,400 calories for dinner Van Rossum et al., 2011).

For attribute "saturated fat", we choose the levels 10 gram, 20 gram, and 30 gram. The recommended daily intake of saturated fat is 24 gram for an average 2200 total calorie intake (U.S. Department of Health and Human Services and U.S. Department of Agriculture, 2005), yet saturated fat is highly overconsumed. At dinner, 42 percent of fat is consumed (Van Rossum et al., 2011). 42 percent of 24 gram implies around 10 gram, which forms our lowest level. To generate wide, evenly-spaced, but still realistic levels the remaining levels are set at 20 and 30 gram (the 95th percentile for Dutch men is 56 gram of saturated fat per day, Van Rossum et al., 2011). For "sodium" the daily recommended intake is currently 2400 milligram (U.S. Food and Drug Administration, 2012). Around 36 percent of salt is consumed for dinner, which implies (2400*0.36) around 900 milligram, which is set as our lowest level. Average daily intake for Dutch men and women is around 3500 milligram of sodium, which corresponds to around 1200 milligram at dinner, which comprises our middle level. Using evenly spaced levels, the highest level is correspondingly set at 1500 milligram, which is still realistic.

Design inputs and priors

In this subsection the experimental design is presented, with a discussion of the choices regarding input and priors.

Scenarios I and II We generate a design with two alternatives (alt1 and alt2), with in total 90 choice sets (rows) divided in 5 blocks. We use a Multinomial Logit Model (mnl) to generate the design. While ideally the design reflects the ultimate model to be estimated, the generation of 90 choice sets using a panel mixed logit specification with Bayesian priors is infeasible given the computational complexity (Bliemer & Rose, 2010, p. 732; Rose & Bliemer, 2013). Instead, we opt for the cross-sectional multinomial logit model with Bayesian priors to generate our design. While this seems like a large departure from a panel mixed logit model, numerous case studies and simulations show that there is only a slight loss in efficiency, and the performance of cross-sectional multinomial logit

is better than cross-sectional mixed logit if the true model is panel mixed logit (Bliemer & Rose, 2010).[24]

The algorithm minimizes the median D-error, uses row swapping, and we set the convergence criterion such that convergence is achieved if no improvement is found in 60 minutes. Since one iteration takes around 0.5 seconds, this implies that 60 minutes handles around 7200 iterations.

Table 4.B.1: Prior specification for Scenario I and II

Price – Base: 10 *Euros*	
2 Euros	$N(0.64; 0.068)$
6 Euros	$N(0.32; 0.034)$
Taste – Base: OK'	
Very Good	$N(0.06; 0.012)$
Excellent	$N(0.26; 0.052)$
Time – Base: 50 *minutes*	
10 minutes	$N(0.4; 0.043)$
30 minutes	$N(0.2; 0.021)$
Calories – Base: 1400	
800 calories	$N(0.2; 0.021)$
1100 calories	$N(0.1; 0.011)$
Sodium – Base: 1500 *milligram*	
900 milligram	$N(0.1; 0.011)$
1200 milligram	$N(0.05; 0.005)$
Saturated Fat – Base: 30 *gram*	
10 gram	$N(0.1; 0.011)$
20 gram	$N(0.05; 0.005)$

Notes: $N(\mu; \sigma)$ refers to a normal distribution with mean μ and standard deviation σ.

Since we have an unlabeled design, all parameters are generic across the alternatives, and there is no constant specified (Hensher et al., 2005, p. 151). The prior values of the parameters are set using bayesian priors using 1000 Halton draws from a normal distribution given in Appendix Table 4.B.1. The mean values are, where possible, based on Kamphuis et al. (2015). For price they use a continuous specification with coefficient -0.08. Using 10 Euro as the baseline category, this translates into a prior of 0.64 for 2 Euro and 0.32 for 6 Euro. The standard deviations of the priors are set such that the

[24] Bliemer & Rose (2010) explain this finding by noting that cross-sectional multinomial logit assumes all observations are from the same person, while cross-sectional mixed logit assumes that all observations are from distinct individuals. The panel mixed logit model is in between, where a single respondent answers a subset of the questions. In our case the subset of questions answered by a given respondent is pretty large, 18, so theoretically the panel mixed logit is better approximated by the cross-sectional multinomial logit rather than the cross-sectional mixed logit model.

order of the levels $U[price = 2] > U[price = 6] > U[price = 10]$ is maintained in 99.9 percent of the cases. The critical value of the normal distribution corresponding to 99.9 percent certainty is $z_{0.999} = 3.1$. Since it seems plausible that the standard deviation is proportional to the mean, for price this implies finding the minimum value k that satisfies

$$0.64 - 3.1 \times \frac{0.64}{k} > 0.32 - 3.1 \times \frac{0.32}{k} > 0$$

which gives $k = 9.4$. Therefore, the standard deviation for 2 Euro is set at $0.64/9.4 = 0.068$ and for 6 Euro it is set at $0.32/9.4 = 0.034$.

With respect to time, Kamphuis et al. (2015) used the separate attributes travel time (coefficient -0.02) and preparation time (coefficient 0.00). We have one attribute for time, and decided to take the average of the coefficients, -0.01. Taking 50 minutes as baseline category, 10 minutes has a prior mean of -0.4, while 30 minutes has a prior of -0.2. With respect to taste, we could directly incorporate the estimates of 0.26 and 0.06, respectively. The standard deviations are set in a similar way as for price.

The priors for the three attributes calories, sodium and saturated fat are more uncertain, and we cannot base them on previous literature. Our prior was that calories are deemed more important compared to saturated fat and sodium, since apart from the health consequences calories may be associated with a weight control motive (Cawley, 2004; Lakdawalla & Philipson, 2009). Therefore, the priors for calories are set slightly higher than the ones for sodium and saturated fat, with standard deviations set following the practice for price.

Scenario III Scenario III contains 4 attributes, each with 3 levels. This implies that the full factorial requires $3^4 = 81$ choice sets. Using 90 choice sets divided into 5 blocks of 18 implies that for scenario III we are able to identify the full factorial, if necessary. Since scenario III is very similar to the study by Kamphuis et al. (2015), we set the mean of our Bayesian prior values equal to their parameter estimates, while standard deviations are again set such that sign reversals are avoided, and the ordering of attribute levels is maintained, in 99.9 percent of the cases.

The priors for price, time, and taste are set equal to their levels in scenario I. The priors for health consequences are mean 1.17 and standard deviation 0.244 for "healthy" and mean 0.24 and standard deviation 0.05 for "health neutral". All standard deviations are set such that the logical order of the attribute levels is maintained in 99.9 percent of the cases, as explained for the coefficient of price in scenario I and II.

Introductory text

Scenario I In this questionnaire we try to understand your food choice behavior. Please respond as honestly as possible and avoid socially desirable answers.

Imagine it is a typical day and you are going to have a usual dinner at home. Depending on your habits, you can cook, you can order take out, or you can buy ready-made food from the grocery store. Eating out is no option. If you often visit a restaurant, we ask you to imagine a day where you would eat dinner at home. In the remainder of this questionnaire we will present you 18 times two meals, and we would like to know: "Which of these two meals would you eat regularly, i.e. at least twice a week?"

The two meals differ in terms of their taste, price, preparation time, calorie content, saturated fat content, and sodium content. These attributes are explained below.

1. Taste: How does the meal taste? Is it (i) ok, i.e. does not taste distinctly good or bad, (ii) good, i.e. tastes pretty good, or (iii) very good, i.e. tastes very good?

2. Price: How much does the meal cost per person? Think about the total cost of the ingredients if it is a self-made dish. Consider the total amount you pay if it is take-out or ready-made food. Cost per person can be (i) 2 Euros, (ii) 6 Euros, or (iii) 10 Euros.

3. Time: How much time does it take before the meal is on your plate – include both travel to the grocery store and preparation time? Total preparation time can be (i) 10 minutes, (ii) 30 minutes, or (iii) 50 minutes.

4. Calories: What is the energy content of the meal in terms of calories? The meal can contain (i) 800 calories, (ii) 1100 calories, or (iii) 1400 calories.

5. Saturated Fat: How many grams of saturated fat does the meal contain? The meal can contain (i) 10 gram, (ii) 20 gram, or (iii) 30 gram of saturated fat.

6. Sodium: How many grams of sodium does the meal contain? The meal can contain (i) 900 milligram, (ii) 1200 milligram, or (iii) 1500 milligram of sodium.

Assume all other characteristics of the meals are the same, e.g. they are equally filling, biological, fair-trade, contain equal amount of carbohydrates and proteins etc. Below you find an example choice set. You do not have to answer this one.

Scenario II Scenario II is identical to scenario I except that descriptions of health attributes – calories, saturated fat, sodium – are supplemented with some health information:

4. Calories: What is the energy content of the meal in terms of calories? Intake of too much calories can lead to overweight, cardiovascular diseases, and type II diabetes. The average recommended intake for calories is around 800 for a dinner meal. The meal can contain (i) 800 calories, (ii) 1100 calories, or (iii) 1400 calories.

5. Saturated Fat: How many grams of saturated fat does the meal contain? Eating too many grams of saturated fat is bad for one's health, and can lead to cardiovascular diseases and high cholesterol. The recommended intake for saturated fat at a dinner is at most 10 gram. The meal can contain (i) 10 gram, (ii) 20 gram, or (iii) 30 gram of saturated fat.

6. Sodium: How many grams of sodium does the meal contain? Salt contains sodium. Eating too much sodium is bad for health and can lead to high blood pressure and cardiovascular disease. The recommend intake for dinner is at most 900 milligram. The meal can contain (i) 900 milligram, (ii) 1200 milligram, or (iii) 1500 milligram of sodium.

Scenario III Scenario III follows scenario I yet replaces the three health attributes with the single attribute "health consequences":

4. Health consequences: How healthy is the alternative? The meal can be (i) healthy, i.e. associated with reduced risk of disease, (ii) health neutral, or (iii) unhealthy, i.e. associated with increased risk of disease.

Appendix 4.C Design for Amazon Mechanical Turk

The design of Amazon Mechanical Turk survey is the same as the original survey design as explained in the previous section, apart from some slight differences. First of all, since we are only interested in identifying the main effects, we generated an efficient design which picks the 18 – instead of 90 as in the original design – most informative choice sets, based on the criteria explained in Section 4.3.3. After careful consideration of the differences between the Netherlands and the United States in terms of price level and consumption habits, we decided to keep all attribute levels as they are, and only replaced euros with dollars.

We generated four different scenarios, each with 18 identical choice sets. The first two scenarios are identical to the original design where we provide respondents in the first scenario ("no information") with no information, and in the second scenario ("health information") with health information on the recommended daily amounts of calories, saturated fat and sodium. In the third scenario ("time information"), we supplemented all

Table 4.C.1: Example choice set MTurk Scenario III

	Meal A	Meal B
Price	2 dollars	6 dollars
Time	10 min	50 min
Taste	Very good	OK
Calories	1400 calories	1100 calories
Sodium	1500 mg	1200 mg
Saturated Fat	10 gram	30 gram

If you spend 10(50) minutes on preparing food, you cannot do anything else in those 10(50) minutes.

choice sets with the uninformative sentence "If you spend 10/30/50 minutes on preparing food, you can not do anything else in those 10/30/50 minutes". Appendix Table 4.C.1 illustrates an example choice set. The idea behind the time information scenario is making the time attribute salient by giving the respondents a sentence about time, but keeping the information content of the sentence at zero. Comparison of the three scenarios will tell us the driving source behind our main result: salience or information.

In the last scenario, we replaced the name of the attribute *taste* with *sensory Appeal* to see whether it makes a difference to describe sensory appeal of the meal in broader terms. We described this new attribute in the introductory text as the following:

1. Sensory appeal: How does the meal taste, look and smell? Is it (i) OK (not distinctly good or bad) (ii) Good (iii) Very good?

To monitor whether our respondents are indeed paying attention and not speeding through questions, we added two attention checks to our survey – one after the 9th choice set, and one, at the end, after the 18th choice set. The first attention check is shown in Appendix Table 4.C.2, and takes the form of a choice set with an objectively dominant alternative. Meal B is cheaper, tastier, and takes less time to prepare. Health related attributes are kept the same for both meals to avoid heterogeneity in dietary preferences leading to controversy over which alternative is dominant. Respondents who chose Meal A are not shown the rest of the survey and their responses have been discarded.

The second attention check tests whether respondents read the instructions, and goes as follows:

> *Recent research on decision making shows that choices are affected by context. Differences in how people feel, their previous knowledge and experience, and their environment can affect choices. To help us understand how people make decisions, we are interested in information about you. Specifically, we are interested in whether you actually take the time to read the*

Table 4.C.2: Attention check for Scenario I

	Meal A	Meal B
Price	10 dollars	6 dollars
Time	30 min	10 min
Taste	OK	Good
Calories	1100 calories	1100 calories
Sodium	1200 mg	1200 mg
Saturated Fat	20 gram	20 gram

directions; if not, some results may not tell us very much about decision making in the real world. To show that you have read the instructions, please ignore the question below ***about*** *how you are feeling and instead check only the "none of the above" option as your answer. Thank you very much.*

Please check all words that describe how you are currently feeling.

- *Interested*
- *Distressed*
- *Excited*
- *Upset*
- *Strong*
- *Guilty*
- *Scared*
- *Hostile*
- *Enthusiastic*
- *Proud*
- *Irritable*
- *Alert*
- *Ashamed*
- *Inspired*
- *Nerveous*
- *Determined*
- *Attentive*
- *Jittery*
- *Active*
- *Afraid*
- *None of the above*

Respondents who fail to choose "none of the above" or choose other other options besides none of the above are screened out of the survey and their responses are discarded.

At the end of the survey, we collected background information on the respondents' gender, race, level of education and income, household size.

MTurk respondents come from 190 different countries, India and United States being the two biggest pools. Thinking that the U.S., rather than India, is more similar to the Netherlands in terms of price levels and life styles, we chose to restrict our pool of respondents to individuals residing in the United States. To keep the original and MTurk samples as similar as possible, we did not let respondents below the age of 18 participate in our survey. Each respondent is randomly assigned to one of the four scenarios, where (s)he answers 18 randomly presented choice sets. We prevented the same individual from

participating in more than one scenario. Moreover respondents who fail an attention check were not allowed to come back again.

Appendix 4.D Additional Tables

Table 4.D.1: WTP estimates derived from mixed logit models with non-random price coefficient

	No health info.	Health info.	Explicit health info.
Time=10 min.	2.90	3.74	4.30
	(0.27)	(0.31)	(0.20)
Time=30 min	2.64	2.59	2.87
	(0.24)	(0.27)	(0.17)
Taste=good	4.29	3.77	1.94
	(0.23)	(0.25)	(0.15)
Taste=very good	6.84	6.23	4.16
	(0.32)	(0.33)	(0.19)
Calories=800 cals.	6.23	8.83	
	(0.32)	(0.44)	
Calories=1100 cals.	3.15	3.90	
	(0.21)	(0.25)	
Saturated fat=10 gr.	3.10	4.69	
	(0.24)	(0.32)	
Saturated fat=20 gr.	1.78	3.13	
	(0.20)	(0.25)	
Sodium=900 mg.	4.91	7.27	
	(0.30)	(0.42)	
Sodium=1200 mg.	2.35	3.26	
	(0.21)	(0.25)	
Health conseq.=neutral			11.94
			(0.33)
Health conseq.=healthy			17.52
			(0.49)
No. of observations	35,064	33,012	35,208
No. of respondents	974	917	978

Notes: Estimates obtained from a mixed logit model where attribute "price" enters linearly and has a fixed coefficient. The WTP is the ratio of the coefficient on the relevant attribute to the coefficient on the price attribute (e.g. Train, 2009). Omitted categories: Time: 50 minutes., Taste: OK., Calories: 1400 cals., Saturated fat: 30 gr., Sodium: 1500 mg., Health: Unhealthy. Standard errors are in parentheses.

Table 4.D.2: Average marginal effects calculated from mixed logit models by education groups

	No health info.			Health info.			Explicit health info.		
	Low educated	High educated	Δ	Low educated	High educated	Δ	Low educated	High educated	Δ
Price=6 euro	-0.079***	-0.088***	-0.009	-0.074***	-0.076***	-0.001	-0.093***	-0.090***	0.003
	(0.009)	(0.011)	(0.014)	(0.009)	(0.012)	(0.015)	(0.008)	(0.011)	(0.014)
Price=10 euro	-0.213***	-0.188***	0.025	-0.165***	-0.183***	-0.018	-0.246***	-0.227***	0.019
	(0.020)	(0.022)	(0.030)	(0.018)	(0.025)	(0.032)	(0.022)	(0.026)	(0.036)
Time=30 min.	0.000	-0.031***	-0.031***	-0.012	-0.036***	-0.024*	-0.040***	-0.046***	-0.006
	(0.008)	(0.009)	(0.012)	(0.009)	(0.010)	(0.014)	(0.006)	(0.008)	(0.010)
Time=50 min.	-0.063***	-0.115***	-0.052**	-0.060***	-0.127***	-0.066**	-0.117***	-0.190***	-0.073***
	(0.014)	(0.021)	(0.025)	(0.015)	(0.023)	(0.027)	(0.014)	(0.023)	(0.027)
Taste=good	0.094***	0.123***	0.029**	0.072***	0.101***	0.029*	0.050***	0.071***	0.020
	(0.007)	(0.010)	(0.012)	(0.011)	(0.010)	(0.016)	(0.018)	(0.008)	(0.020)
Taste=very good	0.159***	0.181***	0.022	0.121***	0.146***	0.025	0.111***	0.148***	0.036**
	(0.012)	(0.016)	(0.020)	(0.013)	(0.017)	(0.022)	(0.013)	(0.012)	(0.017)
Calories=1100 cals.	-0.064***	-0.081***	-0.017	-0.091***	-0.108***	-0.017			
	(0.007)	(0.008)	(0.011)	(0.007)	(0.009)	(0.011)			
Calories=1400 cals.	-0.138***	-0.181***	-0.043**	-0.172***	-0.203***	-0.031			
	(0.012)	(0.016)	(0.020)	(0.013)	(0.017)	(0.020)			
Saturated fat=20 gr.	-0.025***	-0.051***	-0.026**	-0.037***	-0.043***	-0.006			
	(0.007)	(0.009)	(0.012)	(0.008)	(0.011)	(0.014)			
Saturated fat=30 gr.	-0.067***	-0.098***	-0.032**	-0.101***	-0.101***	0.000			
	(0.008)	(0.012)	(0.014)	(0.011)	(0.011)	(0.015)			
Sodium=1200 mg.	-0.046***	-0.068***	-0.022**	-0.083***	-0.071***	0.012			
	(0.006)	(0.008)	(0.010)	(0.007)	(0.009)	(0.011)			
Sodium=1500 mg.	-0.106***	-0.125***	-0.018	-0.157***	-0.132***	0.025			
	(0.010)	(0.013)	(0.016)	(0.011)	(0.013)	(0.017)			
Health=neutral							0.329***	0.338***	0.010
							(0.022)	(0.015)	(0.027)
Health=healthy							0.443***	0.479***	0.035
							(0.023)	(0.019)	(0.029)
No. of observations	22,248	12,816		21,960	11,052		24,768	10,440	
No. of respondents	618	356		610	307		688	290	

Notes: Omitted categories: *Calories*: 800 cals. *Saturated fat*: 10 gr. *Sodium*: 900 mg. *Health*: Unhealthy. The standard errors for the difference of the marginal effects are obtained using 500 bootstrap iterations. * p<0.1, ** p<0.05, *** p<0.001. Standard errors are in parentheses.

Chapter 5

Behavioral Response to Chronic Health Conditions

Chronic diseases are the most common and costly of all the problems. Upon diagnosis, basic lifestyle changes such as quitting smoking, cutting back on alcohol, losing weight and getting physically active are strongly recommended by health care providers to alleviate future medical consequences. In this paper we use 8 biennial waves of the Panel Study of Income Dynamics (PSID) between 1999 and 2013 to study the short- and long-term evolution of smoking, alcohol consumption, physical exercise and weight status in response to a new diagnosis with high blood pressure, diabetes, cardiovascular disease and cancer. Our findings suggest that diagnosed individuals only partially adapt their lifestyles to recommended changes. For the majority of the chronic diseases that we study, people quit smoking and stop drinking after a diagnosis. However, we do not observe that they lose weight (except for diabetes) and increase their physical activity level.

This chapter is based upon:

Hullegie, P. and Koç, H. (2015). *Behavioral Response to Chronic Health Conditions.* (Mimeo). Erasmus University Rotterdam.

5.1 Introduction

This paper examines whether people change their health behaviors after having been diagnosed with a new chronic condition, like diabetes or hypertension. In 2012, an estimated 38 million deaths or 68 percent of all deaths were due to chronic diseases, making them the leading cause of death around the globe (World Health Organization, 2014). According to the World Health Organization, more than 40 percent of these deaths were premature, and most of the time preventable. Medical evidence demonstrates that mortality and morbidity from chronic conditions can be greatly reduced via lifestyle modifications like quitting smoking, limiting alcohol consumption, increasing physical activity, and changing dietary composition (U.S. Department of Health and Human Services, 1990; Mancia et al., 2013).

Given the benefits, a new diagnosis with a chronic condition comes with a considerable amount of information/advice on health promoting lifestyle changes. For example, individuals newly diagnosed with hypertension are advised by their doctors to lower their sodium intake, exercise regularly and maintain a healthy weight (American Heart Association, 2014b). Similarly, newly diagnosed diabetics are advised to lower their calorie and fat intake, quit smoking, decrease alcohol consumption, and increase physical activity (Franz et al., 2002).[1] Moreover, the information conveyed through doctors is personalized. It is directed specifically to the patient in a face-to-face consultation, and is tailored to his needs whenever possible (such as setting a target weight or making a smoking cessation plan). It is also repetitive due to regular contact with the health care system for check-ups. It has been demonstrated that better knowledge on the relation between health behaviors and health outcomes leads to a healthier lifestyle (Kenkel, 1991), therefore the recommendations highlight the role of health behaviors in shaping future health outcomes. Apart from lifestyle recommendations, a new diagnosis informs an individual about his true health status. For instance, after a heart attack, an individual may realize he is not as healthy as he thought and take this as a wake-up call for change. As a result of the diagnosis, an individual may also update his beliefs about the health consequences of his past health behaviors (K. V. Smith et al., 2001; Khwaja et al., 2006). Altogether, diagnosis of a new chronic illness can be thought of as a strong information treatment at the individual level which carries the potential to trigger a behavioral response.

However, besides information, there may be two additional factors that might influence health behaviors after a new diagnosis: (i) the physical constraints and physiological

[1] These lifestyle modifications are evidence based and endorsed by independent expert panels like the United States Preventive Services Task Force. However, there may be heterogeneity among doctors in recommendation behavior (Frank & Kunovichfrieze, 1995). We do not have information on what happens during a doctor-patient consultation.

changes arising from the condition and (ii) medication. For example, an individual who had a (severe) stroke or cancer may be physically limited in the amount of exercise that he can carry out. It is also well known that certain commonly prescribed drugs to treat chronic disease may induce weight loss/gain (Leslie et al., 2007). These alternative explanations may be more relevant for some health behaviors, such as BMI and exercise, than for others, such as smoking and alcohol consumption as the latter requires less health capital. Despite these alternative explanations, the existing literature, in particular Slade (2012) and Oster (2015), emphasize the role of information in generating behavior change after a new diagnosis.

Analyzing two repeated cross sections of the NHANES from 1976–1980 and 1988–1994, Kahn (1999) concludes that diagnosed diabetics are in general making healthier choices compared to diabetics who are not yet aware of their condition, highlighting the value of information. He finds that diagnosed diabetics smoke and drink less, consume fewer sugar products and drink more diet drinks than their non-diagnosed counterparts. Clark & Etile (2002) use the British Household Panel between 1991 and 1997 to study changes in smoking behavior in response to health changes. They find that smokers who experienced a deterioration in their own self-assessed health are more likely to quit or lower cigarette consumption in the future. However, there is no response when other smoking members of the household start to have health problems. The authors take this finding as a sign for the larger potential of personalized, targeted information in promoting change rather than providing impersonal, aggregate level information. Using the 1992 to 2008 biennial waves of HRS, Slade (2012) evaluates the role of a new diagnosis and medication in inducing behavioral change among diabetics 50 years and older. He finds that relative to pre-diabetics, i.e., non-diagnosed individuals with elevated levels of blood sugar, diagnosed diabetics initially increase exercise, lose weight, and curb smoking and drinking. However, the increase in exercise and the reduction in weight are only temporary. Oster (2015) employs household scanner data to investigate the effect of a diabetes diagnosis on food purchases. She estimates that households engage in minor, although statistically significant calorie reductions in response to a new diagnosis. During the month immediately after the diagnosis, individuals adjust their purchases in accordance with the doctor's advice: i.e. an increase in fruits and vegetables consumption and a decrease in unhealthy foods. In the longer term, the decrease in unhealthy food consumption persists, while the increase in healthy food consumption does not.

In this paper we study the short- and long-term evolution of smoking, alcohol consumption, physical exercise and weight status, in response to a new diagnosis with high blood pressure, diabetes, cardiovascular disease – heart attack, heart disease and stroke – and cancer. A key challenge in identifying the effect of a new chronic disease is the issue

of whether the diagnosis can be considered as an unanticipated event. J. P. Smith (1999) argues that much of the actual realization and especially the timing of the diagnosis is unanticipated. However, he also adds that, to increase confidence in such an analysis, predictors of new diagnoses should be included in the model. We therefore analyze the effects of a new diagnosis with a chronic disease in the context of an individual-level fixed effects model. The individual fixed effects help in controlling for permanent unobservable characteristics that determine who experiences an onset, such as family history, and (conditional on onset) who receives a diagnosis, such as health consciousness. We further control for time-varying characteristics, such as age, and pre-diagnosis health behaviors. The identifying assumption here is that, given the set of controls, diagnosis of a new chronic condition is an unanticipated health shock. Under this assumption, we evaluate whether people respond to a new diagnosis of a chronic disease by modifying their lifestyle. Although our aim is to identify the role of information in driving change, we acknowledge that it is practically very difficult to single out the effect of information. Therefore, as discussed previously, our results will be a mixture of the effect of (i) the information received after a diagnosis, (ii) the physical and physiological changes arising from the condition, and (iii) medication.[2]

Our data come from the biennial waves of Panel Study of Income Dynamics (PSID) between 1999 and 2013, a longitudinal survey of a nationally representative sample of the whole United States (U.S.) population. By contrast, most previous work used data from 50+ populations with the justification that chronic disease incidence is higher in this age group (Kahn, 1998; Falba, 2005; Slade, 2012). However, for many chronic conditions incidence among the younger population is comparable to that among the elderly (see Figure 5.1). For example, an almost equal share of respondents in our sample were diagnosed with diabetes when aged 40-49 and 50-59 years (around 25 percent). Alternatively, more than half (62 percent) of our respondents were already diagnosed with high blood pressure before turning 50. Furthermore, not engaging in behavior change might be more costly for the young and chronically ill due to forgone productivity and more years in the disease state. Health returns to lifestyle modifications might also be smaller for individuals who are diagnosed at a later age. Therefore, the first contribution of this paper is to expand the analysis to younger populations and observe the behavioral impact of a new diagnosis over the life-cycle. The second contribution comes from the variety of conditions studied. Several papers in the literature examined the behavioral

[2] Even in the case where one compares a group of individuals that has the condition and a diagnosis with a group that has the condition but not diagnosed yet, differences in health behaviors could be explained by the combined effect of information and medication. Moreover, the physiological changes experienced by the undiagnosed group may be less "severe" than those experienced by the diagnosed group. To isolate the effect of medication, one would ideally conduct an experiment which assigns medication status randomly among diagnosed patients. Such an experiment is obviously infeasible.

response to a diabetes diagnosis (Kahn, 1999; Slade, 2012; Oster, 2015). Looking at different conditions provides information about the type and extent of lifestyle changes in the face of different diseases. Such information can be useful in designing better targeted policies. Finally, the panel nature of our data allow us to study behavior change in the long-run. Much of the existing literature provides evidence based on cross-sectional data which can only inform us about the short-run. However, adopting a lifelong approach to chronic disease management is crucial for survival. Therefore, it is necessary to study the evolution of health behaviors in the long-term.

Our results suggest that individuals who are diagnosed with a chronic condition partially adopt recommended lifestyle changes. For the majority of the chronic diseases that we study, people quit smoking and stop drinking after a diagnosis. However, we do not find that they lose weight (except for diabetes) and increase their physical activity level. One possible explanation for this finding is that providing people with information about possible health outcomes is not always sufficient in generating behavioral change.

The remainder of this paper is organized as follows. Section 5.2 provides background information on chronic diseases and their management. Section 5.3 describes the data and construction of the sample. Section 5.4 illustrates our estimation method and identification strategy. Section 5.5 presents the results and section 5.6 concludes.

5.2 Background on Chronic Diseases and Chronic Disease Management

5.2.1 Background

High blood pressure (or hypertension) is a common and serious condition, in which the blood exerts too much pressure on the walls of the arteries. If left untreated, high blood pressure can lead to a heart attack, a stroke, kidney damage, and many other complications. High blood pressure affected one out of three adults in the U.S. in 2012 (Nwankwo et al., 2013). However, not everyone who carries the disease is aware of it, because, except in the most extreme cases, high blood pressure has no signs or symptoms. It is estimated that, in 2012, one out of five American adults had high blood pressure without being aware of it. Around 65,000 American deaths (2.5 percent of all deaths) in 2011 had high blood pressure as the leading cause (Mozaffarian et al., 2015). High blood pressure is estimated to have cost the American economy $46 billion in 2011 in terms of health care services, medication and missed days of work (Mozaffarian et al., 2015).

Diabetes is a metabolic disorder in which the body does not produce insulin at all (type I) or is not able to use insulin effectively (type II). Type I diabetes is usually

diagnosed in children and young adults, and accounts for 5 percent of all diabetes cases among adults. Since the body is unable to produce any insulin, management of type I diabetes mainly includes insulin injections. Type II diabetes is the most common type of diabetes and makes up almost all of the rest of the cases among adults. It is treated with changes in lifestyle, oral medications, and (in the severest case) insulin. If left untreated (or poorly managed) diabetes may affect several parts of the body and lead to serious complications like heart disease, stroke, blindness, kidney failure and lower-limb amputation. In 2012, one out of 10 individuals in the U.S. had diabetes (Centers for Disease Control and Prevention, 2014). Like hypertension, diabetes can also develop very slowly and, as a result, often remain undetected for years until serious damage to the body has already been done. The Centers for Disease Control and Prevention (2014) estimates that, as of 2012, one out of three people with diabetes in the U.S. are undiagnosed. In 2011, diabetes killed around 74,000 people (3 percent of all deaths) and was the seventh leading cause of death in the United States (Heron, 2015). The total cost of diabetes to American economy was $245 billion in 2012 (Mozaffarian et al., 2015).

Cardiovascular disease is the collection of diseases that involve the heart and blood vessels. The most common types are heart disease, heart attack and stroke. Cardiovascular disease is caused by the build-up of fat, cholesterol and other substances in the walls of the arteries, making it difficult for blood to flow through. An estimated one out of twenty American adults have a cardiovascular disease, excluding hypertension (Mozaffarian et al., 2015). It is the number one cause of death in the U.S., killing about 600,000 people every year which comprises around a quarter of all deaths (Heron, 2015). Cardiovascular disease is also a leading cause of disability in the United States (Autor & Duggan, 2006; Mozaffarian et al., 2015). The American economy spent $320 billion on cardiovascular disease (including high blood pressure) in 2011 (Mozaffarian et al., 2015).

Cancer is a group of diseases characterized by abnormal cells that grow uncontrollably and invade other parts of the body to destroy normal tissue. In 2012, around one out of twenty Americans had a history of cancer (American Cancer Society, 2014). Cancer is the second leading cause of death in the United States. In 2011, about 580,000 individuals died of cancer which comprises 23 percent of all deaths (Heron, 2015). The Agency for Healthcare Research and Quality (AHRQ) estimates that the total of all health care expenditures for cancer in the U.S. in 2011 were $88.7 billion (American Cancer Society, 2015). The National Cancer Institute predicts that, in 2020, the number of cancer survivors will reach 18.1 million (an increase of 30 percent compared to 2010), and the associated cost of cancer care will be $160 billion in 2010 dollars (National Cancer Institute, 2011).

5.2.2 Chronic Disease Management

The lifestyle modifications that are recommended to keep each of the above mentioned four chronic diseases under control are essentially the same: quit smoking, limit alcohol consumption, lose weight and be physically active.

Hypertensive patients and diabetics are advised to quit smoking to reduce their elevated risk of cardiovascular disease (Chobanian et al., 2003; Centers for Disease Control and Prevention, 2014). Among survivors of a heart attack or a cardiac surgery, the probability of death is reduced by at least one third among the ones who stop smoking relative to those who do not (Critchley & Capewell, 2003). Cancer patients are recommended to abstain from smoking because it increases the number and severity of treatment related complications, and the risk of disease progression (Hayashi et al., 1999; Johnston-Early et al., 1980; Kawahara et al., 1998).

Moderate alcohol consumption has been shown to have a protective effect on the heart (Opie & Lecour, 2007). Therefore, drinking is allowed among the chronically ill who choose to drink conditional that the amount is limited to at most one glass per day for women (10 gr. alcohol) and two glasses per day for men (20 gr. alcohol).

In all chronic diseases, patients are recommended to follow a healthy diet and maintain a healthy weight. Typically, a healthy diet limits consumption of salt, trans- and saturated fat, and emphasizes fruits, vegetables, low-fat dairy products and whole grains. Weight loss in overweight or obese patients with hypertension and diabetes have favorable effects on blood pressure, dyslipidemia (an abnormal amount of lipids, i.e. cholesterol and/or fat, in the blood) and blood glucose, and may reduce cardiovascular risk (Neter et al., 2003; Evert et al., 2013; Mancia et al., 2013). Both overweight and obesity are associated with increased mortality from cardiovascular disease (Prospective Studies Collaboration, 2009; Berrington de Gonzalez et al., 2010). Obesity is associated with recurrence among breast and prostate cancer survivors (Rock & Demark-Wahnefried, 2002; Freedland et al., 2004).

Finally, all individuals with a chronic disease are encouraged to engage in regular physical activity. For hypertensive patients, physical activity is shown to be inversely related with cardiovascular and all-cause mortality (Rossi et al., 2012). Regular aerobic physical activity such as brisk walking is recommended at least for 30 minutes on most days of the week (Chobanian et al., 2003). Regular physical exercise not only helps to lose weight but it also reduces long-term blood sugar levels and risk of complications in diabetics (Boulé et al., 2001; Sigal et al., 2004). The American Diabetes Association recommends aerobic and resistance training at moderate intensity on at least 2-3 times per week (Sigal et al., 2004). In the case of cardiovascular disease, the recommended level of exercise is mostly personalized. In general, for patients with low-risk, 30 minute aerobic

exercise training of moderate to vigorous intensity 3-5 times per week is recommended. Even in severe cases, small amounts of training can help patients keep their independence and counteract depressive symptoms (Mancia et al., 2013). Among breast and colorectal cancer survivors, post-treatment physical activity is shown to reduce the relative risk of recurrence, and disease-specific and all-cause mortality (Holmes et al., 2005; Haydon et al., 2006).

It should be noted that lifestyle modifications are also recommended to patients under medication. Following a healthy lifestyle may enhance the efficacy of the drug treatment or reduce the dosage (Chobanian et al., 2003; Mancia et al., 2013). Therefore medication and lifestyle changes should be seen as complements rather than substitutes.

5.3 Data and Descriptive Analysis

5.3.1 Data

Our analysis uses 8 biennial waves of the Panel Study of Income Dynamics (PSID) between 1999 and 2013. The PSID is a longitudinal survey of a nationally representative sample of the U.S. population. It covers nearly the entire life cycle of individuals, from early adulthood through old age, and collects rich data on demographic characteristics, economic variables, health outcomes and behaviors.

Key to our analysis is information on the diagnosis status and timing of the diagnosis (if applicable) of the four chronic diseases that we consider in this paper: high blood pressure, diabetes, cardiovascular disease – heart disease, stroke, heart attack – and cancer. The PSID has been collecting these data since 1999. Specifically, respondents are asked: "Has a doctor ever told you that you have or had ...". If the response is positive, then a follow-up question about the date of first diagnosis is asked. In the 1999, 2001, and 2003 waves, this information was obtained by asking "How long have you had this condition?" As of 2005 the question changed into "How old were you when you were first diagnosed with ...?" In Appendix 5.A we describe how we merged the information from these two questions to determine the date of diagnosis. The PSID additionally gathers data on health behaviors. Respondents are asked whether they smoke, drink and their frequency of exercise. Self-reported measures of height and weight are also collected.

In the PSID, the household head answers the questions on the diagnosis status and the timing of the diagnosis on behalf of himself and his spouse. J. P. Smith (2007) checks for possible differences in the accuracy of the response to the questions about the timing of the diagnosis when the household head answers for himself and for his spouse. He finds that household heads are more likely to give focal responses – like five, ten,

fifteen years ago – when they are asked about the diagnosis dates of their spouses. This tendency is greater for less serious conditions like hypertension and diabetes compared to more serious conditions like stroke or cancer. Given the importance of the accuracy of the timing information for our analysis, we therefore choose to restrict our sample to household heads only.

Before turning to a discussion of the behavioral response to the onset of a new chronic condition, we briefly look at trends in the incidence and prevalence of chronic diseases in the United States. To represent the American population as closely as possible, we base our calculations on all PSID household heads. Furthermore, we correct for inconsistencies in reporting behavior by defining a person as chronically ill in the year of onset (as computed) and all subsequent years. Finally, disease prevalence statistics are age-adjusted to the U.S. Census 2000 population by the direct method. Table 5.1 presents our findings for each of the PSID waves between 1999 and 2013. High blood pressure is the most common chronic condition in the U.S. across all years. Age-adjusted prevalence of high blood pressure among American adults rose continuously between 1999 and 2013, from 18 percent to 28 percent. A similar story holds true for diabetes. The share of diabetics in the American population increased almost steadily from 6 percent in 1999 to 9 percent in 2013. In comparison to other chronic conditions, high blood pressure and diabetes saw the largest increase (50 percent) in the age-adjusted prevalence during the period 1999–2013. In contrast, the prevalence of cardiovascular diseases – stroke, heart attack and heart disease – remained relatively stable. This is an interesting observation given the accompanying increase in the prevalence of the biggest behavioral risk factors, i.e. hypertension and diabetes. This is suggestive of the success in keeping these risk factors under control, via medication or behavioral change or both, and preventing heart related complications. Finally, we see a constant increase in the share of American adults diagnosed with cancer between 1999 and 2009, followed by a moderate decrease between 2009 and 2013.[3] A possible explanation for this increase in cancer prevalence is the success in treating cardiovascular diseases (Honoré & Lleras-Muney, 2006).[4]

Table 5.1 also reports the average age at first diagnosis, which has been significantly decreasing for all conditions considered. For example, in 1999 people became diabetic

[3] At first sight it may seem odd that we observe a decrease in age-adjusted prevalence rates given the way we construct them, i.e. once diagnosed, always diagnosed. This is possible due to compositional differences across PSID waves resulting from entry, for example by getting married with a PSID member, or exit, for example by death.

[4] Our estimates for the age-adjusted disease prevalence are quite similar to those documented by Andreski et al. (2009) who calculate the yearly prevalence of six specific conditions (stroke, hypertension, diabetes, cancer, heart attack and asthma) by using the PSID and the National Health Interview Survey (NHIS). The NHIS is the most widely used nationally representative health survey. It increases our confidence in the quality of the health data collected by the PSID to see that our estimates are similar to those calculated based on the NHIS.

at an average age of 53 in comparison to 50 in 2013. In another example, average age for a stroke was 62 in 1999, while it reduced to 53 in 2013. Figure 5.1 further shows that a considerable share of the PSID respondents got their first diagnosis of a chronic condition at their prime-age, i.e. 25-54 years old. Repercussions of these developments for the economy are large. The Milken Institute calculated that the total cost of the seven most common chronic conditions to the U.S. were $1.3 trillion in 2003. 80 percent of this amount results from output losses due to absenteeism and reduced productivity at work. If current trends in health behaviors and treatment continue, total costs are projected to reach $4.2 trillion in 2023 (DeVol et al., 2007).

The final set of rows in Table 5.1 documents trends in health behaviors. To begin with, we see a decline in the smoking rates over time, especially in 2011 and 2013. This finding is in line with Cawley & Ruhm (2011) who document a constant decreasing trend in smoking prevalence in the U.S. since 1974. However, we do not observe such a favorable trend in other health improving behaviors. For example, the share of overweight/obese individuals in the adult American population increased continuously from 49 percent in 1999 to 55 percent in 2013. According to World Health Organization, excess weight is the third most health damaging behavior in high-income countries, responsible for 8 percent of deaths (World Health Organization, 2009). It is also associated with high levels of morbidity as a large risk factor for many chronic diseases like diabetes, cardiovascular disease, high blood pressure etc. The share of individuals who exercise frequently, i.e. at least 3 times a week reduced from 52 percent in 1999 to 46 percent in 2007, but started to rise again afterwards, reaching 51 percent in 2013. Finally, we observe alcohol consumption gaining prevalence between 1999 and 2013. It should be noted that we do not distinguish between heavy versus light drinking where the latter may have health improving effects (Opie & Lecour, 2007). Therefore, it is difficult to argue with certainty about the overall health impact of the observed positive trend in alcohol consumption. The main conclusion from the discussion above is that in the U.S. the chronic disease burden is increasing and people are diagnosed with a chronic disease at much younger ages than they were 15 years ago. Meanwhile, results are mixed in the health behaviors front. We observe a health improving, negative trend in smoking prevalence, combined with a noticeable increase in the prevalence of overweight/obesity and decrease in physical activity.

5.3.2 Sample Selection

To analyze the behavioral impact of a new diagnosis, we construct a separate sample for each of the four chronic conditions that we study in this paper. Additionally, we create a "general" sample where we consider a respondent to be chronically ill if he is diagnosed

with at least one of those conditions. In this case we take the diagnosis date of the first diagnosed condition as the relevant date of diagnosis, i.e. the "diagnosis event". In all samples, we exclude respondents who fail to report their diagnosis status for any of the diseases in any of the years in the sampling period, as well as those who did not report their date of diagnosis despite being diagnosed. We do this in order to be able to account for co-morbidities.

Since we study possible changes in health behaviors after a diagnosis, in every sample we require respondents who have been diagnosed with the disease of interest to have at least one observation before and one after the diagnosis. This implies that, in each sample, respondents who were diagnosed with the disease of interest in 1999 or 2013 are dropped. We further require all respondents to have at least three observations. Finally we drop observations with missing covariates like gender and education, and missing outcomes.

5.3.3 Descriptive Statistics

Table 5.2 presents summary statistics of several demographic and health variables for the general sample. Separate statistics are calculated for respondents who have never been diagnosed with high blood pressure, diabetes, cardiovascular disease or cancer, and for respondents who were diagnosed with at least one of those conditions between 2000 and 2012. For the latter group, pre- and post-diagnosis statistics are reported. Table 5.2 shows that two thirds of the respondents in the general sample have never been diagnosed with a chronic condition. These respondents are more likely to be young, white, and have at least a college diploma in comparison to the chronically ill. In the pre-diagnosis period, a statistically higher share of the (to be) chronically ill were current smokers and had excess weight than the never diagnosed. They were also less likely to exercise and consume alcohol.

Columns 3 and 4 of Table 5.2 shows a comparison of the pre- and post-diagnosis periods among respondents who were diagnosed with at least one chronic condition during their time in the PSID. Individuals are most likely to be diagnosed with high blood pressure and diabetes. A considerable share (16 percent) is also diagnosed with cardiovascular disease.

The evolution of health behaviors in response to a diagnosis is mixed. We observe a tendency towards healthier behaviors on some dimensions, but not others. Respondents tend to quit smoking and refrain from alcohol after being diagnosed with a chronic condition. However, they stop with being physically active and put on weight.[5] While interpreting these findings, one should bear in mind that health behaviors like smoking

[5] All pre- and post-diagnosis differences in health behaviors are statistically significant at the 1 percent level.

and drinking do not require any health capital to carry out in contrast to exercise, for example. Therefore, it could be the case that the chronically ill find it easier to change their smoking and drinking habits while they may simply not be healthy enough to keep their existing exercise rhythm, let alone increase it, even if they want to. Weight gain could be a consequence of the decrease in the level of physical activity. It could also be a side effect of medication. For example, taking insulin is known to increase weight (Mäkimattila et al., 1999).

Table 5.3 documents the summary statistics for disease specific samples. Statistics reported under the columns headed by "never" are based on data from respondents who have never been diagnosed with the corresponding disease. The columns headed by "pre" and "post" present the pre- and post-diagnosis statistics for respondents who were diagnosed with the corresponding chronic condition between 2000 and 2012. The differences in demographics and pre-diagnosis health behaviors between those never diagnosed and the chronically ill that we observed in the general sample are also visible in the disease specific samples. Irrespective of the condition, people diagnosed with a chronic condition are older, more often non-white, and have lower family income and educational attainment compared to those who have never been diagnosed with that condition. Moreover, they are statistically more likely to smoke in the case of high blood pressure and cardiovascular disease, and be overweight or obese for all conditions except cancer, before diagnosis than those who never received a dignosis. In the post-diagnosis period, we see a significant reduction in the share of smokers for all conditions (e.g. from 26 percent to 17 percent among cancer patients). After the receipt of a diagnosis with any of the conditions, smoking prevalence goes even below the level among those who did not receive such a diagnosis. In the case of diabetes and cancer, such reaction to smoking is accompanied by a modest, but not significant, reduction in the prevalence of overweight/obesity after diagnosis. While all these are positive developments, we also see a large, significant drop in the prevalence of exercise. For all conditions, the chronically ill, who were already less physically active in the pre-diagnosis period than the never diagnosed, reduce their level of physical activity even further in the post-diagnosis period.

Finally, Table 5.3 reveals that there is large number of co-morbidities to each condition. For example, before being diagnosed, the (to be) diabetics already had a higher prevalence of high blood pressure (42 percent) and cardiovascular disease (15 percent) relative to those never diagnosed with diabetes. After the diagnosis, the prevalence of co-morbid conditions increases (to 73 percent for high blood pressure and 28 percent for cardiovascular disease in the case of diabetes). In fact, we can find several cases where prevalence doubles after having been diagnosed. These observations highlight the importance of controlling for co-morbid conditions in our models.

5.4 Empirical Implementation

A natural starting point to estimate the impact of a new diagnosis on health behaviors is to compare average health behaviors of individuals who experienced a diagnosis event to the average among all individuals who did not experience such an event. This comparison would provide an unbiased estimate of a new diagnosis only if receipt of it was randomly distributed in the population. However, as shown in Section 5.3.3, chronically ill people are systematically different from people who are not afflicted with a chronic disease in several aspects which make them more likely to experience the onset of a disease. They are, on average, older, less educated, and engage more often in unhealthy behaviors. They smoke more, exercise less and have a higher BMI. In addition, conditional on onset, individuals who get diagnosed may be systematically different than individuals who do not. The former group may be better educated, contact with the healthcare sector more often and more health conscious than the latter. Therefore, a simple comparison of health behaviors between the diagnosed and non-diagnosed populations does not provide an unbiased estimate of how health behaviors respond to the receipt of a new diagnosis.

However, by conditioning on a set of genetic, behavioral and socioeconomic factors – including past health behaviors and pre-existing conditions – J. P. Smith (1999) argues that it may still be possible to isolate the random element in the receipt of a new diagnosis. Therefore, we estimate a model which controls for pre-diagnosis health behaviors as well as fixed and time-varying individual characteristics:

$$Y_{it} = \alpha_i + \gamma_t + \beta' X_{it} + \sum_{k=-m}^{M} \delta_k D_{it}^k + \epsilon_{it}. \tag{5.1}$$

In Model 5.1, Y_{it} represents the binary outcomes of interest, α_i the individual fixed effect that helps in controlling for time-invariant factors – such as family history, race, education or health consciousness – that determine who experiences an onset and gets a diagnosis, and the γ's are time dummies for each calendar year in the sample period that capture the general evolution of health behaviors over time. The vector X_{it} consists of time-varying covariates like age, marital status, self-reported health, family income, and other diagnosed chronic conditions. In Model 5.1, D_{it}^k is an event time indicator, with $D_{it}^k = 1$ if, in year t it has been k years since individual i was diagnosed (with the disease of interest). If k is negative, $D_{it}^k = 1$ means that year t precedes the year of diagnosis by k years. These dummy variables, $D_{it}^k, k = -m, \ldots, 0, \ldots, M$, jointly represent the event of diagnosis and the corresponding δ_k are the parameters of main interest. For those who are not diagnosed, for all k, $D_{it}^k = 0$. We assume that health behaviors older than 6 years prior to the diagnosis have no predictive power for a diagnosis event. Therefore, we group

all years that precede the diagnosis more than 6 years in the reference dummy. Our last event time dummy includes all years more than 6 years after diagnosis.[6]

Identification of the δ parameters rests on the assumption that, given the set of controls, having been diagnosed with a new chronic condition is a random health event. More specifically, conditional on all controls, a new diagnosis does not have to be an unanticipated event at the time of diagnosis but at more than 6 years before the diagnosis. This implies that health behaviors among the diagnosed would follow the same "post-diagnosis" trend with the non-diagnosed, had they not received a diagnosis.

The identifying assumption is violated if there are omitted variables that are correlated both with the outcome variable and the receipt of a new diagnosis. Individual fixed effects control for time-invariant variables like gender, race and socioeconomic status that are known to be important correlates of health behaviors, chronic disease onset and, given onset, receipt of a diagnosis (Carson et al., 2011; Mackenbach, 2006) . We also explicitly control for other important time-varying correlates, such as age, insurance status and presence of other chronic conditions. Thus, our estimated impact of a new diagnosis on health behaviors will be biased only if there are unobserved, time-varying factors that are excluded from our model. In case the identifying assumption holds, δ_k gives us an estimate of how health behaviors would differ at k periods from the diagnosis relative to the counterfactual case of no diagnosis. The δ_k parameters can also be interpreted as the change in health behaviors in period k compared to the reference period more than 6 years before the diagnosis.

As explained in the introduction, the effect we identify might be a mixture of (i) the information provided by the diagnosis, (ii) the physical constraints and physiological changes arising from the condition, and (iii) medication. For example, diabetes drugs may induce weight loss/gain (Mäkimattila et al., 1999). Beta blockers, commonly used in treatment of high blood pressure, may reduce anxiety, which in turn may affect the need for smoking and alcohol consumption (Katzung et al., 2012). Patients who are under cancer treatment may experience weight loss/gain due to treatment (Doyle et al., 2006). Survivors of a severe stroke or cancer may be physically constrained in their capacity to increase exercise. However, it should be noted that behavior change is recommended despite such possible adverse effects. In fact, patients who suffer from side effects of medication and/or treatment are advised to take extra measures, such as dietary counseling, to keep their lifestyles in line with the guidelines. Exercise intensity and frequency can

[6] Because the PSID is a biennial survey, our event time indicators span two years. This does not track the evolution of the behavioral response in as much detail as a one-year interval. However, using one-year intervals would divide identification into two groups: the response in even numbered years would be identified from those individuals diagnosed in "PSID survey years" (1999, 2001, ...), whereas odd numbered years would be identified from those individuals diagnosed in "non-PSID survey years" (2000, 2002, ...). A two-year interval avoids this issue.

be tailored to the specific condition of each patient.

5.5 Results

We first discuss the short- and long-term behavioral response among household heads who have been diagnosed with any of the following chronic conditions: high blood pressure, diabetes, cardiovascular disease or cancer. More specifically, we evaluate the response to the first condition that is diagnosed. In this case the comparison group consists of household heads who are not diagnosed with any of the mentioned conditions during the sample period. Figure 5.2 plots the estimated trajectory of all four health behaviors against years since diagnosis (i.e. the δ-coefficients from model (5.1); Appendix Table 5.B.1 presents the coefficient estimates).

Panel 5.2a shows that an average household head's propensity to smoke is about 2 percentage points lower immediately following his diagnosis compared to the case where he had not been diagnosed. Size of this reduction remains similar and statistically different from zero over the next five years, but it loses its statistical significance from the sixth year onwards. The results can be alternatively interpreted as a drop in the head's propensity to smoke by 2 percentage points during the first five years following the diagnosis relative to the period that precedes the diagnosis more than 6 years.

A slightly different picture emerges when we move to Panel 5.2b, which presents the evolution of alcohol consumption. There is a 3 to 4 percentage points decrease in the probability to consume any alcohol among diagnosed household heads in the aftermath of the diagnosis. However, in contrast to smoking, this reduction is not realized immediately. Only two years after the diagnosis, a statistically and economically significant decrease is observed in the probability to drink. There is no sign of recidivism as the reduction remains stable both in terms of magnitude and significance during the coming years.

The evolution of overweight or obesity, as shown in Panel 5.2c, reveals that chronically ill household heads tend to lose weight at an increasing magnitude as they move further away from the point of diagnosis. Compared to household heads who have not experienced a diagnosis event, prevalence of overweight/obesity is 2 percentage points lower among the chronically ill heads in the first 2 to 3 years following diagnosis. Weight loss continues in the long-run with the probability of being overweight/obese going down by almost 3.5 percentage points in the 6 to 13th year of diagnosis.

Finally, Panel 5.2d illustrates that, in the face of a diagnosis, chronically ill household heads do not shift their behavior towards (more) frequent exercise (at least three times a week). Taken together with the results on weight status, this suggests that weight loss is achieved mainly through adjusting dietary behavior rather than taking up exercise.

Our findings suggest that, the chronically ill behave overall in compliance with the guidelines: They quit smoking and drinking, and lose weight. Diagnosis gives the chronically ill a stimulus to engage in more healthy behaviors to prevent future complications. The fact that we do not observe any statistically significant differences in any of the outcomes prior to diagnosis elevates confidence in our identifying assumption that, conditional on our controls, a new diagnosis with a chronic condition is an unexpected event.

5.5.1 Do Effects Differ by Chronic Condition

So far we have examined the evolution of health behaviors after a diagnosis with a chronic condition without making a distinction between conditions. However, it may be that a patient who is diagnosed with diabetes responds differently to the diagnosis than a patient who is diagnosed with high blood pressure, for example because not adhering to behavioral guidelines in case of diabetes may be (or perceived to be) much more consequential for the patient's health. Another reason why the response may differ across diseases is the difference in the frequency of feedback. For example, diabetics or hypertensive patients frequently receive information on the progression of their disease through doctor visits and self-monitoring, whereas patients with cardiovascular disease are likely to get updated about their current health status less frequently (American Heart Association, 2014a). In this section, we therefore examine heterogeneity in the response across chronic conditions.

High blood pressure

Figure 5.3 shows the evolution of health behaviors before and after a diagnosis with high blood pressure (see Appendix Table 5.B.2 for the coefficient estimates), which is by far the most common chronic condition among the American population (see Table 5.1). According to Panel 5.3a, there is hardly a response in terms of smoking. None of the estimated coefficients for the post-diagnosis period are statistically different from zero, in addition to being very small in size.

Panel 5.3b shows that, conditional on controls, prior to the diagnosis, prevalence of any alcohol consumption is no different between household heads who are going to be diagnosed with high blood pressure in the future and those who are not. The situation is similar also in the year of diagnosis or the years afterwards. Only as of the second year of diagnosis alcohol prevalence goes down among high blood pressure patients, by 3 percentage points. A possible explanation for the lack of an immediate response might be the binary nature of our measure of alcohol consumption, i.e. any alcohol or otherwise. High blood pressure patients may gradually reduce the amount of alcohol they consume, instead of quitting abruptly. The reduction in alcohol consumption persists in the fol-

lowing few years, but not throughout. As of the sixth year of diagnosis, the reduction in prevalence falls back to 2 percentage points, suggesting that some household heads take up drinking again.

Panels 5.3c and 5.3d reveal that individuals with high blood pressure do not succeed in making changes in their weight status and level of physical activity, either in the short- or long-run. This finding is consistent with Slade & Kim (2014) who fail to find an economically and statistically significant reduction in total energy intake among diagnosed hypertensive patients.

These results suggest that a high-blood-pressure diagnosis does not lead to substantial lifestyle modifications, with the exception of drinking. This is an interesting finding given the strength of the (medical) evidence on the effectiveness of behavior change on lowering blood pressure and cardiovascular disease risk (see Mancia et al. (2013) and the references therein). One possible explanation is that the success of anti-hypertensive medications in treating high blood pressure could give patients a false feeling of safety and deter them from making costly changes in their habits, despite the fact that medication should only be seen as a complement to behavior change rather than a substitute (Chobanian et al., 2003). In a recent article, Slade & Kim (2014) show that individuals who have high blood pressure for a long time and use anti-hypertensive medication consume more sodium than individuals who do not have high blood pressure but are at risk for it. Evaluating the impact of statin use on health behaviors among high cholesterol patients, Kaestner et al. (2014) report that statin use worsened diet, increased alcohol consumption and led to higher obesity rates. Moreover, as high blood pressure usually causes no symptoms and anti-hypertensive medication does not have serious side effects, it seems unlikely that the lack of behavioral change is a result of physical limitations arising from the condition or side effects of the medication. Hence, it can be inferred that high blood pressure patients choose not to adhere to doctor recommendations, which may indeed be due to the belief that the negative consequences of their non-compliance will be counterbalanced by medication.

Diabetes

Figure 5.4 plots the trajectory of the four health behaviors for household heads who are diagnosed with diabetes (Appendix Table 5.B.3 presents the coefficient estimates). According to panel 5.4a, the propensity to smoke is 4 percentage points lower among newly diagnosed diabetics relative to the situation had they not been diagnosed. In subsequent years, some household heads fall back to old habits and start smoking again but the prevalence of smoking remains significantly lower in the long-run.

We also see from Panel 5.4b that the likelihood of consuming any alcohol is 6 per-

centage points lower among household heads who have been diabetic for 2-3 years in comparison to those who have not been diagnosed with diabetes. As mentioned above, the lack of a significant response immediately following the diagnosis might be an artifact of the binary alcohol consumption variable. It seems that diabetics manage to abstain from alcohol throughout the whole post-diagnosis period in the PSID, the reduction in prevalence reaching to 7 percentage points during 6 to 13 years from diagnosis. It is interesting to observe that the behavioral response to drinking is stronger than the response to smoking, since the latter can lead to more serious complications than the former.

Panel 5.4c reveals that diagnosed diabetics are also able to significantly reduce their weight, both in the short- and the long-run. The likelihood of being overweight or obese falls by 4 percentage points among diabetics immediately after diagnosis compared to the case where had they not been diagnosed. The decrease in weight continues in the long-run, reaching to a 6 percentage point reduction in the prevalence of overweight/obesity among household heads with diabetes. This is likely to be due to a reduction in the amount of calories consumed, since diagnosed diabetics are not increasing the amount of exercise as can be seen from Panel 5.4d. Oster (2015) presents evidence supportive of this hypothesis by showing that, after a diabetes diagnosis, household heads engage in reductions in caloric intake sufficient to lose 6 to 11 pounds per year. Our findings are also broadly in line with Slade (2012), who finds a decline among diabetics in smoking, drinking and weight. However, according to his results weight loss is achieved only for a temporary period and individuals put on weight in the long-run. Finally, our findings are consistent with Kahn (1999), who observes that diagnosed diabetics are more likely to make dietary changes (including drinking) rather than behavioral changes (including smoking and exercise).

Some of the weight loss observed among the newly diagnosed diabetics can be attributed to physiological activities of the medication. Some diabetes drugs like metformin can induce weight loss. Therefore, it is not possible to say with certainty that a new diagnosis with diabetes serves as an impetus for losing weight among the diabetic population who are usually overweight or obese at the point of diagnosis. We are not aware of any relationship between diabetes medication and the remaining outcomes. It is also unlikely that diabetes interferes with fitness of the body to the extent that one cannot smoke, drink or exercise. Therefore, we argue that the reduction in smoking and drinking propensity among the newly diagnosed household heads can be attributed to the health information obtained during diagnosis. With respect to exercise, diabetics fail to follow the doctor's recommendation despite proven effectiveness of mild to moderate exercise in lowering blood sugar.

Cardiovascular disease

Figure 5.5 displays the estimated trajectory of the four health behaviors in response to a diagnosis with cardiovascular disease (see Appendix Table 5.B.4 for the coefficient estimates). According to Panel 5.5a, prevalence of smoking goes down by 4 to 5 percentage points during the first five years following the diagnosis. However, not all of this favorable reduction in smoking can be sustained in the long-run. Six to thirteen years later, household heads with a history of cardiovascular disease are only 2 percentage points less likely to smoke than household heads who does not have a cardiovascular disease history.

A similar story holds true also for drinking alcohol. The 4 to 6 percentage point reduction, observed over the next 5 years following the diagnosis, partially vanishes during 6 to 13 years after diagnosis (Panel 5.5b).

When it comes to weight status and regular exercise, Panels 5.5c and 5.5d reveal that survivors of cardiovascular disease neither lose weight nor take up regular exercise. In fact, if anything, they even reduce their level of exercise in the long-run. In the case of a heart attack or a stroke, it is expected that we do not directly observe an increase in the level of physical activity because patients may not be fit enough or may be under risk of an exercise-related complication. However, in the long-run patients are expected to recover and go on with their usual rythm. In contrast, we see a reduction in the share of household heads who frequently exercise after a diagnosis with a cardiovascular disease. While this can be due to non-compliance with the doctor's recommendations, it can also be an artifact of our measure regular exercise, i.e. moderate to vigorous exercise at least three times per week. Patients with a cardiovascular disease history may have their own exercise rhythms that can not be captured by our definition of exercise. Inability to lose weight can be a consequence of the exercise routine or the side effect of the medications used to treat cardiovascular disease. As we also mentioned in section 5.1, one does not need a large stock of health capital to smoke or drink alcohol. Therefore we argue that the doctor's advice generated the positive response with respect to these two behaviors in the short-run but in the long-run this response could not be sustained.

Cancer

Finally, Figure 5.6 illustrates the evolution of the four health behaviors after a cancer diagnosis (see Appendix Table 5.B.5 for the coefficient estimates). We see from Panel 5.6a that cancer patients manage to stay away from smoking, both in the short- and long-run after receiving a cancer diagnosis. It is worthwhile noting that the magnitude of the response in terms of smoking is stronger after a diagnosis with cancer or cardiovascular disease than with diabetes or high blood pressure. This observation can be explained with

reference to K. V. Smith et al. (2001) who find that smokers respond most dramatically to smoking-related health shocks. U.S. Surgeon General identifies cardiovascular diseases and several of the most common cancers, like lung and bladder, as conditions related to smoking, while diabetes and high blood pressure are not in this category.

When it comes to alcohol consumption we see from Panel 5.6b that cancer survivors also manage to abstain from alcohol in the short- and long-run. We see a downward trend in the share of household heads who consumes any alcohol in the post-diagnosis period, although magnitude of the reduction is never statistically different from zero except for 4 to 5 years after diagnosis. This is possibly due to the small number of cancer diagnoses observed in the sample.

According to Panel 5.6c, there is a 3 percentage points reduction in the share of overweight/obese household heads in the year of cancer diagnosis and the year thereafter. It is likely that weight loss is a side effect of cancer treatment. This possibility is supported by the observation that in the following years – which can be considered as recovery and post-recovery periods – cancer patients are as likely to be overweight/obese as household heads who do not suffer from cancer.

Finally, Panel 5.6d shows a significant rise in the prevalence of frequent exercise during the year of diagnosis or the year following it. However, in the coming years we see noticeable decline in regular physical activity. As for cardiovascular disease survivors, the absence of an effect for physical exercise may be due to a lack of behavior change, or, alternatively, be an artifact of our measure of regular exercise unable to capture the special exercise routines of cancer survivors.

5.5.2 Do Effects Differ by Educational Attainment or Age

In this section we check whether there is any heterogeneity in response by education and age to a new diagnosis with any of the four chronic conditions. Higher educated individuals may be more successful in implementing lifestyle changes for a number of reasons. First, education may improve one's understanding of the consequences of the disease and the necessity of behavioral modifications. Second, not engaging in behavioral change may be more costly for the higher educated due to lost productivity/wages. Third, higher educated individuals may be better able to comply with the recommendations (e.g., see Goldman & Smith, 2002 for the case of diabetes). However, according to our results (which are not shown), there is no difference in response between those with high education, i.e. college degree or more, and low education, i.e. less than college degree. This may be due to selection: if individuals with high education are generally healthier, then those highly educated who develop a chronic condition may have worse unobservable characteristics which make them less likely to adopt behavioral change. It

is interesting to look at heterogeneity by age because younger individuals may be more likely to make behavioral changes because the (health) returns to changes are higher for them as they are expected to live for more years with the disease. Furthermore, younger individuals may be physically and/or mentally better able to manage a chronic disease through lifestyle modifications. However, also in this case we do not find any difference in response behavior between those who are 50 years or older at diagnosis and those who are younger.

5.6 Conclusion

This paper quantifies the short- and long-term effects of chronic disease incidence on health behaviors. Our results indicate that individuals fall broadly in line with doctor's recommendations about quitting smoking and abstaining from alcohol after a diagnosis. However, we do not observe any changes in weight status (except for diabetes) and exercise behavior despite the large potential benefits.

It has been widely documented that smoking rates respond dramatically to new, aggregate level, information. The demand for cigarettes dropped significantly in the face of new scientific evidence linking smoking to lung cancer in the 1950s. Cigarette consumption further dropped as a result of health warning labels on packages, mass media anti-smoking campaigns, and other public health efforts that disseminate information about health risks of smoking (Chaloupka & Warner, 2000). However, it has also been shown that smokers are too optimistic when it comes to their own longevity expectations K. V. Smith et al. (2001); Khwaja et al. (2007). Therefore, one reason why we estimate a reduction in smoking prevalence could be that a diagnosis makes individuals more likely to internalize information about the harms of smoking.

The share of overweight/obese individuals went down only after a diabetes diagnosis. It is interesting that we do not find a similar reaction for cardiovascular disease and cancer, which are both considered to be more serious conditions than diabetes. This can be due to the fact that diabetics are confronted with their disease on a daily basis. They must engage in high quality self-management persistently, which includes balancing medication dosage with food intake and physical activity. Additionally, they need to measure their blood sugar levels regularly which provides them with high frequency feedback on whether they are managing the disease effectively. In the case of cardiovascular disease, individuals are not reminded of the consequences of their actions so frequently. Annual or biannual check-ups can bring individuals with a history of cardiovascular disease in contact with the medical system and update them about the current status of their disease and associated risks. For cancer survivors, behavioral change may play less of a role in comparison to

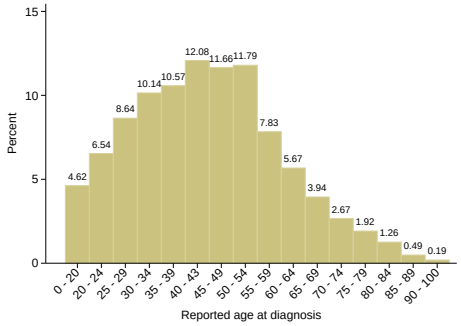
medical treatment in health promotion. Moreover, in the case of cancer, medical evidence that links nutrition and physical activity related lifestyle changes to improved survival is scarce. Therefore, doctors may be reluctant to encourage their patients to make these changes.

One caveat to our findings is the possibility that our estimates do not only result from new information about underlying health status and healthy lifestyles, but also from physical constraints arising from the condition, and medication. Although medications may have adverse effects that interfere with the intended lifestyle modifications, physicians aim to minimize such side effects by adjusting the prescribed drug and/or dosage. Moreover, for a given disease, different medications may have side effects working in opposite directions, which could cancel each other out. For example, while some medications for diabetes induce weight gain, others induce weight loss. Physical constraints may be more relevant in the case of cardiovascular disease and cancer. Patients may be too weak to exercise immediately following a diagnosis. However, after an initial recovery period, absence of a response is more likely to reflect either a failure in following the guidelines or the inability of our measure of regular exercise to capture personalized exercise patterns.

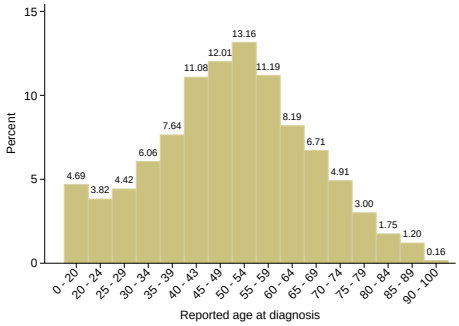
One potential limitation to our analysis is the use of self-reports on past diagnoses and its timing. Unfortunately, PSID respondents do not always consistently report whether they have been diagnosed with a certain chronic condition, nor do they always report the same year of onset. A possible source of such measurement error is that individuals with a lower socioeconomic status may have limited access to medical care and hence are less likely to be diagnosed. Moreover, it is well known that some individuals may report to have health problems as a justification for economic inactivity. It is unclear in which direction our estimates may be biased.

Figures

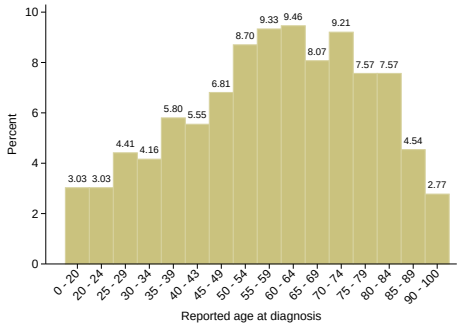
Figure 5.1: Distribution of age at diagnosis for each chronic condition



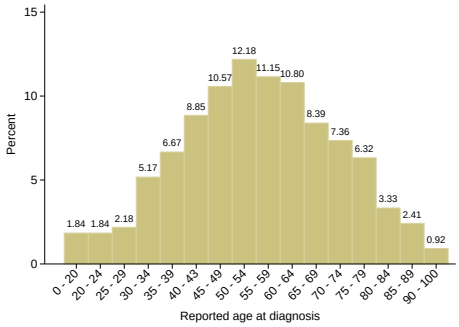
(a) High blood pressure



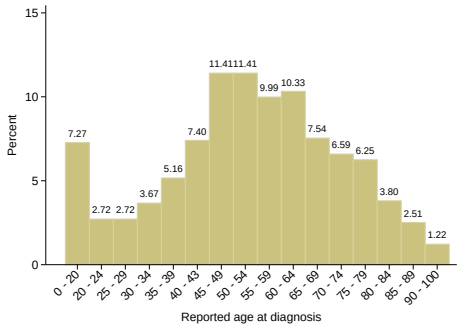
(b) Diabetes



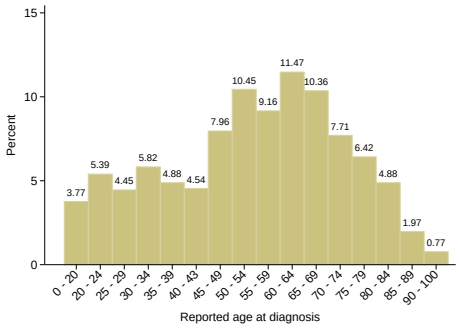
(c) Stroke



(d) Heart attack



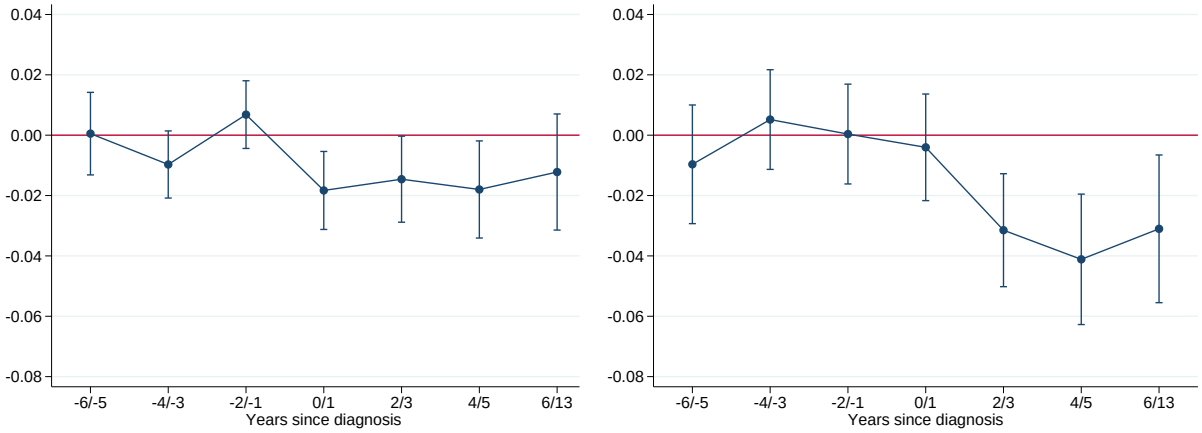
(e) Heart disease



(f) Cancer

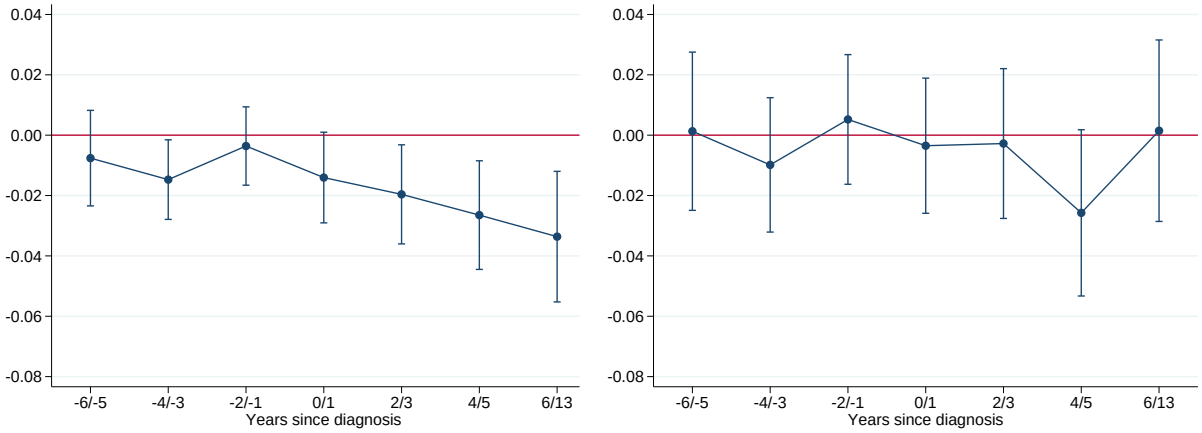
Notes: Each graph pools data from all years on the household heads who are diagnosed with the relevant condition.

Figure 5.2: Evolution of health behaviors after diagnosis of a chronic condition



(a) Smoking

(b) Any alcohol consumption

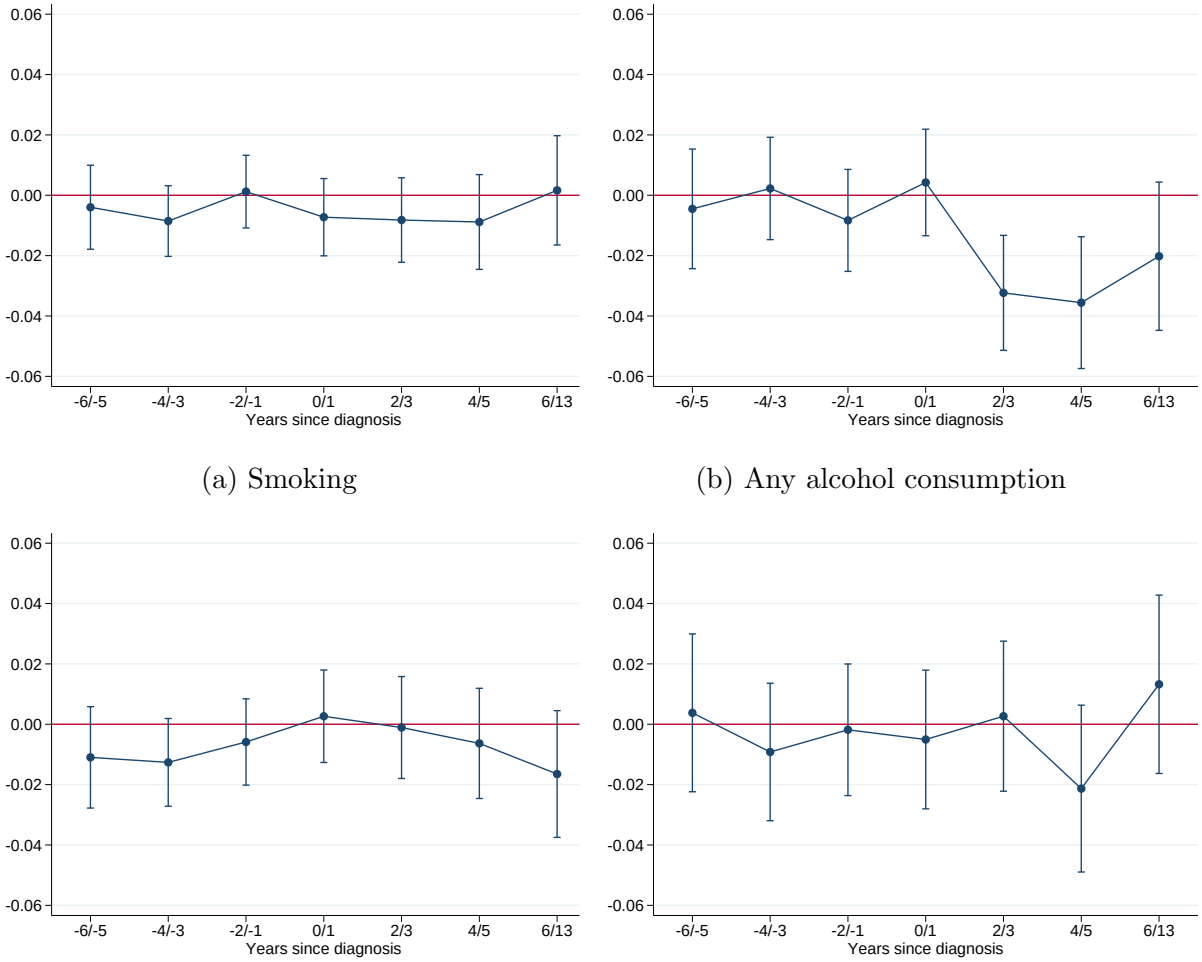


(c) Overweight/Obese status

(d) Frequent exercise

Notes: Figures show coefficients from fixed effects linear probability models. The models control for individual and year fixed effects, age, age-squared, marital status, quartiles of family income, health insurance status and self-reported health. Error bars show 90 percent confidence intervals.

Figure 5.3: Evolution of health behaviors after high blood pressure diagnosis



(a) Smoking

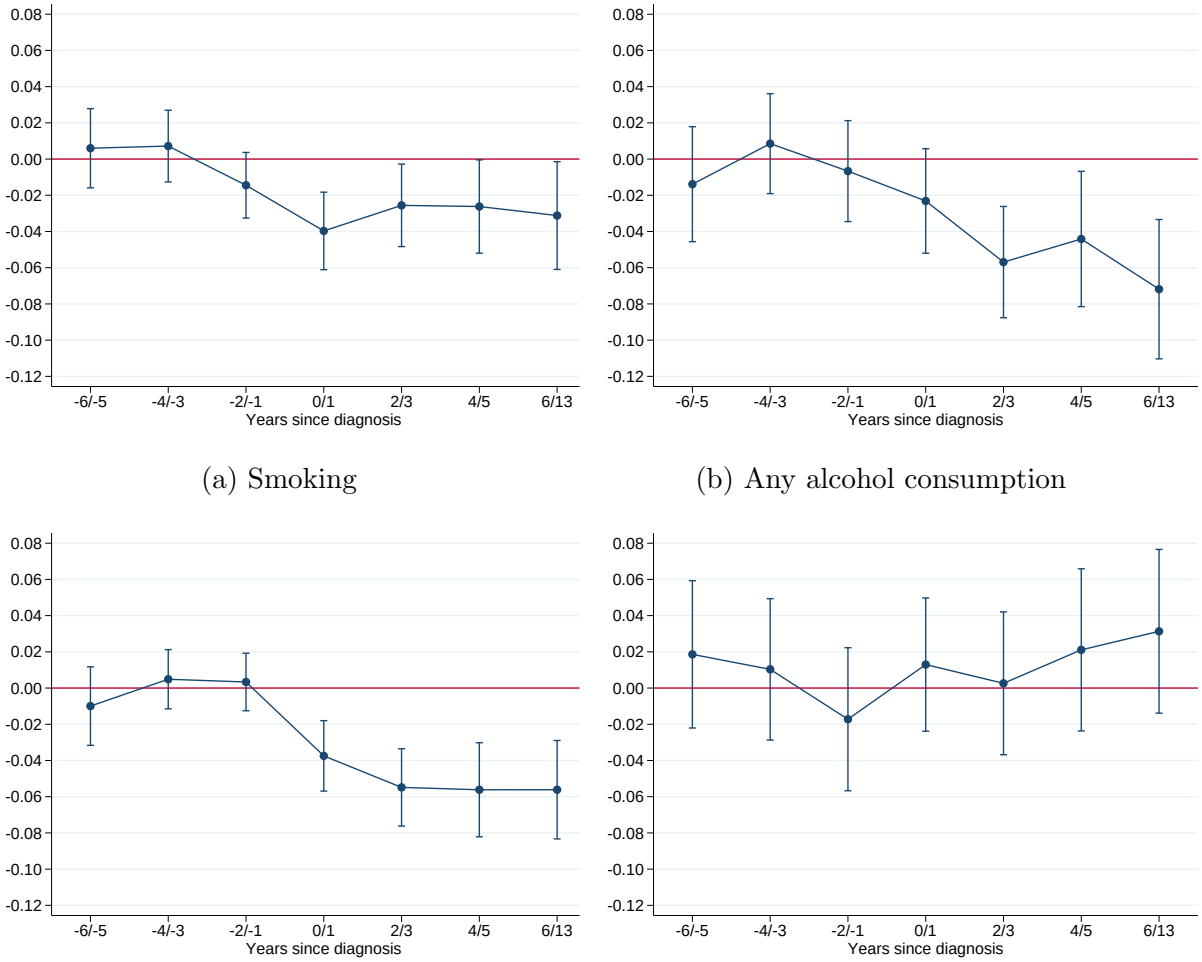
(b) Any alcohol consumption

(c) Overweight/Obese status

(d) Frequent exercise

Notes: Figures show coefficients from fixed effects linear probability models. The models control for individual and year fixed effects, age, age-squared, marital status, quartiles of family income, health insurance status, self-reported health, and current diagnoses of diabetes, cardiovascular disease and cancer. Error bars show 90 percent confidence intervals.

Figure 5.4: Evolution of health behaviors after diabetes diagnosis



(a) Smoking

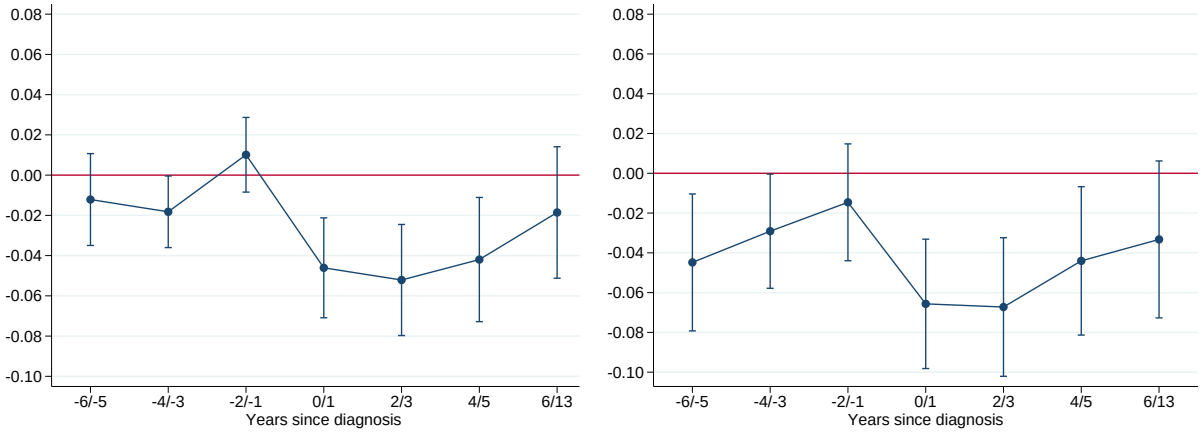
(b) Any alcohol consumption

(c) Overweight/Obese status

(d) Frequent exercise

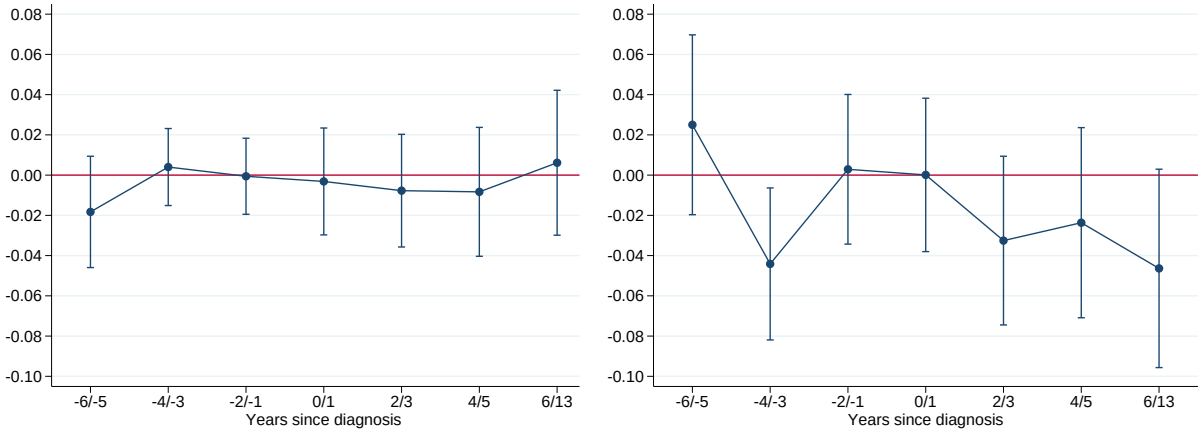
Notes: Figures show coefficients from fixed effects linear probability models. The models control for individual and year fixed effects, age, age-squared, marital status, quartiles of family income, health insurance status, self-reported health, and current diagnoses of high blood pressure, cardiovascular disease and cancer. Error bars show 90 percent confidence intervals.

Figure 5.5: Evolution of health behaviors after cardiovascular disease diagnosis



(a) Smoking

(b) Any alcohol consumption

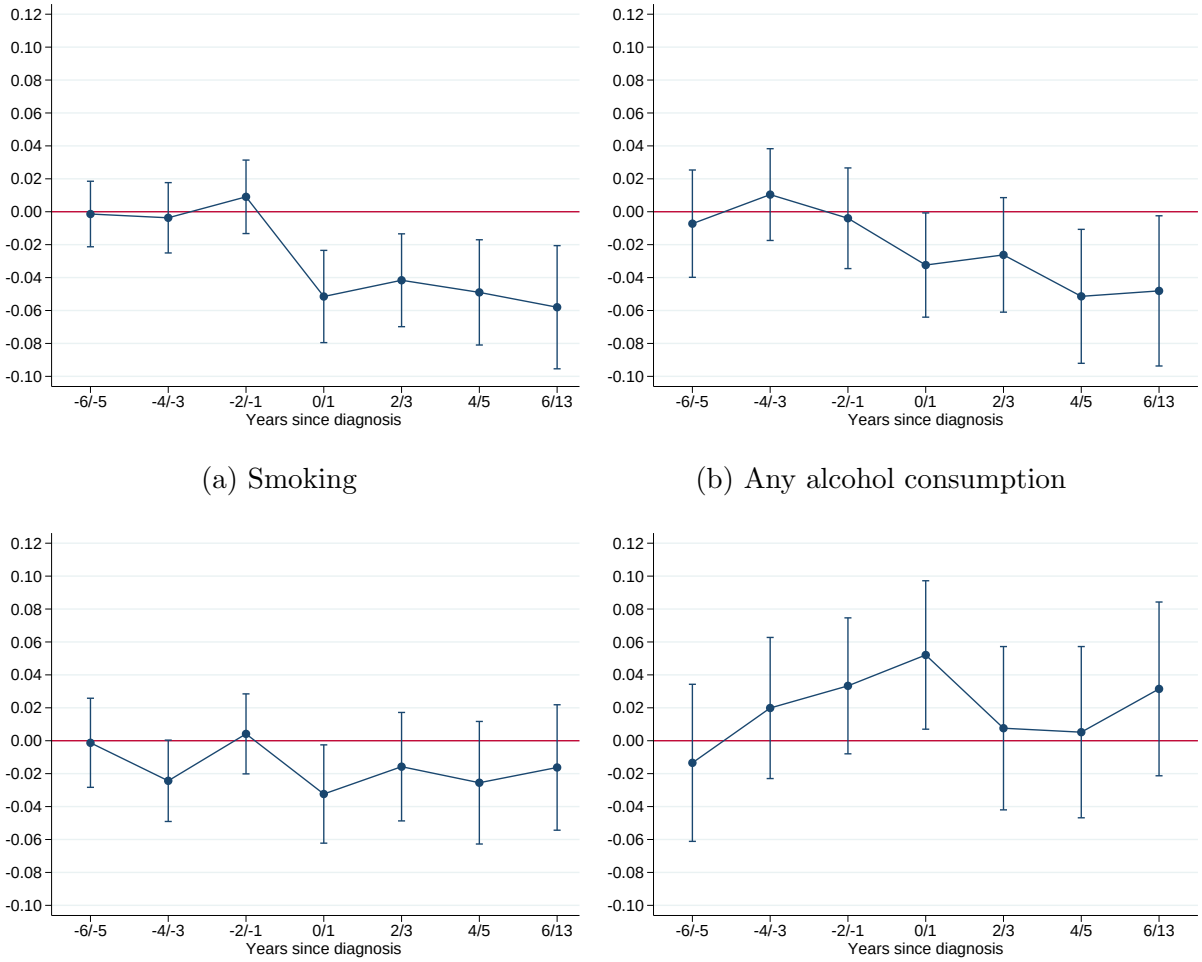


(c) Overweight/Obese status

(d) Frequent exercise

Notes: Figures show coefficients from fixed effects linear probability models. The models control for individual and year fixed effects, age, age-squared, marital status, quartiles of family income, health insurance status, self-reported health, and current diagnoses of high blood pressure, diabetes and cancer. Error bars show 90 percent confidence intervals.

Figure 5.6: Evolution of health behaviors after cancer diagnosis



(a) Smoking

(b) Any alcohol consumption

(c) Overweight/Obese status

(d) Frequent exercise

Notes: Figures show coefficients from fixed effects linear probability models. The models control for individual and year fixed effects, age, age-squared, marital status, quartiles of family income, health insurance status, self-reported health, and current diagnoses of high blood pressure, diabetes and cardiovascular disease. Error bars show 90 percent confidence intervals.

Tables

Table 5.1: Summary statistics

	1999	2001	2003	2005	2007	2009	2011	2013
Disease prevalence								
High blood pressure	0.25	0.29	0.32	0.34	0.36	0.39	0.40	0.40
Diabetes	0.08	0.09	0.11	0.11	0.12	0.12	0.13	0.13
Stroke	0.04	0.04	0.05	0.05	0.05	0.05	0.05	0.05
Heart attack	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.05
Heart disease	0.09	0.10	0.11	0.11	0.11	0.10	0.10	0.09
Cardiovascular disease	0.13	0.14	0.14	0.15	0.14	0.14	0.14	0.13
Cancer	0.05	0.06	0.07	0.07	0.08	0.08	0.08	0.08
Age of onset								
High blood pressure	47.36	46.72	45.83	45.25	44.54	43.87	43.41	43.00
Diabetes	51.33	51.12	50.50	49.98	49.40	49.05	48.39	47.56
Stroke	61.81	60.64	59.49	57.83	56.64	55.82	54.26	52.25
Heart attack	56.62	55.94	55.43	55.29	54.34	53.73	52.96	52.32
heart disease	55.03	54.27	53.45	52.77	51.95	51.47	50.60	49.62
Cardiovascular disease	54.61	53.60	52.88	52.01	51.12	50.56	49.39	48.38
Cancer	57.83	56.77	55.75	55.17	54.27	53.66	52.21	51.13
Health behaviors								
Currently smokes	0.25	0.25	0.25	0.24	0.23	0.23	0.22	0.21
Currently drinks any alcohol	0.62	0.63	0.63	0.63	0.63	0.65	0.65	0.65
Overweight or obese	0.66	0.68	0.69	0.71	0.71	0.72	0.73	0.72
Frequent exercise	0.69	0.68	0.67	0.62	0.62	0.63	0.64	0.65

Notes: In each year, summary statistics for disease prevalence and age at diagnosis are calculated based on the PSID household heads who are currently diagnosed with the relevant condition, except the ones who report the age at diagnosis for high blood pressure, diabetes, cardiovascular disease or cancer as younger than 20 or older than 90 in any of the years. We additionally include household heads who are not diagnosed with any of the mentioned conditions while calculating the summary statistics for health behaviors. Frequent exercise is defined as moderate to vigorous exercise at least three times per week.

Table 5.2: Descriptive statistics for the general sample by diagnosis status

			Diagnosed during PSID	
	Whole sample	Never diagnosed	Pre-diagnosis	Post-diagnosis
Demographics				
Age (in years)	41.75	38.32	43.30	50.76
Female	0.27	0.26	0.28	0.28
White	0.61	0.64	0.57	0.57
Black	0.33	0.31	0.37	0.37
Other race	0.06	0.05	0.06	0.06
Less than high school	0.15	0.14	0.18	0.18
High school or some college	0.56	0.55	0.58	0.56
College or more	0.29	0.31	0.25	0.26
Married	0.58	0.57	0.59	0.59
Good health	0.90	0.94	0.90	0.79
Family income (tens of thousands)	6.84	6.86	6.04	7.40
Health insurance	0.90	0.89	0.91	0.92
Disease prevalence				
Diabetes	0.04	0.00	0.00	0.20
High blood pressure	0.17	0.00	0.00	0.81
Stroke	0.01	0.00	0.00	0.05
Heart attack	0.01	0.00	0.00	0.05
Heart disease	0.02	0.00	0.00	0.10
Cardiovascular disease	0.03	0.00	0.00	0.16
Cancer	0.02	0.00	0.00	0.11
Health behaviors				
Currently smokes	0.24	0.24	0.28	0.22
Currently drinks any alcohol	0.67	0.71	0.64	0.61
Overweight or obese	0.67	0.61	0.73	0.77
Frequent exercise	0.67	0.70	0.66	0.61
Number of observations	36,053	22,626	5,806	7,621
Number of respondents	6,074	4,113	1,916	1,916

Notes: Never diagnosed refers to respondents who have not been diagnosed with high blood pressure, diabetes, cardiovascular disease or cancer prior to or during their time in the PSID. Frequent exercise is defined as moderate to vigorous exercise at least three times per week

Table 5.3: Descriptive statistics for disease specific samples by diagnosis status

	Diabetes			High blood pressure			Cardiovascular disease			Cancer		
	Never	Pre-	Post-	Never	Pre-	Post-	Never	Pre-	Post-	Never	Pre-	Post-
Demographics												
Age (in years)	43.54	48.59	56.06	40.43	43.67	50.83	42.49	52.94	59.02	43.59	54.47	60.67
Female	0.29	0.26	0.27	0.27	0.30	0.30	0.28	0.31	0.31	0.29	0.27	0.29
White	0.62	0.55	0.53	0.65	0.55	0.55	0.60	0.62	0.61	0.59	0.79	0.77
Black	0.33	0.37	0.39	0.30	0.39	0.39	0.34	0.34	0.34	0.35	0.18	0.20
Other race	0.05	0.08	0.07	0.05	0.06	0.06	0.06	0.04	0.05	0.05	0.03	0.03
Less than high school	0.15	0.20	0.22	0.14	0.19	0.19	0.15	0.29	0.24	0.17	0.14	0.16
High school or some college	0.56	0.57	0.55	0.56	0.58	0.56	0.56	0.50	0.55	0.56	0.55	0.53
College or more	0.29	0.22	0.23	0.30	0.23	0.25	0.29	0.21	0.21	0.27	0.31	0.31
Married	0.56	0.62	0.60	0.57	0.57	0.57	0.57	0.53	0.55	0.56	0.62	0.60
Good health	0.88	0.79	0.64	0.92	0.87	0.76	0.89	0.77	0.58	0.86	0.84	0.68
Family income (tens of thousands)	6.70	5.76	6.23	6.90	5.76	7.01	6.71	5.86	6.32	6.53	7.05	7.25
Health insurance	0.90	0.93	0.93	0.90	0.91	0.92	0.90	0.92	0.94	0.90	0.95	0.97
Disease prevalence												
Diabetes	0.00	0.00	1.00	0.04	0.07	0.19	0.08	0.17	0.29	0.10	0.12	0.23
High blood pressure	0.29	0.42	0.73	0.00	0.00	1.00	0.28	0.45	0.76	0.32	0.42	0.60
Stroke	0.03	0.04	0.10	0.01	0.02	0.06	0.00	0.00	0.29	0.03	0.05	0.10
Heart attack	0.04	0.06	0.12	0.02	0.02	0.07	0.00	0.00	0.34	0.04	0.08	0.14
Heart disease	0.07	0.11	0.21	0.04	0.05	0.14	0.00	0.00	0.68	0.08	0.14	0.24
Cardiovascular disease	0.10	0.15	0.28	0.05	0.07	0.19	0.00	0.00	1.00	0.11	0.20	0.31
Cancer	0.06	0.06	0.13	0.05	0.04	0.10	0.05	0.08	0.16	0.00	0.00	1.00
Health behaviors												
Currently smokes	0.24	0.25	0.19	0.24	0.28	0.22	0.24	0.30	0.23	0.24	0.26	0.17
Currently drinks any alcohol	0.67	0.58	0.49	0.69	0.63	0.61	0.67	0.59	0.51	0.65	0.62	0.57
Overweight or obese	0.66	0.87	0.85	0.62	0.74	0.79	0.68	0.71	0.73	0.69	0.68	0.66
Frequent exercise	0.67	0.62	0.56	0.69	0.66	0.61	0.67	0.65	0.58	0.66	0.68	0.62
Number of observations	42,297	21,83	2,516	27,126	5,443	7,175	41,444	1,932	2,241	44,821	1,499	1,490
Number of respondents	7,231	676	676	4,4842	1,851	1,851	7,081	599	599	7,622	429	429

Notes: For each condition, never diagnosed refers to respondents who have not been diagnosed with that condition prior to or during their time in the PSID. Frequent exercise is defined as moderate to vigorous exercise at least three times per week

Appendix 5.A: Measurement of Chronic Illness

In this appendix we describe how we construct the date of onset of a chronic disease. The PSID collects detailed information on disease diagnosis status, and timing of the diagnosis since 1999. Specifically, the PSID asks respondents: "Has a doctor ever told you that you have or had ..." for a number of chronic conditions. If the response is affirmative, then a follow-up question about the date of first diagnosis is asked. In the 1999, 2001, and 2003 waves this information was obtained by asking "How long have you had this condition?". The respondent could answer in days, weeks, months, and years. From 2005 and onwards, the question became "How old were you when you were first diagnosed with ...?".

We take the following approach to infer the date of onset. For each chronic condition and survey wave in the period 1999–2003, we first transform the number of days, weeks, months, and years that a respondent reports to have the condition into the number of days. This number is then subtracted from the interview date to determine the date and year of first diagnosis. For the 2005-2013 period, the difference between the current age and the age at first diagnosis, which roughly equals the number of years since diagnosis, is subtracted from the year of the interview.

At this point it should be noted that respondents who have stayed for more than one wave in the PSID may have reported different dates of onset across different waves. To come up with a single date of onset in such cases, we followed Chung (2013). For each chronic disease reported during the 1999–2003 surveys, if the reported number of days since the onset of the disease is less than 365 days, then this date is considered to be the "true" onset date. For the 2005–2013 surveys, if the age at the time of the interview is the same as the reported age of onset, then the corresponding calender year is considered to be the "true" onset year. At the end, if a respondent still has more than one onset date, than the earliest one is taken as the true onset date. For respondents who fall outside of this rule (i.e. reported number of days since diagnosis is greater than 365 or reported age of onset is different than the age during the interview), we assume that the mode of reported dates is the "true" onset date. If there is more than one mode, then we take the earliest one.

Appendix 5.B: Full Sets of Coefficient Estimates

Table 5.B.1: Behavioral response to diagnosis of a chronic condition

	Smoking	Any alcohol consumption	Overweight/Obesity status	Frequent exercise
Years since diagnosis				
-6/-5	0.001	-0.010	-0.008	0.001
	(0.008)	(0.012)	(0.010)	(0.016)
-4/-3	-0.010	0.005	-0.015*	-0.010
	(0.007)	(0.010)	(0.008)	(0.014)
-2/-1	0.007	0.000	-0.004	0.005
	(0.007)	(0.010)	(0.008)	(0.013)
0/1	-0.018**	-0.004	-0.014	-0.003
	(0.008)	(0.011)	(0.009)	(0.014)
2/3	-0.015*	-0.031***	-0.020**	-0.003
	(0.009)	(0.011)	(0.010)	(0.015)
4/5	-0.018*	-0.041***	-0.026**	-0.026
	(0.010)	(0.013)	(0.011)	(0.017)
6/13	-0.012	-0.031**	-0.034**	0.001
	(0.012)	(0.015)	(0.013)	(0.018)
Calendar years				
1999	base	base	base	base
2001	-0.026**	0.002	0.053***	-0.004
	(0.012)	(0.016)	(0.013)	(0.021)
2003	-0.052**	-0.011	0.099***	-0.036
	(0.021)	(0.030)	(0.024)	(0.039)
2005	-0.081***	-0.018	0.152***	-0.082
	(0.031)	(0.043)	(0.036)	(0.056)
2007	-0.115***	-0.031	0.199***	-0.098
	(0.041)	(0.057)	(0.047)	(0.075)
2009	-0.139***	-0.022	0.243***	-0.088
	(0.051)	(0.071)	(0.059)	(0.093)
2011	-0.176***	-0.022	0.288***	-0.094
	(0.061)	(0.085)	(0.071)	(0.111)
2013	-0.207***	-0.035	0.316***	-0.087
	(0.071)	(0.099)	(0.082)	(0.130)
Demographics				
Age (in years)	0.011**	0.006	0.009	0.008
	(0.005)	(0.007)	(0.006)	(0.009)
Age (in years) squared	-0.000	-0.000*	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Married	-0.023***	-0.036***	0.025***	-0.059***
	(0.008)	(0.010)	(0.010)	(0.013)
Low income	-0.000	-0.047***	-0.003	-0.010
	(0.007)	(0.009)	(0.008)	(0.011)
Low-middle income	-0.009*	-0.011*	-0.004	-0.008
	(0.005)	(0.006)	(0.006)	(0.008)
High-middle income	base	base	base	base
High income	-0.004	0.008	-0.003	0.016*
	(0.005)	(0.006)	(0.006)	(0.008)
Health insurance	-0.008	0.020**	0.010	0.029***
	(0.007)	(0.009)	(0.007)	(0.011)
Poor health	-0.005	-0.071***	-0.007	-0.182***
	(0.016)	(0.021)	(0.019)	(0.025)
Fair health	0.022***	-0.010	0.013	-0.100***
	(0.008)	(0.011)	(0.009)	(0.014)
Good health	0.022***	0.021***	0.025***	-0.055***
	(0.005)	(0.007)	(0.006)	(0.009)
Very good health	0.017***	0.016***	0.020***	-0.014*
	(0.004)	(0.006)	(0.005)	(0.007)
Excellent health	base	base	base	base
Constant	-0.062	0.512**	0.567***	0.570*
	(0.175)	(0.245)	(0.204)	(0.317)
Number of observations	36,053	36,053	36,053	36,053
Number of respondents	6,074	6,074	6,074	6,074

Notes: Coefficients from fixed effects linear probability models are reported. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses.

Table 5.B.2: Behavioral response to high blood pressure diagnosis

	Smoking	Any alcohol consumption	Overweight/Obesity status	Frequent exercise
Years since diagnosis				
-6/-5	-0.004	-0.004	-0.011	0.004
	(0.008)	(0.012)	(0.010)	(0.016)
-4/-3	-0.009	0.002	-0.013	-0.009
	(0.007)	(0.010)	(0.009)	(0.014)
-2/-1	0.001	-0.008	-0.006	-0.002
	(0.007)	(0.010)	(0.009)	(0.013)
0/1	-0.007	0.004	0.003	-0.005
	(0.008)	(0.011)	(0.009)	(0.014)
2/3	-0.008	-0.032***	-0.001	0.003
	(0.009)	(0.012)	(0.010)	(0.015)
4/5	-0.009	-0.036***	-0.006	-0.021
	(0.010)	(0.013)	(0.011)	(0.017)
6/13	0.002	-0.020	-0.016	0.013
	(0.011)	(0.015)	(0.013)	(0.018)
Calendar years				
1999	base	base	base	base
2001	-0.025**	-0.010	0.053***	-0.004
	(0.011)	(0.016)	(0.013)	(0.020)
2003	-0.051**	-0.031	0.096***	-0.035
	(0.020)	(0.029)	(0.023)	(0.037)
2005	-0.080***	-0.050	0.143***	-0.078
	(0.029)	(0.042)	(0.034)	(0.054)
2007	-0.112***	-0.070	0.186***	-0.097
	(0.039)	(0.056)	(0.045)	(0.072)
2009	-0.139***	-0.069	0.225***	-0.089
	(0.049)	(0.069)	(0.056)	(0.089)
2011	-0.175***	-0.075	0.267***	-0.090
	(0.058)	(0.083)	(0.067)	(0.107)
2013	-0.207***	-0.096	0.292***	-0.083
	(0.068)	(0.096)	(0.078)	(0.124)
Demographics				
Age (in years)	0.010**	0.011	0.011*	0.010
	(0.005)	(0.007)	(0.006)	(0.009)
Age (in years) squared	-0.000	-0.000**	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Married	-0.020**	-0.035***	0.025***	-0.059***
	(0.008)	(0.010)	(0.009)	(0.012)
Low income	-0.001	-0.040***	-0.006	-0.007
	(0.006)	(0.009)	(0.007)	(0.010)
Low-middle income	-0.008*	-0.009	-0.007	-0.009
	(0.005)	(0.006)	(0.005)	(0.008)
High-middle income	base	base	base	base
High income	-0.005	0.005	0.001	0.011
	(0.004)	(0.006)	(0.005)	(0.008)
Health insurance	-0.007	0.016*	0.012*	0.028***
	(0.006)	(0.008)	(0.007)	(0.010)
Poor health	0.003	-0.064***	-0.002	-0.187***
	(0.014)	(0.018)	(0.016)	(0.023)
Fair health	0.022***	-0.014	0.017**	-0.108***
	(0.007)	(0.011)	(0.008)	(0.013)
Good health	0.021***	0.019***	0.026***	-0.053***
	(0.005)	(0.007)	(0.006)	(0.009)
Very good health	0.017***	0.016***	0.019***	-0.014*
	(0.004)	(0.006)	(0.005)	(0.007)
Excellent health	base	base	base	base
Comorbid conditions				
Diabetes	-0.023*	-0.026	-0.070***	0.017
	(0.013)	(0.016)	(0.012)	(0.019)
Cardiovascular disease	-0.046***	-0.039**	-0.003	-0.033*
	(0.016)	(0.017)	(0.016)	(0.019)
Cancer	-0.057***	-0.029	-0.034**	0.005
	(0.017)	(0.019)	(0.017)	(0.023)
Constant	-0.054	0.355	0.519***	0.539*
	(0.171)	(0.245)	(0.199)	(0.313)
Number of observations	39,744	39,744	39,744	39,744
Number of respondents	6,693	6.693	6,693	6,693

Notes: Coefficients from fixed effects linear probability models are reported. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses.

Table 5.B.3: Behavioral response to diabetes diagnosis

	Smoking	Any alcohol consumption	Overweight/Obesity status	Frequent exercise
Years since diagnosis				
-6/-5	0.006	-0.014	-0.010	0.019
	(0.013)	(0.019)	(0.013)	(0.025)
-4/-3	0.007	0.009	0.005	0.010
	(0.012)	(0.017)	(0.010)	(0.024)
-2/-1	-0.014	-0.007	0.003	-0.017
	(0.011)	(0.017)	(0.010)	(0.024)
0/1	-0.040***	-0.023	-0.037***	0.013
	(0.013)	(0.018)	(0.012)	(0.022)
2/3	-0.026*	-0.057***	-0.055***	0.003
	(0.014)	(0.019)	(0.013)	(0.024)
4/5	-0.026*	-0.044*	-0.056***	0.021
	(0.016)	(0.023)	(0.016)	(0.027)
6/13	-0.031*	-0.072***	-0.056***	0.031
	(0.018)	(0.023)	(0.017)	(0.027)
Calendar years				
1999	base	base	base	base
2001	-0.028***	-0.007	0.051***	-0.007
	(0.010)	(0.014)	(0.012)	(0.019)
2003	-0.056***	-0.026	0.089***	-0.032
	(0.018)	(0.026)	(0.021)	(0.035)
2005	-0.086***	-0.045	0.129***	-0.077
	(0.027)	(0.038)	(0.031)	(0.050)
2007	-0.123***	-0.067	0.169***	-0.094
	(0.036)	(0.051)	(0.041)	(0.067)
2009	-0.150***	-0.064	0.206***	-0.084
	(0.045)	(0.063)	(0.052)	(0.083)
2011	-0.192***	-0.073	0.242***	-0.086
	(0.053)	(0.076)	(0.062)	(0.099)
2013	-0.225***	-0.090	0.262***	-0.076
	(0.062)	(0.088)	(0.072)	(0.115)
Demographics				
Age (in years)	0.011**	0.012*	0.012**	0.012
	(0.005)	(0.007)	(0.005)	(0.008)
Age (in years) squared	0.000	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Married	-0.019***	-0.029***	0.028***	-0.058***
	(0.007)	(0.009)	(0.008)	(0.011)
Low income	0.003	-0.031***	-0.005	0.001
	(0.006)	(0.008)	(0.007)	(0.009)
Low-middle income	-0.005	-0.005	-0.004	-0.003
	(0.004)	(0.006)	(0.005)	(0.007)
High-middle income	base	base	base	base
High income	-0.003	0.008	0.001	0.011
	(0.004)	(0.006)	(0.005)	(0.008)
Health insurance	-0.009	0.016**	0.009	0.029***
	(0.006)	(0.008)	(0.006)	(0.010)
Poor health	0.002	-0.064***	-0.020	-0.183***
	(0.012)	(0.016)	(0.013)	(0.019)
Fair health	0.017***	-0.014	0.017**	-0.107***
	(0.007)	(0.009)	(0.007)	(0.012)
Good health	0.023***	0.014**	0.023***	-0.051***
	(0.005)	(0.006)	(0.005)	(0.008)
Very good health	0.017***	0.013**	0.018***	-0.014**
	(0.004)	(0.005)	(0.005)	(0.007)
Excellent health	base	base	base	base
Comorbid conditions				
High blood pressure	-0.004	-0.010	0.007	-0.004
	(0.006)	(0.008)	(0.007)	(0.010)
Cardiovascular disease	-0.041***	-0.031**	-0.001	-0.013
	(0.013)	(0.015)	(0.013)	(0.017)
Cancer	-0.058***	-0.030*	-0.022	0.015
	(0.015)	(0.016)	(0.015)	(0.020)
Constant	-0.119	0.313	0.504***	0.504*
	(0.165)	(0.235)	(0.192)	(0.306)
Number of observations	46,996	46,996	46,996	46,996
Number of respondents	7,907	7,907	7,907	7,907

Notes: Coefficients from fixed effects linear probability models are reported. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses.

Table 5.B.4: Behavioral response to cardiovascular disease diagnosis

	Smoking	Any alcohol consumption	Overweight/Obesity status	Frequent exercise
Years since diagnosis				
-6/-5	-0.012	-0.045**	-0.018	0.025
	(0.014)	(0.021)	(0.017)	(0.027)
-4/-3	-0.018*	-0.029*	0.004	-0.044*
	(0.011)	(0.017)	(0.012)	(0.023)
-2/-1	0.010	-0.015	-0.001	0.003
	(0.011)	(0.018)	(0.011)	(0.023)
0/1	-0.046***	-0.066***	-0.003	0.000
	(0.015)	(0.020)	(0.016)	(0.023)
2/3	-0.052***	-0.067***	-0.008	-0.033
	(0.017)	(0.021)	(0.017)	(0.025)
4/5	-0.042**	-0.044*	-0.008	-0.024
	(0.019)	(0.023)	(0.019)	(0.029)
6/13	-0.019	-0.033	0.006	-0.046
	(0.020)	(0.024)	(0.022)	(0.030)
Calendar years				
1999	base	base	base	base
2001	-0.031***	0.002	0.049***	-0.007
	(0.010)	(0.015)	(0.012)	(0.019)
2003	-0.059***	-0.011	0.087***	-0.031
	(0.019)	(0.027)	(0.021)	(0.035)
2005	-0.088***	-0.022	0.126***	-0.072
	(0.028)	(0.039)	(0.031)	(0.051)
2007	-0.126***	-0.036	0.163***	-0.089
	(0.037)	(0.052)	(0.042)	(0.068)
2009	-0.155***	-0.028	0.199***	-0.081
	(0.045)	(0.065)	(0.052)	(0.084)
2011	-0.196***	-0.027	0.234***	-0.083
	(0.054)	(0.078)	(0.062)	(0.101)
2013	-0.229***	-0.040	0.253***	-0.073
	(0.063)	(0.090)	(0.072)	(0.117)
Demographics				
Age (in years)	0.011**	0.008	0.012**	0.010
	(0.005)	(0.007)	(0.005)	(0.009)
Age (in years) squared	-0.000	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Married	-0.019**	-0.035***	0.029***	-0.055***
	(0.008)	(0.009)	(0.009)	(0.011)
Low income	0.003	-0.036***	-0.004	-0.001
	(0.006)	(0.008)	(0.007)	(0.010)
Low-middle income	-0.005	-0.007	-0.004	-0.005
	(0.004)	(0.006)	(0.005)	(0.007)
High-middle income	base	base	base	base
High income	-0.003	0.010*	0.001	0.013*
	(0.004)	(0.006)	(0.005)	(0.008)
Health insurance	-0.010*	0.016**	0.013**	0.029***
	(0.006)	(0.008)	(0.006)	(0.010)
Poor health	0.007	-0.074***	-0.013	-0.181***
	(0.012)	(0.017)	(0.014)	(0.020)
Fair health	0.022***	-0.016	0.016**	-0.097***
	(0.007)	(0.010)	(0.007)	(0.012)
Good health	0.024***	0.014**	0.021***	-0.053***
	(0.005)	(0.007)	(0.005)	(0.008)
Very good health	0.017***	0.014***	0.018***	-0.017**
	(0.004)	(0.005)	(0.005)	(0.007)
Excellent health	base	base	base	base
Comorbid conditions				
Diabetes	-0.026**	-0.037**	-0.055***	0.024
	(0.011)	(0.014)	(0.010)	(0.017)
High blood pressure	-0.001	-0.008	0.006	-0.006
	(0.006)	(0.008)	(0.007)	(0.010)
Cancer	-0.049***	-0.032*	-0.020	0.018
	(0.016)	(0.017)	(0.016)	(0.020)
Constant	-0.126	0.467**	0.503***	0.529*
	(0.165)	(0.236)	(0.189)	(0.304)
Number of observations	45,617	45,617	45,617	45,617
Number of respondents	7,680	7,680	7,680	7,680

Notes: Coefficients from fixed effects linear probability models are reported. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses.

Table 5.B.5: Behavioral response to cancer diagnosis

	Smoking	Any alcohol consumption	Overweight/Obesity status	Frequent exercise
Years since diagnosis				
-6/-5	-0.001	-0.007	-0.001	-0.013
	(0.012)	(0.020)	(0.016)	(0.029)
-4/-3	-0.004	0.010	-0.024	0.020
	(0.013)	(0.017)	(0.015)	(0.026)
-2/-1	0.009	-0.004	0.004	0.033
	(0.014)	(0.019)	(0.015)	(0.025)
0/1	-0.051***	-0.032*	-0.032*	0.052*
	(0.017)	(0.019)	(0.018)	(0.027)
2/3	-0.042**	-0.026	-0.016	0.008
	(0.017)	(0.021)	(0.020)	(0.030)
4/5	-0.049**	-0.051**	-0.026	0.005
	(0.019)	(0.025)	(0.023)	(0.032)
6/13	-0.058**	-0.048*	-0.016	0.031
	(0.023)	(0.028)	(0.023)	(0.032)
Calendar years				
1999	base	base	base	base
2001	-0.030***	-0.001	0.053***	-0.011
	(0.010)	(0.014)	(0.011)	(0.019)
2003	-0.060***	-0.016	0.093***	-0.042
	(0.018)	(0.026)	(0.021)	(0.035)
2005	-0.091***	-0.033	0.135***	-0.092*
	(0.027)	(0.038)	(0.031)	(0.050)
2007	-0.129***	-0.048	0.177***	-0.116*
	(0.036)	(0.051)	(0.041)	(0.067)
2009	-0.157***	-0.042	0.217***	-0.114
	(0.044)	(0.064)	(0.051)	(0.083)
2011	-0.201***	-0.046	0.254***	-0.120
	(0.053)	(0.076)	(0.061)	(0.100)
2013	-0.234***	-0.061	0.275***	-0.115
	(0.062)	(0.088)	(0.071)	(0.116)
Demographics				
Age (in years)	0.011**	0.010	0.010*	0.016*
	(0.005)	(0.007)	(0.005)	(0.008)
Age (in years) squared	0.000	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Married	-0.018**	-0.030***	0.026***	-0.060***
	(0.007)	(0.009)	(0.008)	(0.011)
Low income	0.003	-0.033***	-0.007	-0.000
	(0.006)	(0.008)	(0.006)	(0.009)
Low-middle income	-0.005	-0.007	-0.006	-0.002
	(0.004)	(0.006)	(0.005)	(0.007)
High-middle income	base	base	base	base
High income	-0.003	0.008	0.000	0.014*
	(0.004)	(0.006)	(0.005)	(0.008)
Health insurance	-0.010*	0.019**	0.010	0.031***
	(0.006)	(0.008)	(0.006)	(0.010)
Poor health	0.007	-0.066***	-0.014	-0.184***
	(0.011)	(0.015)	(0.012)	(0.018)
Fair health	0.018***	-0.009	0.019***	-0.105***
	(0.007)	(0.009)	(0.007)	(0.012)
Good health	0.023***	0.016**	0.022***	-0.051***
	(0.005)	(0.006)	(0.005)	(0.008)
Very good health	0.016***	0.016***	0.018***	-0.015**
	(0.004)	(0.005)	(0.005)	(0.007)
Excellent health	base	base	base	base
Comorbid conditions				
Diabetes	-0.026**	-0.042***	-0.057***	0.019
	(0.011)	(0.014)	(0.010)	(0.017)
High blood pressure	-0.007	-0.009	0.007	-0.005
	(0.006)	(0.008)	(0.007)	(0.010)
Cardiovascular disease	-0.040***	-0.036**	-0.004	-0.001
	(0.013)	(0.015)	(0.013)	(0.017)
Constant	-0.147	0.387	0.575***	0.375
	(0.164)	(0.236)	(0.190)	(0.307)
Number of observations	47,810	47,810	47,810	47,810
Number of respondents	8,051	8,051	8,051	8,051

Notes: Coefficients from fixed effects linear probability models are reported. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parentheses.

Chapter 6

Conclusion

This thesis consists of four studies that seek to understand different aspects of prevention and health behaviors from a health economist's perspective.

The first study evaluates the impact of the Dutch breast cancer screening program on breast cancer and all-cause mortality. This program is a typical example of the population-based organized screening programs that were initiated in the late 1980s and early 1990s in many European countries to combat rising mortality rates from breast cancer. The results show that in the Dutch case the program has successfully reduced mortality from breast cancer. At the end of a follow-up period of 17 years, women who had a delay of at least 24 months in receiving the first invitation for a mammography from the program, and consequently received, on average, one less screen in a lifetime, are more likely to die from breast cancer than their same-aged peers who experienced a shorter delay in their first invitation. This study adds to a polarized literature divided between proponents of screening, who claim that regular screening saves lives, and opponents, who argue that there is not enough reliable evidence to make such a claim. Opponents additionally argue that, in the modern era, i.e. the period after the introduction of adjuvant systemic therapy, early detection has become less important mainly due to advances in cancer treatment (Bleyer & Welch, 2012).[1] By following the same cohort of women over time, the analysis in this study accounts for the role of improved treatment in reducing breast cancer mortality and concludes that screening itself contributes to a lower death rate from breast cancer.

The results are less favorable when it comes to the impact of the screening program on mortality from all causes. The study finds no statistically significant difference between the overall mortality rate of women who waited for at least 24 months before entering

[1] In the Netherlands, adjuvant therapy was introduced in the early 1980s for women with node-*positive* breast cancer. However, women with node-*negative* disease did not receive it before 2000 (Otto et al., 2003).

the program and that of women waited for less than 24 months. Then the question is which outcome, all-cause mortality or cancer-specific mortality, should we use to judge the effectiveness of cancer screening programs. Penston (2011) argues that all-cause mortality is a more reliable outcome as accuracy of disease specific mortality depends on the correct identification of the cause of death. Black et al. (2002) support this argument by claiming that in screening trials deaths from other causes are more likely to be falsely attributed to the target cancer among the screened group as they are more likely to be diagnosed with target cancer relative to the non-screened group. On the other hand, the target cancer is less likely to be assigned as a cause of death among the non-screened group since the target cancer had not been previously diagnosed. These two different misclassifications combined would exert downward bias on the estimated benefit of screening when cancer-specific mortality is used as an outcome. Moreover, use of all-cause mortality can accommodate possible harms of screening. Fatal complications may arise during an invasive procedure carried out to assess a suspicious screening result further. While such deaths from screening are not included in cancer-specific mortality figures, they are certainly relevant for the evaluation of benefits of screening. Another relevant point is mortality from treatment of early disease which mainly concerns the cases of over-diagnosis who would have never received the treatment that killed them had there be no screening. Gøtzsche & Olsen (2000) document that in RCTs from Malmo and Stockholm, both surgery and radiotherapy were more common among women screened for breast cancer. According to Early Breast Cancer Trialists' Collaborative Group (2000), use of radiotherapy in women with early detected breast cancer increases overall mortality in the long-run, mainly due to an elevated risk of cardiovascular death. As long as screening and treatment related deaths are not counted in cancer-specific mortality, using it as an outcome will overestimate the benefits of screening. However such deaths will be reflected in all-cause mortality figures, making overall mortality a more comprehensive outcome to look at.

On the other hand, deaths from even common cancers make up a very small portion of the total number of deaths. Therefore, Steele & Brewster (2011) argue that it would take impractically large RCTs to show a statistically significant reduction in all-cause mortality due to cancer screening. Of course, this is more of practical issue rather than an argument for the superiority of cancer-specific mortality over all-cause mortality as an outcome. Maybe a more appealing argument is that cancer screening is not designed to reduce overall mortality. If all medical interventions were to be evaluated with respect to their impact on all-cause mortality, then some of the medical interventions used today would perhaps be given a second thought as well.

Although all-cause mortality is admittedly a stringent outcome to look at while evalu-

ating cancer screening programs, it reflects the policy maker's ultimate goal of improving overall survival. Maybe what is worthwhile is to take additional outcomes into account, such as the rate of overdiagnosis and false-positives, stage at diagnosis and life expectancy. These statistics would provide the policy maker with a middle ground by giving her the chance to look beyond cancer-specific mortality without being as stringent as overall mortality. For example, improved stage at diagnosis, i.e. detection of cancer at an earlier stage, might be a beneficial outcome if treatment is less costly at earlier stages relative to later stages. Then depending on the cost of the program compared to treatment costs, a policy maker may be in favor of keeping the program despite its lack of effect on overall mortality. As for life expectancy, it can be that the program does not reduce the cumulative number of deaths at the end of a given time period but it extends life.

The second and third studies can be unified under the theme of educational disparities in health with a special focus on preventive care and health behaviors as potential mechanisms. The second study documents that there are no educational differences in breast cancer screening uptake in the Netherlands where a population-based organized screening program invites women of a certain age group for a free screening. In contrast to the Netherlands, there are large differences in take-up rate in the United States which does not operate such a program. The reason why higher educated women in the United States are more likely to get screened seems to be that they earn a higher income and have better insurance coverage relative to the lower educated. Perhaps contrary to expectations, educational differences in information possession (e.g. benefits and harms of mammograms) and risk perceptions (e.g. expected probability of getting breast cancer) seem to play only a minor role in generating the education gradient. Therefore, the study concludes that, based on evidence from the Netherlands and the United States, educational disparities in access to medical care rather than in awareness/knowledge is the main driver behind educational disparities in screening participation.

In contrast to the second study, the third study underlines the large role that differences in nutrition knowledge plays in driving differences in dietary behavior across education groups. Higher educated individuals have better knowledge than the lower educated on the adverse health consequences of unhealthy consumption and choose healthier options. Providing nutrition information (e.g. diseases associated with the overconsumption of calories and the recommended daily amount of calories) largely eliminates these education differentials suggesting that what makes the lower educated make poor food choices is lack of information. The finding that knowledge carries a larger importance in explaining the education gradient in food choice relative to breast cancer screening might be because information on breast cancer is more widespread than information on dietary behavior. Moreover there is more uncertainty surrounding healthy eating pat-

terns compared to breast cancer risk factors and screening methods. These factors might give the higher educated a comparative advantage as they are more efficient than the lower educated in collecting and processing information. Nonetheless, even after levelling off all knowledge differentials between education groups, the higher educated tend to choose healthier diets suggesting that they value their health more. Auxiliary analyses suggest that at least part of these differences in the valuation of health derives from higher incomes among the higher educated.

The finding that the lower educated lack relevant information while making dietary choices implies that there is room for government intervention. One option is that the government supplies the missing information or requires the food producers to do so. A widely considered alternative to information provision is introducing a tax on unhealthy food items to make consumers internalize the costs (e.g. health care costs associated with obesity, cardiovascular disease etc.) that they impose on themselves and on the society. A small counterfactual analysis suggests that while taxes on unhealthy food would make healthier options more attractive for lower educated individuals, the required taxes would have to be implausibly large, i.e. as high as 200 percent, to make these options equally attractive as a tasty, cheap, quick and unhealthy meal. Thus, the results from this study suggest that interventions such as calorie labeling, health warnings that aim to distribute nutrition knowledge carry more potential than introducing taxes in inducing the lower educated to eat more healthily.

The last study illustrates the potential of information provision in encouraging healthy behaviors for a particular subset of the population: the chronically ill. Based on the recognition that physicians provide information and advice, along with medical care, the study uses a new diagnosis with a chronic condition to evaluate the impact of information on reducing smoking and alcohol consumption, losing weight and engaging in physical activity. Following the physician's recommendation, individuals abstain from smoking and drinking, both in the short- and long-run, after being diagnosed with a chronic condition, irrespective of the severity of the condition. However, in terms of losing weight and getting physically active, there is hardly any response, if not a tendency towards the opposite of what is recommended. The behavioral response does not differ by education level possibly due a selection effect: individuals who get a chronic condition despite the protective effect on health of their higher education might have unobserved characteristics which separate them from the other higher educated individuals. Findings of this study should be interpreted with caution as they do not reflect the pure effect on health behaviors of information provision but a mixture of the effect of information, medication and physical constraints that come altogether with a diagnosis.

Although the findings of this thesis improved our understanding of prevention and

health behaviors, it is possible to identify several directions for future research. First, more research is required to shed light on the differing impact of the Dutch breast cancer screening program on breast cancer and all-cause mortality as found in the second study. An estimate of the level of over-diagnosis and false-positives in the program should give us an indication of the role of harms of screening in explaining this difference. Second, the third study shows that there are information differences by education in nutrition knowledge but these differences alone cannot fully account for the education gradient in dietary choice. What are possible explanations, in addition to the value of health, for the part of the gradient that is not related with information? Peer effects and the degree of self-control can be fruitful areas to explore. Third, why do the higher educated value their health more? Auxiliary analyses carried out in the third study should be expanded to answer this question. Finally, how can we encourage the chronically ill behave more healthily? What are the possible interventions that can induce these individuals to adopt a healthy diet and physical activity?

Samenvatting

Chronische ziekten vormen één van 's werelds grootste gezondheidsuitdagingen. Zo zijn ze de belangrijkste doodsoorzaak: In 2000 stierven wereldwijd 31 miljoen mensen aan een chronische ziekte, in 2012 waren dat er 38 miljoen (World Health Organization, 2014). Bijna de helft van deze sterfgevallen waren voorbarig, d.w.z. ze troffen mensen jonger dan 70 jaar oud (World Health Organization, 2014). Chronische ziekten zijn niet alleen verantwoordelijk voor een groot aantal sterfgevallen, ze hebben ook een significant negatief effect op de economie. Zo wordt geschat dat in 2003 chronische ziekten de economie van de Verenigde Staten 1.3 biljoen hebben gekost aan behandelingskosten (20 procent) en aan afgenomen arbeidsaanbod en productiviteit (80 procent; DeVol et al., 2007). In 2011 leidde erkenning van de stijgende problematiek ertoe dat wereldleiders hun handtekening zetten onder een VN-verklaring over niet-overdraagbare aandoeningen. In deze verklaring committeerden wereldleiders zich tot het opstellen en versterken van nationale beleidsplannen ter preventie en bestrijding van chronische ziekten (United Nations General Assembly, 2011).

In de geïndustrialiseerde wereld bestrijkt preventie een breed scala aan interventies, van informatiecampagnes die gezondheidsbevorderende activiteiten stimuleren tot interventies die erop gericht zijn ziektes in een vroegtijdig stadium op te sporen. Dit laatste type interventie wordt veelvuldig georganiseerd door middel van bevolkingsonderzoeken waarbij inwoners van een land zich gratis of tegen lage kosten kunnen laten screenen (voorbeelden zijn de mammografie voor borstkanker en de colonoscopie voor darmkanker). Het eerste deel van dit proefschrift – hoofdstukken twee en drie – roept twee vragen op over georganiseerde screening of bevolkingsonderzoeken in hoge-inkomenslanden in het algemeen: (i) zijn ze effectief in het terugdringen van sterfte als gevolg van chronische ziekten, (ii) kunnen zij de breed gedocumenteerde sociaal-economische verschillen in het gebruik van preventieve zorg verklaren (Fletcher & Frisvold, 2009; Mullahy, 1999; Cutler & Lleras-Muney, 2010), en zo ja, hoe?

Hoofdstuk twee evalueert het effect van het Nederlandse bevolkingsonderzoek naar borstkanker, zowel op borstkankersterfte als totale sterfte. De empirische strategie maakt gebruik van de geleidelijke uitbreiding van het programma over gemeenten tussen 1995 en

1997 – dat ertoe leidde dat sommige vrouwen van een bepaalde leeftijd hun eerste uitnodiging voor deelname eerder ontvingen dan andere vrouwen – om het causale effect van het bevolkingsonderzoek op borstkankersterfte en totale sterfte te schatten. De resultaten geven aan dat het programma erin is geslaagd om het aantal sterftegevallen als gevolg van borstkanker terug te dringen: aan het einde van 2011 zijn er onder vrouwen woonachtig in gemeenten die tenminste 24 maanden later begonnen met screenen 170 meer doden per 100,000 vrouwen. Deze vrouwen ontvangen gemiddeld genomen één uitnodiging voor het bevolkingsonderzoek minder. Aan de andere kant laat het onderzoek geen effect zien op de totale sterfte.

Hoofdstuk drie bestudeert of er verschillen bestaan in het gebruik van borstkankerscreening tussen mensen met een ander opleidingsniveau. Hierbij wordt gekeken naar Nederland – dat een bevolkingsonderzoek kent – en de Verenigde Staten – dat geen bevolkingsonderzoek kent. In Nederland bestaan er, onder vrouwen die worden uitgenodigd voor een gratis mammografie in het kader van het bevolkingsonderzoek, geen verschillen in mammografie-gebruik tussen opleidingsniveaus. In de Verenigde Staten bestaan deze verschillen echter wel. Verder onderzoek duidt erop dat het verschil in mammografiegebruik tussen vrouwen met ander opleidingsniveaus vooral gerelateerd is aan verschillen in inkomen, verzekeringsdekking en toegang tot medische zorg.

In het tweede deel van dit proefschrift – hoofdstukken vier en vijf – staan gezonde leefstijlen centraal. Volgens Centers for Disease Control and Prevention (2009) zijn roken, drinken, fysieke inactiviteit en een ongezond dieet verantwoordelijk voor een groot deel van arbeidsongeschiktheid, morbiditeit en sterfte als gevolg van chronische ziekten in de Verenigde Staten. Een vergelijkbaar beeld bestaat ook voor alle andere ontwikkelde landen (World Health Organization, 2009). Hierdoor dringt de vraag zich op waarom mensen zich, ondanks de schadelijke gevolgen, zo ongezond gedragen, en vervolgens hoe wij hen kunnen overtuigen gezonder te gaan leven? Ongezond gedrag komt met name voor onder mensen met een lage sociaal-economische status. (Cutler & Lleras-Muney, 2010) laten bijvoorbeeld zien dat mensen met een laag opleidingsniveau vaker roken, drinken, overgewicht hebben in vergelijking met hoger opgeleiden.

Een van de meest voorkomende verklaringen voor de waargenomen verschillen in gezond gedrag of het gebruik van preventieve zorg is gezondheidskennis. Grossman (1972) stelt dat hoger opgeleiden efficiënter zijn in het produceren van gezondheid. Een mogelijke reden hiervoor is de "allocatieve efficiëntie" hypothese, die stelt dat hoger opgeleiden een grotere kennis over gezond gedrag hebben en daarom beter in staat zijn om gezondere keuzes te maken. Een alternatieve verklaring is "productieve efficiëntie" hypothese, die stelt dat hoger opgeleiden efficiënter gebruik maken van een gegeven hoeveelheid kennis. Hoofdstuk drie onderzoekt of de allocatieve efficiëntie hypothese kan verklaren

waarom er geen gradiënt over opleidingsniveaus bestaat in het gebruik van een mammografie in Nederland, terwijl een dergelijke gradiënt wel bestaat in de Verenigde Staten. De veronderstelling is dat een bevolkingsonderzoek het belang van eventuele verschillen in informatie reduceert. Echter, uit de analyse blijkt dat conditioneren op subjectieve beoordelingen van het screeningsrisico en van het risico op borstkanker nauwelijks van invloed is op de gradiënt over opleidingsniveaus in de Verenigde Staten. Dit suggereert dat verschillen in informatie tussen opleidingsniveaus hooguit een kleine rol spelen in het begrijpen van de onderwijsgradiënt.

In de literatuur zijn slechts een paar studies te vinden die de allocatieve efficiënte hypothese expliciet testen, omdat data over zowel gezondheidskennis als levensstijlen schaars zijn (Kenkel, 1991; Lange, 2011). Hoofdstuk vier lost dit probleem op door gebruik te maken van een zelf ontworpen discreet-keuze experiment om zo de vraag te beantwoorden waarom hoger opgeleiden een gezonder dieet volgen dan lager opgeleiden. Deelnemers aan het experiment werden willekeurig toegewezen aan groepen die ieder een verschillende hoeveelheid informatie ontvingen. Hierdoor kan er onderscheid gemaakt worden tussen gezondheidskennis en gezondheidswaardering. De resultaten van het experiment duiden erop dat hoger opgeleiden een gezonder dieet volgen dan lager opgeleiden, omdat hoger opgeleiden een veel betere gezondheidskennis hebben. Op het moment dat lager opgeleiden informatie krijgen over de negatieve gevolgen van een ongezond dieet en informatie over de dagelijks aanbevolen hoeveelheden, zoals over het aantal calorieën, gaan ze ook gezondere keuzes maken. Extra informatie heeft echter nauwelijks gevolgen voor de keuzes van hoger opgeleiden. Desalniettemin blijven lager opgeleiden ongezondere keuzes maken in vergelijking met hoger opgeleiden, zelfs als ze exact dezelfde hoeveelheid informatie ontvangen. Dit suggereert dat lager opgeleiden minder belang hechten aan een goede gezondheid.

Hoofdstuk vijf onderzoekt het potentieel van informatievoorziening in het aanzetten tot gezond gedrag onder chronisch zieken. Patiënten krijgen na de diagnose van een chronische ziekte een aanzienlijke hoeveelheid informatie. Aan de ene kant krijgen ze het advies van hun arts om hun levensstijl te verbeteren, wat de relatie tussen gezondheidsuitkomsten en gedrag benadrukt. Aan de andere kant informeert een diagnose de patiënt over haar/zijn daadwerkelijke onderliggende gezondheid, wat een wake-up call kan zijn voor sommige patiënten. Een diagnose kan daarom gezien worden als een sterke "information treatment" op het individuele niveau. De resultaten geven aan dat, in overeenstemming met de aanbevelingen van artsen, mensen zich na een diagnose van een chronische ziekte onthouden van roken en drinken, zowel op de korte als lange termijn. Dit geldt zowel voor patiënten met hart- en vaatziekten als voor diabetici. Hierbij moet worden opgemerkt dat de daling van het aantal rokers sterker is onder patiënten met

hart- en vaatziekten. Echter, mensen verliezen nauwelijks gewicht en gaan nauwelijks meer bewegen.

Dit proefschrift bevat waardevolle inzichten voor beleidsdiscussies. Ten eerste, de bevinding dat het Nederlandse bevolkingsonderzoek naar borstkanker sterfte als gevolg van borstkanker vermindert is van groot belang in het kader van twijfels over de effectiviteit van dergelijke programma's. De bevinding is gebaseerd op een solide identificatie strategie en is dus een waardevolle bijdrage aan het huidige, zeer gefragmenteerde, debat over screening, waar niet al het gepresenteerde bewijs van gelijke kwaliteit is. Het ontbreken van een effect van het bevolkingsonderzoek naar borstkanker op totale sterfte laat enige twijfel bestaan over het effect van het bevolkingsonderzoek op de gezondheid van de gehele bevolking. Het in dit proefschrift gepresenteerde bewijs duidt erop dat de rol van gezondheidskennis in het creëren van sociaal-economische verschillen afhangt van het gedrag in kwestie. Hierdoor is het succes van interventies die gezondheidskennis willen bevorderen teneinde sociaal-economische ongelijkheden terug te dringen afhankelijk van het gedrag waarop de interventie is gericht.

References

Adamowicz, V. W. L., & Swait, J. D. (2011). Discrete Choice Theory and Modeling. In J. L. Lusk, J. Roosen, & J. F. Shogren (Eds.), *The Oxford Handbook of the Economics of Food Consumption and Policy.* Oxford: Oxford University Press.

Allcott, H. (2015). Site Selection Bias in Program Evaluation. *The Quarterly Journal of Economics*, *130*(3), 1117–1165.

American Cancer Society. (2011). *Global Cancer Facts & Figures. 2nd Edition* (Tech. Rep.). Atlanta, GA: American Cancer Society.

American Cancer Society. (2012a). *Breast Cancer Facts & Figures 2011-2012* (Tech. Rep.). Atlanta, GA: American Cancer Society.

American Cancer Society. (2012b). *Breast Cancer Prevention and Early Detection* (Tech. Rep.). Atlanta, GA: American Cancer Society.

American Cancer Society. (2014). *Cancer Facts & Figures 2014* (Tech. Rep.). Atlanta, GA: American Cancer Society.

American Cancer Society. (2015). *Cancer Facts & Figures 2015* (Tech. Rep.). Atlanta, GA: American Cancer Society.

American Heart Association. (2014a). *Home Blood Pressure Monitoring.* http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/SymptomsDiagnosisMonitoringofHighBloodPressure/Home-Blood-Pressure-Monitoring_UCM_301874_Article.jsp.

American Heart Association. (2014b). *Prevention & Treatment of High Blood Pressure.* http://www.heart.org/HEARTORG/Conditions/HighBloodPressure/PreventionTreatmentofHighBloodPressure/Prevention-Treatment-of-High-Blood-Pressure_UCM_002054_Article.jsp.

Andreski, P., McGonagle, K., & Schoeni, R. (2009). *An Analysis of the Quality of the Health Data in the Panel Study of Income Dynamics* (PSID Technical Series Paper No. 09-02). Ann Arbor, MI: University of Michigan.

Autor, D. H., & Duggan, M. G. (2006). The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding. *The Journal of Economic Perspectives*, *20*, 71–96.

Avitabile, C., Jappelli, T., & Padula, M. (2011). Cognitive Abilities, Healthcare and Screening Tests. *Journal of Population Ageing*, *4*(4), 251–269.

Barsky, R. B., Kimball, M. S., Juster, T. F., & Shapiro, M. D. (1997). Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Survey. *The Quarterly Journal of Economics*, *112*(2), 537–579.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies*, *81*(2), 608–650.

Bergman, L., Beelen, M. L., Gallee, M. P., Hollema, H., Benraadt, J., van Leeuwen, F. E., & the Comprehensive Cancer Centers' ALERT Group. (2000). Risk and Prognosis of Endometrial Cancer after Tamoxifen for Breast Cancer. *The Lancet*, *356*(9233), 881–887.

Bernheim, B. D., & Rangel, A. (2004). Addiction and Cue-Triggered Decision Processes. *American Economic Review*, *94*(5), 1558–1590.

Berrington de Gonzalez, A., Hartge, P., Cerhan, J. R., Flint, A. J., Hannan, L., MacInnis, R. J., ... Thun, M. J. (2010). Body-Mass Index and Mortality Among 1.46 Million White Adults. *New England Journal of Medicine*, *363*(23), 2211–2219.

Bijwaard, G., & Van Kippersluis, H. (2015). *Efficiency of Health Investment: Education or Intelligence?* (Working paper No. TI-15-004/V). Amsterdam: Tinbergen Institute.

Black, W. C., Haggstrom, D. A., & Welch, G. H. (2002). All-Cause Mortality in Randomized Trials of Cancer Screening. *Journal of the National Cancer Institute*, *94*(3), 167–173.

Bleyer, A., & Welch, H. G. (2012). Effect of Three Decades of Screening Mammography on Breast Cancer Incidence. *New England Journal of Medicine*, *367*(21), 1998–2005.

Bliemer, M. C. J., & Rose, J. M. (2010). Construction of Experimental Designs for Mixed Logit Models Allowing for Correlation across Choice Observations. *Transportation Research Part B: Methodological*, *44*(6), 720–734.

Bollinger, B., Leslie, P., & Sorensen, A. (2011). Calorie Posting in Chain Restaurants. *American Economic Journal: Economic Policy*, *3*(1), 91–128.

Borstkanker Vereniging Nederland. (2005). *Rapport Onderbehandeling van Borstkanker-Patiënten met Uitgezaaide HER2-Positieve Tumoren* (Tech. Rep.). Utrecht: Borstkanker Vereniging Nederland.

Boulé, N. G., Haddad, E., Kenny, G. P., Wells, G. A., & Sigal, R. J. (2001). Effects of Exercise on Glycemic Control and Body-Mass in Type-2 Diabetes Mellitus: A Meta-Analysis of Controlled Clinical Trials. *Journal of the American Medical Association*, *286*(10), 1218–1227.

Carman, K. G., & Kooreman, P. (2014). Probability Perceptions and Preventive Health Care. *Journal of Risk and Uncertainty*, *49*(1), 43–71.

Carson, A. P., Howard, G., Burke, G. L., Shea, S., Levitan, E. B., & Muntner, P. (2011). Ethnic Differences in Hypertension Incidence Among Middle-Aged and Older Adults The Multi-Ethnic Study of Atherosclerosis. *Hypertension*, *57*(6), 1101–1107.

Cawley, J. (2004). An Economic Framework for Understanding Physical Activity and Eating Behaviors. *American Journal of Preventive Medicine*, *27*(3), 117–125.

Cawley, J., & Meyerhoefer, C. (2012). The Medical Care Costs of Obesity: An Instrumental Variables Approach. *Journal of Health Economics*, *31*(1), 219–230.

Cawley, J., & Ruhm, C. J. (2011). The Economics of Risky Health Behaviors. In M. V. Pauly, T. G. McGuire, & P. P. Barros (Eds.), *Handbook of Health Economics* (Vol. 2). Amsterdam: North-Holland.

CentER Data. (2011). *LISS Panel.* http://www.lissdata.nl/lissdata/.

Centers for Disease Control and Prevention. (2009). *The Power of Prevention: Chronic Disease, the Public Health Challenge of the 21st Century* (Tech. Rep.). Atlanta, GA: U.S. Department of Health and Human Services.

Centers for Disease Control and Prevention. (2014). *National Diabetes Statistics Report: Estimates of Diabetes and its Burden in the United States, 2014* (Tech. Rep.). Atlanta, GA: U.S. Department of Health and Human Services.

Chaloupka, F. J., & Warner, K. E. (2000). The Economics of Smoking. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of Health Economics* (Vol. 1). Amsterdam: North-Holland.

Chapman, J.-A. W., Meng, D., Shepherd, L., Parulekar, W., Ingle, J. N., Muss, H. B., ... Goss, P. E. (2008). Competing Causes of Death from a Randomized Trial of Extended Adjuvant Endocrine Therapy for Breast Cancer. *Journal of the National Cancer Institute*, *100*(4), 252–260.

Chobanian, A. V., Bakris, G. L., Black, H. R., Cushman, W. C., Green, L. A., Izzo Jr, J. L., ... the National High Blood Pressure Education Program Coordinating Committee (2003). The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: The JNC 7 Report. *Journal of the American Medical Association*, *289*(19), 2560–2571.

Chowdhury, R., Warnakula, S., Kunutsor, S., Crowe, F., Ward, H. A., Johnson, L., ... Di Angelantonio, E. (2014). Association of Dietary, Circulating, and Supplement Fatty Acids with Coronary Risk: A Systematic Review and Meta-analysis. *Annals of Internal Medicine*, *160*(6), 398-406.

Chung, Y. (2013). *Chronic Health Conditions and Economic Outcomes* (Mimeo).

Clark, A., & Etile, F. (2002). Do Health Changes Affect Smoking? Evidence from British Panel Data. *Journal of Health Economics*, *21*(4), 533–562.

Cohen, J. T., Neumann, P. J., & Weinstein, M. C. (2008). Does Preventive Care Save Money? Health Economics and the Presidential Candidates. *New England Journal of Medicine*, *358*(7), 661–663.

Collette, H. J. A., Rombach, J. J., Day, N. E., & De Waard, F. (1984). Evaluation of Screening for Breast Cancer in a Non-Randomized Study (the DOM project) by Means of a Case-Control Study. *The Lancet*, *323*(8388), 1224–1226.

Commission of the European Countries. (1990). *Europe against Cancer Programme: Report on the Implementation of the First Plan of Action, 1987-1989. Communication from the European Commission to the Council, the European Parliament and the Economic and Social Committee* (Tech. Rep.). Brussells: EU Commission.

Conti, G., & Hansman, C. (2013). Personality and the Education-Health Gradient: A Note on "Understanding Differences in Health Behaviors by Education". *Journal of Health Economics*, *32*(2), 480–485.

Conti, G., Heckman, J., & Urzua, S. (2010). The Education-Health Gradient. *American Economic Review*, *100*(2), 234-38.

Cornfield, J. (1951). A Method of Estimating Comparative Rates from Clinical Data. Applications to Cancer of the Lung, Breast, and Cervix. *Journal of the National Cancer Institute*, *11*, 1269–1275.

Critchley, J. A., & Capewell, S. (2003). Smoking Cessation for the Secondary Prevention of Coronary Heart Disease. *Cochrane Database of Systematic Reviews*(CD003041).

Cutler, D. M. (2008). Are We Finally Winning the War on Cancer? *The Journal of Economic Perspectives*, *22*(4), 3–26.

Cutler, D. M., Glaeser, E. L., & Shapiro, J. M. (2003). Why have Americans Become more Obese? *The Journal of Economic Perspectives*, *17*(3), 93-118.

Cutler, D. M., & Lleras-Muney, A. (2010). Understanding Differences in Health Behaviors by Education. *Journal of Health Economics*, *29*(1), 1–28.

Cutler, D. M., Lleras-Muney, A., & Vogl, T. (2011). Socioeconomic Status and Health: Dimensions and Mechanisms. In S. Glied & P. C. Smith (Eds.), *Oxford Handbook of Health Economics.* Oxford: Oxford University Press.

De Irala-Estevez, J., Groth, M., Johansson, L., Oltersdorf, U., Prattala, R., & Martínez-González, M. A. (2000). A Systematic Review of Socioeconomic Differences in Food Habits in Europe: Consumption of Fruit and Vegetables. *European Journal of Clinical Nutrition*, *54*(9), 706–714.

DeVol, R., Bedroussian, A., Charuworn, A., Chatterjee, A., Kim, I. K., Kim, S., & Klowden, K. (2007). *An Unhealthy America: The Economic Burden of Chronic Disease. Charting a New Course to Save Lives and Increase Productivity and Economic Growth* (Project report). Santa Monica, CA: Milken Institute.

Dodd, G. D. (1992). American Cancer Society Guidelines on Screening for Breast Cancer: An Overview. *CA: A Cancer Journal for Clinicians*, *42*(3), 177–180.

Downs, J. S., Loewenstein, G., & Wisdom, J. (2009). Strategies for Promoting Healthier Food Choices. *The American Economic Review, Papers and Proceedings*, *99*(2), 159–164.

Doyle, C., Kushi, L. H., Byers, T., Courneya, K. S., Demark-Wahnefried, W., Grant, B., ... Andrews, K. S. (2006). Nutrition and Physical Activity During and After Cancer Treatment: An American Cancer Society Guide for Informed Choices. *CA: A Cancer Journal for Clinicians*, *56*(6), 323–353.

Drewnowski, A. (1997). Taste Preferences and Food Intake. *Annual Review of Nutrition*, *17*(1), 237-253.

Drewnowski, A., & Specter, S. E. (2004). Poverty and Obesity: The Role of Energy Density and Energy Costs. *The American Journal of Clinical Nutrition*, *79*(1), 6–16.

Early Breast Cancer Trialists' Collaborative Group. (2000). Favourable and Unfavourable Effects on Long-Term Survival of Radiotherapy for Early Breast Cancer: An Overview of the Randomized Trials. *The Lancet*, *355*(9217), 1757–1770.

Elbel, B., Kersh, R., Brescoll, V. L., & Dixon, L. B. (2009). Calorie Labeling and Food Choices: A First Look at the Effects on Low-Income People in New York City. *Health Affairs*, *28*(6), w1110–w1121.

Etilé, F. (2011). Food Consumption and Health. In J. L. Lusk, J. Roosen, & J. F. Shogren (Eds.), *The Oxford Handbook of the Economics of Food Consumption and Policy.* Oxford: Oxford University Press.

European Union. (2006). *Regulation (EC) No. 1924/2006 of the European Parliament and of the Council on Nutrition and Health Claims Made on Foods.*

Evert, A. B., Boucher, J. L., Cypress, M., Dunbar, S. A., Franz, M. J., Mayer-Davis, E. J., ... others (2013). Nutrition Therapy Recommendations for the Management of Adults With Diabetes. *Diabetes Care*, *36*(11), 3821–3842.

Falba, T. (2005). Health Events and the Smoking Cessation of Middle Aged Americans. *Journal of Behavioral Medicine*, *28*(1), 21–33.

Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J. W. W., Comber, H., ... Bray, F. (2013). Cancer Incidence and Mortality Patterns in Europe: Estimates for 40 Countries in 2012. *European Journal of Cancer*, *49*(6), 1374–1403.

Fiebig, D. G., Keane, M. P., Louviere, J., & Wasi, N. (2010). The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. *Marketing Science*, *29*(3), 393-421.

Fielder, H. M., Brook, D., Cuzick, J., Warwick, J., & Duffy, S. W. (2004). A Case-Control Study to Estimate the Impact on Breast Cancer Death of the Breast Screening Programme in Wales. *Breast Cancer Res*, *6*(Suppl 1), 1–2.

Finkelstein, E. A., Strombotne, K. L., Chan, N. L., & Krieger, J. (2011). Mandatory Menu Labeling in One Fast-Food Chain in King County, Washington. *American Journal of Preventive Medicine*, *40*(2), 122–127.

Fletcher, J. M., & Frisvold, D. E. (2009). Higher Education and Health Investments: Does more Schooling Affect Preventive Health Care Use? *Journal of human capital*, *3*(2), 144–176.

Frank, E., & Kunovichfrieze, T. (1995). Physicians? Prevention Counseling Behaviors: Current Status and Future Directions. *Preventive medicine*, *24*(6), 543–545.

Franz, M. J., Bantle, J. P., Beebe, C. A., Brunzell, J. D., Chiasson, J.-L., Garg, A., ... Wheeler, M. (2002). Evidence-Based Nutrition Principles and Recommendations for the Treatment and Prevention of Diabetes and Related Complications. *Diabetes Care*, *25*(1), 148–198.

Freedland, S. J., Aronson, W. J., Kane, C. J., Presti, J. C., Amling, C. L., Elashoff, D., & Terris, M. K. (2004). Impact of Obesity on Biochemical Control after Radical Prostatectomy for Clinically Localized Prostate Cancer: A Report by the Shared Equal Access Regional Cancer Hospital Database Study Group. *Journal of Clinical Oncology*, *22*(3), 446–453.

Fudenberg, D., & Levine, D. K. (2006). A Dual-Self Model of Impulse Control. *The American Economic Review*, *96*(5), 1449–1476.

Galama, T., & Van Kippersluis, H. (2010). *A Theory of Socioeconomic Disparities in Health over the Life Cycle* (Working paper No. 773). Santa Monica, CA: RAND corporation.

Gérvas, J., Starfield, B., & Heath, I. (2008). Is Clinical Prevention Better than Cure? *The Lancet*, *372*(9654), 1997–1999.

Giles, J. (2011). Social Science Lines Up its Biggest Challenges. *Nature*, *470*(7332), 18–19.

Goldman, D., & Smith, J. P. (2002). Can Patient Self-Management Help Explain the SES Health Gradient? *Proceedings of the National Academy of Sciences*, *99*, 10929–10934.

Gøtzsche, P. C., & Jørgensen, K. J. (2013). Screening for Breast Cancer with Mammography. *Cochrane Database of Systematic Reviews*(CD001877).

Gøtzsche, P. C., & Nielsen, M. (2011). Screening for Breast Cancer with Mammography. *Cochrane Database of Systematic Reviews*(CD001877).

Gøtzsche, P. C., & Olsen, O. (2000). Is Screening for Breast Cancer with Mammography Justifiable? *The Lancet*, *355*(9198), 129–134.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *The Journal of Political Economy*, 223–255.

Hanson, K., McPake, B., Nakamba, P., & Archard, L. (2005). Preferences for Hospital Quality in Zambia: Results from a Discrete Choice Experiment. *Health Economics*, *14*(7), 687–701.

Harris, E. E., Correa, C., Hwang, W.-T., Liao, J., Litt, H. I., Ferrari, V. A., & Solin, L. J. (2006). Late Cardiac Mortality and Morbidity in Early-Stage Breast Cancer Patients after Breast-Conservation Treatment. *Journal of Clinical Oncology*, *24*(25), 4100–4106.

Harteloh, P., de Bruin, K., & Kardaun, J. (2010). The Reliability of Cause-of-Death Coding in the Netherlands. *European Journal of Epidemiology*, *25*(8), 531–538.

Hausman, J. (2012). Contingent Valuation: From Dubious to Hopeless. *The Journal of Economic Perspectives*, *26*(4), 43–56.

Hayashi, K., Fukushima, K., Sagara, Y., & Takeshita, M. (1999). Surgical Treatment for Patients with Lung Cancer Complicated by Severe Pulmonary Emphysema. *The Japanese Journal of Thoracic and Cardiovascular Surgery*, *47*(12), 583–587.

Haydon, A. M., MacInnis, R. J., English, D. R., & Giles, G. G. (2006). Effect of Physical Activity and Body Size on Survival after Diagnosis with Colorectal Cancer. *Gut*, *55*(1), 62–67.

Heck, K. E., & Pamuk, E. R. (1997). Explaining the Relation between Education and Postmenopausal Breast Cancer. *American Journal of Epidemiology*, *145*(4), 366–372.

Hensher, D. A. (2006). How do Respondents Process Stated Choice Experiments? Attribute Consideration under Varying Information Load. *Journal of Applied Econometrics*, *21*(6), 861–878.

Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). *Applied Choice Analysis: A Primer*. Cambridge: Cambridge University Press.

Heron, M. (2015). *Deaths: Leading Causes for 2011* (Vol. 64; National Vital Statistics Reports No. 7). Hyattsville, MD: National Center for Health Statistics.

Hershman, D. L., Buono, D., McBride, R. B., Tsai, W. Y., Joseph, K. A., Grann, V. R., & Jacobson, J. S. (2008). Surgeon Characteristics and Receipt of Adjuvant Radiotherapy in Women with Breast Cancer. *Journal of the National Cancer Institute*, *100*(3), 199–206.

Hoch, S. J., & Loewenstein, G. F. (1991). Time-Inconsistent Preferences and Consumer Self-Control. *Journal of Consumer Research*, *17*(4), 492–507.

Holland, R., Rijken, H., & Hendriks, J. H. C. L. (2007). The Dutch Population-Based Mammography Screening: 30-Year Experience. *Breast Care*, *2*(1), 12–18.

Holmes, M. D., Chen, W. Y., Feskanich, D., Kroenke, C. H., & Colditz, G. A. (2005). Physical Activity and Survival after Breast Cancer Diagnosis. *Journal of the American Medical Association*, *293*(20), 2479–2486.

Honoré, B. E., & Lleras-Muney, A. (2006). Bounds in Competing Risks Models and the War on Cancer. *Econometrica*, *74*(6), 1675–1698.

Huber, J., & Zwerina, K. (1996). The Importance of Utility Balance in Efficient Choice Designs. *Journal of Marketing Research*, *33*(3), 307–317.

Independent UK Panel on Breast Cancer Screening. (2012). The Benefits and Harms of Breast Cancer Screening: An Independent Review. *The Lancet*, *380*(9855), 1778–1786.

International Agency for Research on Cancer. (2002). *IARC Handbooks of Cancer Prevention. Breast Cancer Screening* (Vol. 7; H. Vainio & F. Bianchini, Eds.). Lyon: IARC Press.

Ippolito, P. M. (1981). Information and the Life Cycle Consumption of Hazardous Goods. *Economic Inquiry*, *19*(4), 529–558.

Jatoi, I., & Miller, A. B. (2003). Why is Breast Cancer Mortality Declining? *The Lancet Oncology*, *4*(4), 251–254.

Johansson-Stenman, O. (2011). Health Investments under Risk and Ambiguity. In J. L. Lusk, J. Roosen, & J. F. Shogren (Eds.), *The Oxford Handbook of the Economics of Food Consumption and Policy.* Oxford: Oxford University Press.

Johnston-Early, A., Cohen, M. H., Minna, J. D., Paxton, L. M., Fossieck, B. E., Ihde, D. C., ... Makuch, R. (1980). Smoking Abstinence and Small Cell Lung Cancer Survival: An Association. *Journal of the American Medical Association*, *244*(19), 2175–2179.

Jusot, F., Or, Z., & Sirven, N. (2012). Variations in Preventive Care Utilization in Europe. *European Journal of Ageing*, *9*(1), 15–25.

Kaestner, R., Darden, M., & Lakdawalla, D. (2014). Are Investments in Disease Prevention Complements? The Case of Statins and Health Behaviors. *Journal of Health Economics*, *36*, 151–163.

Kahn, M. E. (1998). Education's Role in Explaining Diabetic Health Investment Differentials. *Economics of Education Review*, *17*(3), 257–266.

Kahn, M. E. (1999). Diabetic Risk Taking: The Role of Information, Education and Medication. *Journal of Risk and Uncertainty*, *18*(2), 147–164.

Kahneman, D. (2011). *Thinking, Fast and Slow.* New York, NY: Farrar, Straus and Giroux.

Kalager, M., Zelen, M., Langmark, F., & Adami, H.-O. (2010). Effect of Screening Mammography on Breast Cancer Mortality in Norway. *New England Journal of Medicine*, *363*(13), 1203–1210.

Kamphuis, C. B. M., de Bekker-Grob, E. W., & van Lenthe, F. J. (2015). Factors Affecting Food Choices of Older Adults from High and Low Socioeconomic Groups: A Discrete Choice Experiment. *The American Journal of Clinical Nutrition*, *101*(4), 768–774.

Katzung, B. G., Masters, S. B., & Trevor, A. J. (2012). *Basic & Clinical Pharmacology* (12th ed.). New York, NY: McGraw-Hill Professional.

Kawahara, M., Ushijima, S., Kamimori, T., Kodama, N., Ogawara, M., Matsui, K., ... Furuse, K. (1998). Second Primary Tumours in more than 2-year Disease-Free Survivors of Small-Cell Lung Cancer in Japan: The Role of Smoking Cessation. *British Journal of Cancer*, *78*(3), 409–412.

Kenkel, D. S. (1991). Health Behavior, Health Knowledge, and Schooling. *Journal of Political Economy*, *99*(2), 287–305.

Khwaja, A., Sloan, F., & Chung, S. (2006). Learning about Individual Risk and the Decision to Smoke. *International Journal of Industrial Organization*, *24*(4), 683–699.

Khwaja, A., Sloan, F., & Chung, S. (2007). The Relationship between Individual Expectations and Behaviors: Mortality Expectations and Smoking Decisions. *Journal of Risk and Uncertainty*, *35*(2), 179–201.

Kjær, T., Bech, M., Gyrd-Hansen, D., & Hart-Hansen, K. (2006). Ordering Effect and Price Sensitivity in Discrete Choice Experiments: Need We Worry? *Health economics*, *15*(11), 1217–1228.

Krantz, D. H. (1991). From Indices to Mappings: The Representational Approach to Measurement. In D. Brown & J. Smith (Eds.), *Frontiers of Mathematical Psychology* (pp. 1–52). New York, NY: Springer-Verlag.

Lairson, D. R., Chan, W., & Newmark, G. R. (2005). Determinants of the Demand for Breast Cancer Screening among Women Veterans in the United States. *Social Science & Medicine*, *61*(7), 1608–1617.

Lakdawalla, D., & Philipson, T. (2009). The Growth of Obesity and Technological Change. *Economics & Human Biology*, *7*(3), 283–293.

Lakins, N. E., LaVallee, R. A., Williams, G. D., & Yi, H. (2008). *Apparent Per Capita Alcohol Consumption: National, State, and Regional Trends, 1977–2006* (Surveillance report No. 85). Arlington, VA: National Institute on Alcohol Abuse and Alcoholism.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, *74*(2), 132–157.

Lancsar, E., & Louviere, J. (2008). Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. *Pharmacoeconomics*, *26*(8), 661–677.

Lange, F. (2011). The Role of Education in Complex Health Decisions: Evidence from Cancer Screening. *Journal of Health Economics*, *30*(1), 43–54.

Lauby-Secretan, B., Scoccianti, C., Loomis, D., Benbrahim-Tallaa, L., Bouvard, V., Bianchini, F., & Straif, K. (2015). Breast Cancer Screening - Viewpoint of the IARC Working Group. *New England Journal of Medicine*, *372*(24), 2353–2358.

Leganger, A., & Kraft, P. (2003). Control Constructs: Do They Mediate the Relation between Educational Attainment and Health Behaviour? *Journal of Health Psychology*, *8*(3), 361–372.

Leslie, W. S., Hankey, C. R., & Lean, M. E. J. (2007). Weight Gain as an Adverse Effect of some Commonly Prescribed Drugs: A Systematic Review. *QJM: An International Journal of Medicine*, *100*(7), 395–404.

Loomis, J. (2011). What is to Know about Hypothetical Bias in Stated Preference Valuation Studies? *Journal of Economic Surveys*, *25*(2), 363–370.

Louviere, J. J., Flynn, T. N., & Carson, R. T. (2010). Discrete Choice Experiments are not Conjoint Analysis. *Journal of Choice Modelling*, *3*(3), 57–72.

Mackenbach, J. P. (2006). *Health Inequalities: Europe in Profile* (Tech. Rep.). London: U.K. Department of Health.

Mäkimattila, S., Nikkilä, K., & Yki-Järvinen, H. (1999). Causes of Weight Gain during Insulin Therapy with and without Metformin in Patients with Type-2 Diabetes Mellitus. *Diabetologia*, *42*(4), 406–412.

Malhotra, A. (2013). Saturated Fat is not the Major Issue. *BMJ*, *347*.

Mancia, G., Fagard, R., Narkiewicz, K., Redon, J., Zanchetti, A., Böhm, M., ... others (2013). 2013 ESH/ESC Guidelines for the Management of Arterial Hypertension: The Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *Blood Pressure*, *22*(4), 193–278.

Mangham, L. J., Hanson, K., & McPake, B. (2009). How to Do (or not to Do) – Designing a Discrete Choice Experiment for Application in a Low-Income Country. *Health Policy and Planning*, *24*(2), 151–158.

Manski, C. F. (2009). *Identification for Prediction and Decision*. Cambridge, MA: Harvard University Press.

Mathios, A. D. (2000). The Impact of Mandatory Disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market. *The Journal of Law and Economics*, *43*(2), 651–678.

Meara, E. (2001). *Why is Health Related to Socioeconomic Status?* (Working paper No. 8231). Cambridge, MA: National Bureau of Economic Research.

Meissner, H. I., Breen, N., Taubman, M. L., Vernon, S. W., & Graubard, B. I. (2007). Which Women are not Getting Mammograms and Why? (United States). *Cancer Causes & Control*, *18*(1), 61–70.

Menvielle, G., Kunst, A. E., Van Gils, C. H., Peeters, P. H., Boshuizen, H., Overvad, K., ... others (2011). The Contribution of Risk Factors to the Higher Incidence of Invasive and In Situ Breast Cancers in Women with Higher Levels of Education in the European Prospective Investigation into Cancer and Nutrition. *American Journal of Epidemiology*, *173*(1), 26–37.

Michael, R. T., & Becker, G. S. (1973). On the New Theory of Consumer Behavior. *The Swedish Journal of Economics*, 378–396.

Miller, A. B., Wall, C., Baines, C. J., Sun, P., To, T., & Narod, S. A. (2014). Twenty Five Year Follow-Up for Breast Cancer Incidence and Mortality of the Canadian National Breast Screening Study: Randomized Screening Trial. *BMJ*, *348*.

Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., ... others (2015). Heart Disease and Stroke Statistics-2015 Update: A Report from the American Heart Association. *Circulation*, *131*(4), e29–e322.

Mullahy, J. (1999). It'll Only Hurt a Second? Microeconomic Determinants of Who Gets Flu Shots. *Health economics*, *8*(1), 9–24.

National Cancer Institute. (2011). *Cancer Prevalence and Cost of Care Projections.* http://costprojections.cancer.gov/.

National Evaluation Team for Breast Cancer Screening. (2009). *National Evaluation of Breast Cancer Screening in the Netherlands: 1990-2007* (12th evaluation report). Rotterdam: Department of Public Health, Erasmus MC.

National Evaluation Team for Breast Cancer Screening. (2014). *National Evaluation of Breast Cancer Screening in the Netherlands 1990-2011/2012* (Tech. Rep.). Rotterdam: Department of Public Health, Erasmus MC.

Nelson, H. D., Tyne, K., Naik, A., Bougatsos, C., Chan, B. K., & Humphrey, L. (2009). Screening for Breast Cancer: An Update for the U.S. Preventive Services Task Force. *Annals of Internal Medicine*, *151*(10), 727–737.

Neter, J. E., Stam, B. E., Kok, F. J., Grobbee, D. E., & Geleijnse, J. M. (2003). Influence of Weight Reduction on Blood Pressure: A Meta-Analysis of Randomized Controlled Trials. *Hypertension*, *42*(5), 878–884.

NIBUD. (2014). *Wat Geeft U uit aan Voeding?* http://www.nibud.nl/uitgaven/huishouden/voeding.html.

Niezen, M. G., Stolk, E. A., Steenhoek, A., & Uyl-De Groot, C. A. (2006). Inequalities in Oncology Care: Economic Consequences of High Cost Drugs. *European Journal of Cancer*, *42*(17), 2887–2892.

Nwankwo, T., Yoon, S. S., Burt, V., & Gu, Q. (2013). *Hypertension among Adults in the United States: National Health and Nutrition Examination Survey, 2011-2012* (Data brief). Hyattsville, MD: National Center for Health Statistics.

Nyström, L., Andersson, I., Bjurstam, N., Frisell, J., Nordenskjöld, B., & Rutqvist, L. E. (2002). Long-Term Effects of Mammography Screening: Updated Overview of the Swedish Randomized Trials. *The Lancet*, *359*(9310), 909–919.

OECD. (2014a). *Health at a Glance: Europe 2014*. Paris: OECD Publishing.

OECD. (2014b). *Obesity Update 2014* (Tech. Rep.). Paris: OECD.

O'Malley, M. S., Earp, J. A., Hawley, S. T., Schell, M. J., Mathews, H. F., & Mitchell, J. (2001). The Association of Race/Ethnicity, Socioeconomic Status, and Physician

Recommendation for Mammography: Who Gets the Message about Breast Cancer Screening? *American Journal of Public Health*, *91*(1), 49–54.

Opie, L. H., & Lecour, S. (2007). The Red Wine Hypothesis: From Concepts to Protective Signalling Molecules. *European Heart Journal*, *28*(14), 1683–1693.

Oreopoulos, P., & Salvanes, K. G. (2011). Priceless: The Nonpecuniary Benefits of Schooling. *The Journal of Economic Perspectives*, 159–184.

Oster, E. (2015). *Diabetes and Diet: Behavioral Response and the Value of Health* (Mimeo).

Otten, J. D., Broeders, M. J., Fracheboud, J., Otto, S. J., de Koning, H. J., & Verbeek, A. L. M. (2008). Impressive Time-Related Influence of the Dutch Screening Programme on Breast Cancer Incidence and Mortality, 1975-2006. *International Journal of Cancer*, *123*(8), 1929–1934.

Otto, S. J., Fracheboud, J., Looman, C. W. N., Broeders, M. J. M., Boer, R., Hendriks, J. H. C. L., ... de Koning, H. J. (2003). Initiation of Population-Based Mammography Screening in Dutch Municipalities and Effect on Breast Cancer Mortality: A Systematic Review. *The Lancet*, *361*(9367), 1411–1417.

Otto, S. J., Fracheboud, J., Verbeek, A. L. M., Boer, R., Reijerink-Verheij, J. C. I. Y., Otten, J. D. M., ... de Koning, H. J. (2012). Mammography Screening and Risk of Breast Cancer Death: A Population-Based Case-Control Study. *Cancer Epidemiology Biomarkers & Prevention*, *21*(1), 66–73.

Palència, L., Espelt, A., Rodríguez-Sanz, M., Puigpinós, R., Pons-Vigués, M., Pasarín, M. I., ... Borrell, C. (2010). Socio-Economic Inequalities in Breast and Cervical Cancer Screening Practices in Europe: Influence of the Type of Screening Program. *International Journal of Epidemiology*, *39*, 757–765.

Penston, J. (2011). Should We Use Total Mortality rather than Cancer Specific Mortality to Judge Cancer Screening Programmes? Yes. *BMJ*, *343*.

Picone, G., Sloan, F., & Taylor Jr, D. (2004). Effects of Risk and Time Preference and Expected Longevity on Demand for Medical Tests. *Journal of Risk and Uncertainty*, *28*(1), 39–53.

Prospective Studies Collaboration. (2009). Body-Mass Index and Cause-Specific Mortality in 900,000 Adults: Collaborative Analyses of 57 Prospective Studies. *The Lancet*, *373*(9669), 1083–1096.

RAND Corporation. (2011). *American Life Panel.* https://alpdata.rand.org/.

Revelt, D., & Train, K. (2000). *Customer-Specific Taste Parameters and Mixed Logit: Households' Choice of Electricity Supplier* (Working paper No. E00-274). Berkeley, CA: University of California at Berkeley.

Rijksinstituut voor Volksgezondheid en Milieu. (2015). *Bevolkingsonderzoek Borstkanker voor Professionals.* http://www.rivm.nl/Onderwerpen/B/Bevolkingsonderzoek_borstkanker_voor_professionals/Achtergrond_organisatie_kosten_en_geschiedenis/Geschiedenis.

Roberto, C. A., Larsen, P. D., Agnew, H., Baik, J., & Brownell, K. D. (2010). Evaluating the Impact of Menu Labeling on Food Choices and Intake. *American Journal of Public Health*, *100*(2), 312.

Rock, C. L., & Demark-Wahnefried, W. (2002). Nutrition and Survival after the Diagnosis of Breast Cancer: A Review of the Evidence. *Journal of Clinical Oncology*, *20*(15), 3302–3316.

Rose, J. M., & Bliemer, M. C. J. (2013). Sample Size Requirements for Stated Choice Experiments. *Transportation*, *40*(5), 1021-1041.

Rossi, A., Dikareva, A., Bacon, S. L., & Daskalopoulou, S. S. (2012). The Impact of Physical Activity on Mortality in Patients with High Blood Pressure: A Systematic Review. *Journal of Hypertension*, *30*(7), 1277–1288.

Russell, L. B. (1986). *Is Prevention Better than Cure?* Washington, DC: The Brookings Institution.

Russell, L. B. (1993). The Role of Prevention in Health Reform. *New England Journal of Medicine*, *329*(5), 352–354.

Russell, L. B. (2009). Preventing Chronic Disease: An Important Investment, But do not Count on Cost Savings. *Health Affairs*, *28*(1), 42-45.

Saffer, H. (2014). *Self-Regulation and Health* (Working Paper No. 20483). Cambridge, MA: National Bureau of Economic Research.

Sambamoorthi, U., & McAlpine, D. D. (2003). Racial, Ethnic, Socioeconomic, and Access Disparities in the Use of Preventive Services among Women. *Preventive Medicine*, *37*(5), 475–484.

Savage, L., & Widener, A. (2008). Physician Characteristics are Associated with Quality of Cancer Care. *Journal of the National Cancer Institute*.

Schouten, L., Meijer, H., Huveneers, J., & Kiemeney, L. (1996). Urban-Rural Differences in Cancer Incidence in the Netherlands, 1989–1991. *International Journal of Epidemiology*, *25*(4), 729–736.

Selvin, E., & Brett, K. M. (2003). Breast and Cervical Cancer Screening: Sociodemographic Predictors among White, Black, and Hispanic Women. *American Journal of Public Health*, *93*(4), 618–623.

Shapiro, S., Venet, W., Strax, P., Venet, L., & Roeser, R. (1982). Ten-to Fourteen-Year Effect of Screening on Breast Cancer Mortality. *Journal of the National Cancer Institute*, *69*(2), 349–355.

Shiv, B., & Fedorikhin, A. (1999). Heart and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making. *Journal of Consumer Research*, *26*(3), 278–292.

Sigal, R. J., Kenny, G. P., Wasserman, D. H., & Castaneda-Sceppa, C. (2004). Physical Activity/Exercise and Type-2 Diabetes. *Diabetes Care*, *27*(10), 2518–2539.

Simon, H. A. (1956). Rational Choice and the Structure of the Environment. *Psychological Review*, *63*(2), 129–138.

Slade, A. N. (2012). Health Investment Decisions in Response to Diabetes Information in Older Americans. *Journal of Health Economics*, *31*(3), 502–520.

Slade, A. N., & Kim, H. (2014). Dietary Responses to a Hypertension Diagnosis: Evidence from the National Health and Nutrition Examination Survey (NHANES) 2007–2010. *Behavioral Medicine*, *40*(1), 1–13.

Slovic, P. (1995). The Construction of Preference. *American Psychologist*, *50*(5), 364–371.

Smith, J. P. (1999). Healthy Bodies and Thick Wallets. *The Journal of Economic Perspectives*, *13*, 145–166.

Smith, J. P. (2007). The Impact of Socioeconomic Status on Health over the Life-Course. *Journal of Human Resources*, *42*(4), 739–764.

Smith, K. V., Taylor Jr, D. H., Sloan, F. A., Johnson, F. R., & Desvousges, W. H. (2001). Do Smokers Respond to Health Shocks? *Review of Economics and Statistics*, *83*(4), 675–687.

Steele, R. J. C., & Brewster, D. H. (2011). Should We Use Total Mortality rather than Cancer Specific Mortality to Judge Cancer Screening Programmes? No. *BMJ*, *343*.

Steptoe, A., Pollard, T. M., & Wardle, J. (1995). Development of a Measure of the Motives Underlying the Selection of Food: The Food Choice Questionnaire. *Appetite*, *25*(3), 267–284.

Tabár, L., Vitak, B., Chen, T. H.-H., Yen, A. M.-F., Cohen, A., Tot, T., ... Duffy, S. W. (2011). Swedish Two-County Trial: Impact of Mammographic Screening on Breast Cancer Mortality During 3 Decades. *Radiology*, *260*(3), 658–663.

Thompson, F. E., & Subar, A. F. (2008). Dietary Assessment Methodology. In A. M. Coulston, C. J. Boushey, & M. G. Ferruzzi (Eds.), *Nutrition in the Prevention and Treatment of Disease* (3rd ed.). San Diego, CA: Academic Press.

Train, K. E. (2009). *Discrete Choice Methods with Simulation* (2nd ed.). Cambridge: Cambridge University Press.

Unesco Institute for Statistics. (2011). *ISCED: International Standard Classification of Education.* http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx.

United Nations General Assembly. (2011). *Political Declaration of the High-Level Meeting of the General Assembly on the Prevention and Control of Non-Communicable Diseases.*

Urban, N., Anderson, G. L., & Peacock, S. (1994). Mammography Screening: How Important is Cost as a Barrier to Use? *American Journal of Public Health*, *84*(1), 50–55.

U.S. Centers for Medicare and Medicaid Services. (2011). *Free Preventive Care.* https://www.healthcare.gov/health-care-law-protections/free-preventive-care/.

U.S. Department of Agriculture and U.S. Department of Health and Human Services. (2010). *Dietary Guidelines for Americans* (7th Edition ed.). Washington, DC: U.S. Government Printing Office.

U.S. Department of Health and Human Services. (1990). *The Health Benefits of Smoking Cessation* (A Report of the Surgeon General). Rockville, MD: U.S. Department of Health and Human Services.

U.S. Department of Health and Human Services and U.S. Department of Agriculture. (2005). *Dietary Guidelines for Americans 2005* (6th ed.). Washington, DC: U.S. Government Printing Office.

U.S. Food and Drug Administration. (2012). *Sodium in your Diet. Using the Nutrition Facts Label to Reduce your Intake.*

U.S. Preventive Services Task Force. (1989). *Guide to Clinical Preventive Services: An Assessment of the Effectiveness of 169 Interventions. Report of the U.S. Preventive Services Task Force.* Philadelphia: Williams & Wilkins.

U.S. Preventive Services Task Force. (2009). Screening For Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals of Internal Medicine*, *151*(10), 716–726.

Van Leeuwen, F. E., Van den Belt-Dusebout, A. W., Benraadt, J., Diepenhorst, F. W., Van Tinteren, H., Coebergh, J. W. W., ... Bontenbal, M. (1994). Risk of Endometrial Cancer After Tamoxifen Treatment of Breast Cancer. *The Lancet*, *343*(8895), 448–452.

Van Rossum, C. T. M., Fransen, H. P., Verkaik-Kloosterman, J., Buurma-Rethans, E. J. M., & Ocke, M. C. (2011). *Dutch National Food Consumption Survey 2007-2010 : Diet of children and adults aged 7 to 69 years* (Tech. Rep.). Bilthoven: Rijksinstituut voor Volksgezondheid en Milieu RIVM.

Van Schoor, G., Moss, S. M., Otten, J. D. M., Donders, R., Paap, E. D. E. N., den Heeten, G. J., ... Verbeek, A. L. M. (2011). Increasingly Strong Reduction in Breast Cancer Mortality Due to Screening. *British Journal of Cancer*, *104*(6), 910–914.

Variyam, J. N. (2008). Do Nutrition Labels Improve Dietary Outcomes? *Health Economics*, *17*(6), 695–708.

Walsh, B., Silles, M., & O'Neill, C. (2011). The Importance of Socio-Economic Variables in Cancer Screening Participation: A Comparison between Population-Based and Opportunistic Screening in the EU-15. *Health Policy*, *101*(3), 269–276.

Wansink, B., & Chandon, P. (2006). Can "Low-Fat" Nutrition Labels Lead to Obesity? *Journal of Marketing Research*, *43*(4), 605–617.

Welch, G. H., Schwartz, L., & Woloshin, S. (2011). *Overdiagnosed: Making People Sick in the Pursuit of Health.* Boston, MA: Beacon Press.

Willett, W. C., Sampson, L., Stamper, M. J., Rosner, B., Bain, C., Witschi, J., ... Speizer, F. E. (1985). Reproducibility and Validity of a Semiquantitative Food Frequency Questionnaire. *American Journal of Epidemiology*, *122*(1), 51-65.

Wisdom, J., Downs, J. S., & Loewenstein, G. (2010). Promoting Healthy Choices: Information versus Convenience. *American Economic Journal: Applied Economics*, *2*(2), 164–178.

World Health Organization. (2009). *Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks*. Geneva: WHO Press.

World Health Organization. (2011). *10 Facts on Cancer*. http://www.who.int/features/factfiles/cancer/facts/en/index.html.

World Health Organization. (2014). *Global Status Report on Noncommunicable Diseases* (Tech. Rep.). Geneva: World Health Organisation.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

596. S.L. BLAUW, *Well-to-Do or Doing Well? Empirical Studies of Well-Being and Development*

597. T.A. MAKAREWICZ, *Learning to Forecast: Genetic Algorithms and Experiments*

598. P. ROBALO, *Understanding Political Behavior: Essays in Experimental Political Economy*

599. R. ZOUTENBIER, *Work Motivation and Incentives in the Public Sector*

600. M.B.W. KOBUS, *Economic Studies on Public Facility Use*

601. R.J.D. POTTER VAN LOON, *Modeling Non-Standard Financial Decision Making*

602. G. MESTERS, *Essays on Nonlinear Panel Time Series Models*

603. S. GUBINS, *Information Technologies and Travel*

604. D. KOPÁNYI, *Bounded Rationality and Learning in Market Competition*

605. N. MARTYNOVA, *Incentives and Regulation in Banking*

606. D. KARSTANJE, *Unraveling Dimensions: Commodity Futures Curves and Equity Liquidity*

607. T.C.A.P. GOSENS, *The Value of Recreational Areas in Urban Regions*

608. Ł.M. MARĆ, *The Impact of Aid on Total Government Expenditures*

609. C. LI, *Hitchhiking on the Road of Decision Making under Uncertainty*

610. L. ROSENDAHL HUBER, *Entrepreneurship, Teams and Sustainability: A Series of Field Experiments*

611. X. YANG, *Essays on High Frequency Financial Econometrics*

612. A.H. VAN DER WEIJDE, *The Industrial Organization of Transport Markets: Modeling Pricing, Investment and Regulation in Rail and Road Networks*

613. H.E. SILVA MONTALVA *Airport Pricing Policies: Airline Conduct, Price Discrimination, Dynamic Congestion and Network Effects*

614. C. DIETZ, *Hierarchies, Communication and Restricted Cooperation in Cooperative Games*

615. M.A. ZOICAN, *Financial System Architecture and Intermediation Quality*

616. G. ZHU, *Three Essays in Empirical Corporate Finance*

617. M. PLEUS, *Implementations of Tests on the Exogeneity of Selected Variables and their Performance in Practice*

618. B. VAN LEEUWEN, *Cooperation, Networks and Emotions: Three Essays in Behavioral Economics*

619. A.G. KOPÁNYI-PEUKER, *Endogeneity Matters: Essays on Cooperation and Coordination*

620. X. WANG, *Time Varying Risk Premium and Limited Participation in Financial Markets*

621. L.A. GORNICKA, *Regulating Financial Markets: Costs and Trade-Offs*

622. A. KAMM, *Political Actors Playing Games: Theory and Experiments*

623. S. VAN DEN HAUWE, *Topics in Applied Macroeconometrics*

624. F.U. BRÄUNING, *Interbank Lending Relationships, Financial Crises and Monetary Policy*

625. J.J. DE VRIES, *Estimation of Alonso's Theory of Movements for Commuting*

626. M. POPLAWSKA, *Essays on Insurance and Health Economics*

627. X. CAI, *Essays in Labor and Product Market Search*

628. L. ZHAO, *Making Real Options Credible: Incomplete Markets, Dynamics, and Model Ambiguity*

629. K. BEL, *Multivariate Extensions to Discrete Choice Modeling*

630. Y. ZENG, *Topics in Trans-Boundary River Sharing Problems and Economic Theory*

631. M.G. WEBER, *Behavioral Economics and the Public Sector*

632. E. CZIBOR, *Heterogeneity in Response to Incentives: Evidence from Field Data*

633. A. JUODIS, *Essays in Panel Data Modelling*

634. F. ZHOU, *Essays on Mismeasurement and Misallocation on Transition Economies*

635. P. MULLER, *Labor Market Policies and Job Search*

636. N. KETEL, *Empirical Studies in Labor and Education Economics*

637. T.E. YENILMEZ, *Three Essays in International Trade and Development*

638. L.P. DE BRUIJN, *Essays on Forecasting and Latent Values*

639. S. VRIEND, *Profiling, Auditing and Public Policy: Applications in Labor and Health Economics*

640. M.L. ERGUN, *Fat Tails in Financial Markets*

641. T. HOMAR, *Intervention in Systemic Banking Crises*

642. R. LIT, *Time Varying Parameter Models for Discrete Valued Time Series*

643. R.H. KLEIJN, *Essays on Bayesian Model Averaging using Economic Time Series*

644. S. MUNS, *Essays on Systemic Risk*

645. B.M. SADABA, *Essays on the Empirics of International Financial Markets*