Network for Studies on Pensions, Aging and Retirement

Netspar THESES

Patrick Hullegie

Essays on Health and Labor Economics

PhD Thesis 2012-016

Essays on health and labor economics

Patricius Gerhardus Johannes Hullegie

Essays on health and labor economics

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University, op gezag van de rector magnificus, prof.dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op vrijdag 5 oktober 2012 om 10.15 uur door

Patricius Gerhardus Johannes Hullegie

geboren op 17 oktober 1982 te Deventer.

Promotor:	Prof. dr. P. Kooreman
Copromotor:	Dr. T. J. Klein
Promotiecommissie:	Prof. dr. J. H. Abbring
	Prof. dr. M. Lindeboom
	Prof. dr. J. Maurer
	Prof. dr. A. H. O. van Soest

Acknowledgements

This thesis contains the results of the research I carried out as a Ph.D. student at Tilburg University. During this period I have benefited from the advice, knowledge, and support of many people.

First of all, I would like to thank my advisors Peter Kooreman and Tobias Klein. During a talk in his office at the University of Groningen, Peter asked me whether I would be interested in pursuing a Ph.D. at Tilburg University, where he would be a professor as of the next academic year. My answer should be clear. Thanks to Peter I have been able to experience the joy of doing research and the great research environment at Tilburg University. I am also grateful to Peter for the advice and support he gave over the years. Tobias has taught me valuable lessons about empirical research, and his guidance and optimism have been important throughout my Ph.D.

Next, I would like to thank the other members of my committee, Jaap Abbring, Maarten Lindeboom, Jürgen Maurer, and Arthur van Soest for their extensive and excellent feedback on my thesis.

I am grateful to Arie Kapteyn for giving me the opportunity to visit the RAND Corporation for three months, and for generously letting me be a guest in his beautiful house in Topanga. At RAND I worked with Titus Galama and Erik Meijer, who are wonderful co-authors.

I am also grateful to Jan van Ours, for facilitating a longer stay at Tilburg University and for introducing me to the field of empirical labor economics. Working with Jan was very pleasant, and I hope to have the opportunity to work with him in the future as well.

I would like to thank Hanka Voňková for being a wonderful co-author and good friend. Room K431 has been a very pleasant office throughout my Ph.D. thanks to Nathanaël Vellekoop. I also would like to thank Martin Salm for all the useful advice he has given over the last couple of years.

My move from the beautiful city of Groningen *(Er gaat immers niets boven Groningen)* to the less beautiful city of Tilburg didn't seem particularly attractive in the beginning, but turned out to be really enjoyable thanks to Aleks, Chris, Christian, David, Gerard, Hanka, Jan, Kenan, Kim, Luc, Marta, Martin, Miguel, Nathanaël, Peter, Pedro, Raposo, Salima, Sander, and Tunga.

I would like to thank Frans & Marion, Gertjan & Karin, Henk & Simon, Herman, Jasper, Jeffrey & Maaike, Johan & Gerrie, Michiel & Marita, Sander & Hedwig, Willibrord, Wouter & Marijke, and Wouter & Naomie for giving a non-academic perspective on life and for sharing many good times.

Most importantly, I wouldn't be where I am now without the love and support of my family; papa & mama, Manon & Roy & Finn, Ellen & Jasper, heel erg bedankt!

Contents

Chapter 1

Introduction and summary

This thesis consists of three parts. The first part (chapter 2) examines the validity of a method that aims at improving the interpersonal comparability of self-reports in surveys. The second part (chapters 3 and 4) is concerned with the questions how the demand for medical care is related to health insurance, and to health, respectively. The third part (chapter 5) studies whether job search requirements help older workers to find a job more quickly. The chapters in this thesis are based on the following research papers:

Chapter 2:

Voňková, H., & Hullegie, P. (2011). Is the anchoring vignette method sensitive to the domain and choice of the vignette? *Journal of the Royal Statistical Society Series A*, Vol. 174, No. 3, pp. 597–620.

Chapter 3:

Hullegie, P., & Klein, T. J. (2010). The effect of private health insurance on medical care utilization and self-assessed health in Germany. *Health Economics*, Vol. 19, No. 9, pp. 1048–1062.

and

Hullegie, P., & Klein, T. J. (2011). The effect of private health insurance on doctor visits, hospital nights and self-assessed health: Evidence from the German Socio-Economic Panel. *Schmollers Jahrbuch (Journal of Applied Social Science Studies)*, Vol. 131, No. 2, pp. 395–407.

Chapter 4:

Galama, T. J., Hullegie, P., Meijer, E., & Outcault, S. (2012). Is there empirical evidence for decreasing returns to scale in a health capital model? *Health Economics*, Vol. 21, No. 9, pp. 1080–1100.

Chapter 5:

Hullegie, P. & Van Ours, J. C. (2012). Seek and ye shall find: how search requirements affect job finding rates of older workers. *Working Paper*.

1.1 Interpersonal comparability of self-reports in surveys

Researchers in social sciences working with surveys often ask for respondents' self-assessment of a concept of interest. The following question on political efficacy is an example (King, Murray, Salomon, & Tandon, 2004)

> How much say do you have in getting the government to address issues that interest you?
> – no say at all, little say, some say, a lot of say, unlimited say.

Another example is this question on work disability (Kapteyn, Smith, & Van Soest, 2007)

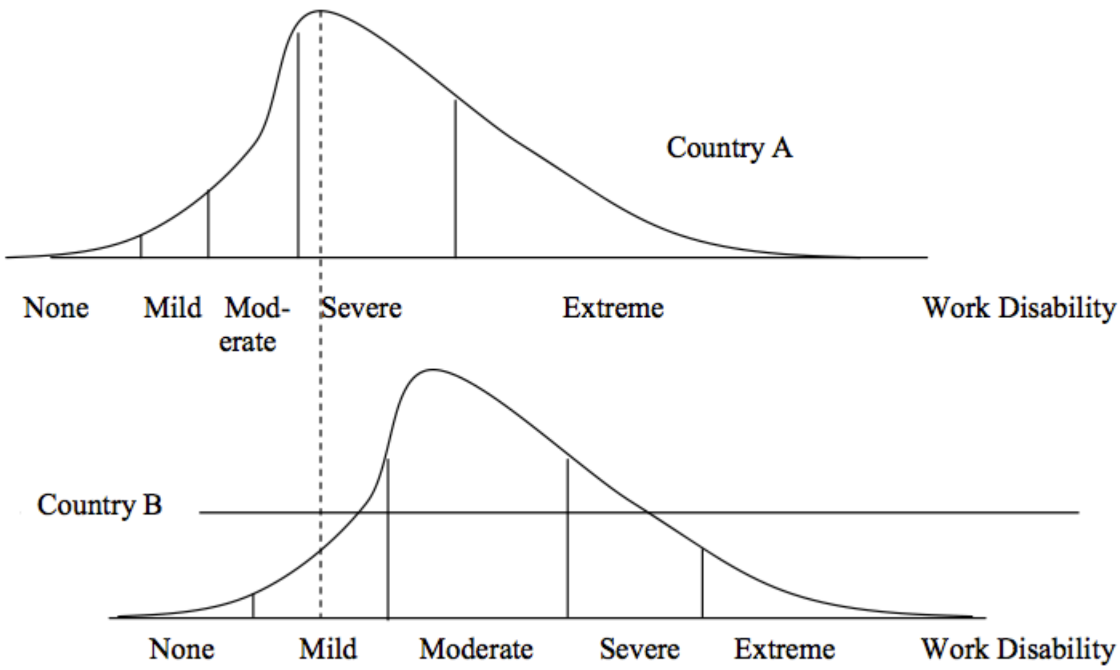
> Do you have any impairment or health problem that limits the kind or amount of paid work you can do? – no, not at all; yes, I am mildly limited; yes, I am moderately limited; yes, I am severely limited; yes, I am extremely limited, cannot work.

Researchers collect such subjective data because it may not be easy (or even impossible) to objectively measure the concept of interest (e.g., pain or political efficacy) or because it is too costly to obtain the objective measure in large surveys (e.g., visual acuity or work disability).

Analyses just based on such self-assessments are, however, not without problems. In particular, when the goal is to draw conclusions about actual differences, researchers have to assess to which extent the answers of survey respondents are interpersonally comparable. Do different respondents interpret the phrases "no say at all, little say, some say, a lot of say, unlimited say" in the same way? How do respondents interpret the phrase "any impairment or health problem?" The existing literature, does not give a reason to think that responses should be or are comparable, and in fact suggests that relying on them can be (extremely) misleading. For example, while comparing different states in India, Sen (2002) finds that the state of Kerala, which has the highest levels of literacy and longevity also has the highest rate of reported morbidity. By contrast, the state of Bihar, which has very poor medical and educational facilities, has the lowest rates of reported morbidity. King et al. (2004) provide two other examples, one on political efficacy and another one on visual acuity. See Manski (2004) and the references therein for a discussion of the (in)comparability of verbal expectations data.

To illustrate the problem of interpersonal incomparability of responses consider Figure 1, which is taken from Kapteyn et al. (2009). This figure gives the distribution of a continuous measure of health related work disability in two hypothetical countries. The distribution in country A is to the left of that in country B, implying that, on average, the population of country A suffers less from health related work disability than that of country B. All persons in both countries are asked to self-assess their health related work disability on a five-point scale (see the example question above). In Figure 1, persons in country A are much more pessimistic about a given amount of health related work disability than persons in country B. For example, a person in country A whose actual work disability is given by the dashed line would report to be "severely" disabled, whereas a person in country B with the same actual work disability would report to be only "mildly" disabled. Analyses based on the (frequency distribution of) self-assessments only would falsely conclude that persons in country A suffer more from health related work disability than those in country B.

Figure 1.1: Comparing self-reported health related work disability



Source: (Kapteyn, Smith, & Van Soest, 2009)

King et al. (2004) introduced anchoring vignettes as a tool to make the otherwise (interpersonally) incomparable self-assessments more comparable. An anchoring vignette is a short description of aspects of a hypothetical person's life which are relevant to the concept of interest. An example is the following "story" on work disability (Kapteyn et al., 2007)

> Mark has pain in his back and legs, and the pain is present almost all the time. It gets worse while he is working. Although medication helps, he feels uncomfortable when moving around, holding and lifting things at work. Does Mark have any impairment or health problem that limits the amount or kind of paid work he can do? – no, not at all; yes, he is mildly limited; yes, he is moderately limited; yes, he is severely limited; yes, he is extremely limited.

Suppose a vignette is constructed that describes a level of health related work disability for the hypothetical person corresponding to the dashed line in Figure 1. Persons in country A would evaluate the person in the vignette to be "severely" disabled, whereas persons in country B would report him to be "mildly" disabled. Since the actual level of health related work disability of the vignette person is the same for people in both countries, differences in evaluations must be due to heterogeneity in reporting behavior. Therefore, vignette evaluations help to identify the differences in reporting behavior. For example, using the scale of country A as the benchmark, the evaluations of country B can be expressed on this scale. This would lead to the correct conclusion that, on average, the population in country B suffers more from health related work disability than that in country A.

The anchoring vignette method requires two assumptions: (1) *response consistency,* which is the assumption that persons use the same reporting behavior for self-assessments and vignette evaluations; (2) *vignette equivalence,* which is the assumption that the level of the variable represented in the vignette is understood in the same way by all respondents.

The focus of a number of recent papers is on testing these assumptions and the "performance" of the anchoring vignette method (e.g., Bago d'Uva, Van Doorslaer, Lindeboom, & O'Donnell, 2011; Kapteyn, Smith, Van Soest, & Voňková, 2011; Van Soest, Delaney, Harmon, Kapteyn, & Smith, 2011). Chapter 2 fits into this line of research as it tests the sensitivity of the vignette method to the choice of the vignette. First, we study whether different vignettes within a certain domain (e.g., political efficacy or work disability) lead to similar adjusted self-assessments. We adjust self-assessments using one vignette at a time and then compute the correlation coefficient between any pair of adjusted self-assessments. If different vignettes lead to similar adjusted self-assessment, then this correlation coefficient will be close to one. This approach requires that more than one vignette is collected within a domain. Second, we study whether different adjusted self-assessments (again adjusted using one vignette at a time) are closer to the actual situation than the unadjusted self-assessments. The conclusions are mixed: for a sample of older persons (Survey of Health, Ageing, and Retirement in Europe) we find that the vignette method is sensitive to the choice of the vignette but that, in some health domains, a single vignette can make the self-assessments more comparable.

1.2 Health insurance and demand for medical care

The design of a health insurance system faces several challenges, among which problems of *asymmetric information* are the most important (see Cutler & Zeckhauser, 2000 and Zweifel & Manning, 2000 for surveys). When a person has health insurance, the price of medical services is partly or fully paid for by others. As a consequence, persons with more generous insurance coverage have less incentives to reduce the probability of illness through preventive care (ex-ante moral hazard), and they may also use more medical services than they would do when having less generous coverage (ex-post moral hazard). Competition in the health insurance market suffers from the problem that, persons who expect to use more medical care, ceteris paribus, choose a more generous insurance policy than those who expect to use fewer medical services. If insurance companies could charge a premium based on expected medical care use, then the market would efficiently sort people. However, such practices are generally not allowed since there is wide agreement that it is not fair to make people pay more just because they are sick. Insurance companies therefore can only charge average prices. Economic models of adverse selection predict that healthier people will be driven out of the market (Akerlof,

1970) or that they will be underinsured (Rothschild & Stiglitz, 1976).

Both models of moral hazard and adverse selection predict a positive correlation between insurance coverage and expenditures, which forms the basis for an empirical test of asymmetric information (Chiappori & Salanie, 2000). The failure to find a positive correlation, as for example in long-term care insurance (Finkelstein & McGarry, 2006), led to the idea that private information may be multi-dimensional rather than one-dimensional as assumed in the classic models. For example, persons may have information about both their risk type and their risk aversion. If those who are more risk averse buy more insurance and have lower risks this leads to what has been called "advantageous selection" (see e.g., Fang, Keane, & Silverman, 2008).

To test for moral hazard, it is ideal to use data from a randomized natural experiment, such as the RAND Health Insurance Experiment (Newhouse, 1993). Conducting such a randomized experiment is, however, typically not feasible because of financial constraints, ethical considerations, or other reasons. Non-experimental studies are particularly valuable when persons experience an unexpected and exogenous shock in the incentive structure they face.

In chapter 3 we exploit an unusual feature of the German health insurance system which allows us to control for selection into private insurance: as soon as income in the last year exceeds the so-called compulsory insurance threshold, persons become eligible to opt out of the public health insurance system and buy private insurance instead. Random variation in income around this compulsory insurance threshold provides a natural experiment that allows us to conduct a regression discontinuity analysis. Controlling for selection into private insurance, we find a significant negative effect of being privately insured on the number of doctor visits for those who visit the doctor at least once in a three month period. At the same time, we find no significant effect on the number of nights spent in a hospital, which can arguably be influenced less by a person, and a positive effect on self-assessed health. This suggests that privately insured patients receive better or more intense treatment each time they see a doctor, or that they invest more in prevention.

1.3 Health and demand for medical care

Traditional demand theory assumes that each consumer has a utility function that allows him or her to rank alternative combinations of goods and services purchased in the market. The consumer is supposed to purchase the combination of goods and services that maximizes his or her utility function subject to a budget constraint. It has been long recognized that this traditional model may not provide a satisfactory explanation of the demand for medical goods and services, because what consumers demand when they purchase these services are not these services per se but rather better health. Although long recognized, until Grossman (1972a, 1972b) the distinction between health as an output and medical care as an input had not been formalized. In Grossman's human capital framework a person invests in health (e.g., invest time and consume medical goods and services) for the consumption benefits (health provides utility) as well as production benefits (healthy persons have greater earnings).

The model provides a conceptual framework for interpretation of the demand for health and medical care in relation to a person's resource constraints, preferences, and consumption needs over the life cycle.

While Grossman's model has great theoretical and intuitive appeal and has led to a rich body of literature and many useful insights in health economics, it also has several limitations. For example, (i) in empirical work it is generally found that health and the demand for medical care are negatively related, whereas Grossman's model predicts a positive relationship (Wagstaff, 1986a; Zweifel & Breyer, 1997; Galama & Kapteyn, 2011); and (ii) empirically, health declines faster for persons with lower socio-economic status, and the model does not predict this (Case & Deaton, 2005). See chapter 4 for other limitations that have been identified within this literature. Grossman (2000) provides a review and rebuttal of some of the limitations.

Despite the limitations, theoretical extensions and competing economic models are still relatively few. Promising adaptations of the model are the models of Ehrlich and Chuma (1990) and Galama (2011), who have extended the Grossman model to include a health production process that is characterized by decreasing returns to scale (DRTS), whereas the standard model assumes a linear health production function with constant returns to scale (CRTS). Introducing decreasing returns removes the limitations of the Grossman model mentioned above (Galama, 2011): The model with DRTS predicts (i) a negative correlation between health investment and health; and (ii) that the wealthy and educated live longer and experience slower declines in health.

Empirical tests of the health production literature have thus far been based on the equilibrium equation derived under the assumption of a linear health production process (e.g., Grossman, 1972b; Wagstaff, 1986a). In chapter 4 we test the predictions of a theory of health capital with decreasing returns to scale in health production. To this end, we employ an equation for health investment that was derived by Galama (2011). We estimate this equation using the Panel Study of Income Dynamics (PSID) and contrast our findings with those of a relatively small existing empirical literature.

The contribution of this chapter is as follows. First, to the best of our knowledge, no prior attempts have been made to empirically confront the predictions of a theory of health capital with decreasing returns to scale. Second, we carefully account for the endogenous nature of health in the demand for medical care. Only a few papers in the empirical literature have estimated direct relationships between medical care and health, and none of the papers that have tested the predictions of health capital theory have attempted to address the inherent endogeneity of health.

We obtain a statistically significant negative coefficient of health when the number of nights spent in a hospital is the dependent variable and when we do not take the endogeneity of health into account. Similarly, in the models for out-of-pocket medical expenditures or total medical expenditures, the dollar amounts are negatively related to health and statistically significant, although we do not find an effect for the participation equations (i.e., whether expenditures are positive or zero). This closely resembles the methodology and the findings in the literature and appears to support decreasing returns to scale, which predicts a negative coefficient, whereas constant returns to scale is associated with a positive coefficient. However, when we attempt to control for the endogeneity of health by using instrumental variables methods, using childhood health and parental smoking during childhood as instruments, the coefficients become statistically insignificant and not consistently negative.

1.4 Job search requirements and job finding rates

Receiving UI benefits is, in the Netherlands, as in most countries, conditional upon meeting criteria such as "availability for work" and "actively searching for a job" (Grubb, 2001). Such job search requirements may affect the job finding rate in several ways. First, they increase job search intensity among UI recipients for whom the optimal search intensity is less than the required minimum number of contacts (because of an increase of the probability to get a sanction). UI recipients for whom the optimal search intensity is higher than the required minimum number of contacts are not affected by the introduction of the requirements. Second, UI recipients who perceive the search requirements as a burden experience a raise in the non-monetary costs of continued receipt of UI benefits. This lowers the value of unemployment, thereby increasing search effort, or reducing the reservation wage, or both. The increased costs of continued UI benefit receipt may decrease search effort among persons who already met the new search requirements by means of informal search (Van den Berg & Van der Klaauw, 2006).

Meyer (1995) gives an overview of a number of field experiments that have been conducted in the United States to evaluate alternative job search policies. These experiments typically combined more strict search requirements with job finding services, making it difficult to determine the relative importance of each of the measures. A few field experiments, which are discussed in some detail in chapter 5 have been conducted that were explicitly designed to study the effects of alternative search requirements. The conclusions are mixed: Johnson and Klepinger (1994) and Klepinger, Johnson, and Joesch (2002) conclude that stricter search requirements reduce the length of the unemployment spell, but Ashenfelter, Ashmore, and Deschênes (2005) do not find evidence for this.

Job search criteria are, however, often less stringent for older workers in many European countries (OECD, 2006). In the Netherlands, UI recipients were for a long time exempted from the requirement to actively search for a job when they reached the age of 57.5. The reason for this was a relatively high unemployment rate among the young and the belief that older workers were holding their jobs.

In chapter 5 we study how prior to January 2004 the exemption from the search requirement affected job search behavior of the UI recipients involved. We find that the removal of the search requirement had a large negative effect on the job finding rate. Furthermore, there is some evidence that already some time before the search requirement was removed the job finding rate goes down. Unemployed workers who are getting close to the age of 57.5 seem to reduce their search intensity in anticipation of the removal of the search requirement. Doing a similar analysis for workers who became unemployed after 1 January 2004 we do not find such effects.

Although the absolute increase in the job finding rates among older workers for whom the search requirement is reinstalled is rather small, the fact that there is an increase at all is remarkable given the relatively weak labor market position of older workers. Apparently, even older workers have some influence over their job finding. Extrapolating our findings to younger age categories it is clear that it is important to have well specified search requirements which should be enforced to make sure that UI recipients keep searching for a job irrespective of how long they have been unemployed. Even workers with seemingly poor job prospects seem to benefit from the requirement to actively search for a job.

Chapter 2

Is the anchoring vignette method sensitive to the domain and choice of the vignette?

2.1 Introduction

Survey respondents are commonly asked to self-assess their health, level of work disability, job/life satisfaction, and other concepts. Consider, for example, the typical survey question that asks respondents to self-assess their health: *"Would you say your health is ...,"* with answers ranging from *"very bad"* to *"very good."* Researchers frequently use the answers to these questions to study differences between countries or between groups within a country. When the goal is to draw conclusions about actual differences, the results from direct comparison of self-assessments may be biased when respondents use the response categories in different ways. This interpersonal incomparability is referred to in the literature as differential item functioning (DIF) or as heterogeneity in reporting behavior.

King et al. (2004) introduced anchoring vignettes as a tool to correct self-assessments for heterogeneity in reporting behavior. An anchoring vignette is a short description of aspects of a hypothetical person's life which are relevant to the domain of interest. In practice this means that survey respondents not only assess their own situation but also the situation described in the vignette. Intuitively, the method can be understood as follows: since the situation described in the vignette is the same for every respondent, vignette evaluations provide information about response styles of respondents. Therefore, we can identify heterogeneity in reporting behavior and adjust self-assessments for it. The anchoring vignette method requires the following two assumptions: (1) *response consistency,* which is the assumption that persons use the same reporting behavior for self-assessments and vignette evaluations; (2) *vignette equivalence,* which is the assumption that the level of the variable represented in the vignette is understood in the same way by all respondents.

The vignette method has been applied in different domains like politics (e.g., King et al. (2004), Hopkins and King (2010)), health (e.g., Salomon, Tandom, and Murray (2004), Bago d'Uva, Van Doorslaer, Lindeboom, and O'Donnell (2008)), work disability (e.g., Kapteyn et al. (2007)), satisfaction with the health care system (e.g., Murray et al. (2003), Sirven, Santos-Eggimann, and Spagnoli (2008)). See Gary King's website (http://gking/harvard.edu/vign/eg) for a large collection of vignettes used in different settings.

This chapter studies the validity of the parametric model for the anchoring vignette method, called the CHOPIT model. This model is used most often in applications of the method. In addition to the response consistency and vignette equivalence assumptions, the CHOPIT model makes functional form and distributional assumptions. See section 2.3 for more details. If any assumptions of the model do not hold, we can get wrongly adjusted self-assessments. Rather than testing the assumptions of the model separately, we take the following two approaches to assess the validity of the anchoring vignette method.

First, we study whether different vignettes within a certain domain lead to similar adjusted self-assessments. This idea requires that more than one vignette is collected within a domain, which is the case for the data we use. We estimate the CHOPIT model using one vignette at a time and we then compute the correlation coefficient between any pair of DIF-adjusted self-assessments within a domain. If different vignettes lead to similar adjusted self-assessments then the correlation coefficient between any pair of DIF-adjusted self-assessments will be close to one. As far as we are aware this approach has not been taken before.

The first approach is uninformative about the question whether different DIF-adjusted self-assessments are closer to the actual situation than unadjusted self-assessments. That is what we study in the second approach, details of which are discussed in section 2.3. Assessing the validity of the anchoring vignette method by means of a measure of the actual situation has been suggested before in the literature. In fact, we follow Van Soest et al. (2011). The novelty of our approach is that we study the performance of a single vignette, as we did in the first approach. The credibility of this approach is closely connected with the quality of the chosen objective measure(s) as well as with the assumptions of the CHOPIT model. The quality of the objective measure depends on how closely it corresponds with the health dimension elicited in the self-assessment and vignette question(s). If the correspondence is strong, then the results of this approach show whether or not the DIF-adjusted self-assessments are closer to the actual situation than the unadjusted self-assessments. If the correspondence is weak, then the results of this approach may not be valid. In that case, or in the case that there is no measure of the actual situation available, results of our first approach will still indicate whether or not the vignette method is sensitive to the choice of the vignette (as long as more than one vignette is collected).

Studying whether the method is sensitive to the choice of the vignette and studying whether DIF-adjusted self-assessments are closer to an objective measure than unadjusted self-assessment are both important issues, because researchers who apply the method should be confident that the method works properly.

The first comparison of unadjusted and DIF-adjusted self-assessments with a measure of the actual situation is, as far as we are aware, reported by King et al. (2004). They use self-assessments and vignette evaluations on visual acuity collected by the WHO for China and Slovakia. On average, the self-assessments do not show a significant difference in visual acuity between Chinese and Slovak respondents. However the measured test for vision - the Snellen Eye Chart test - shows that respondents from China have, on average, substantially worse vision than those from Slovakia. Self-assessments are adjusted using the eight vignette evaluations simultaneously. Comparison of these DIF-adjusted self-assessments confirms the conclusion from the measured test.

Van Soest et al. (2011) propose a formal test for the response consistency assumption. Additionally, they show whether the distribution of DIF-adjusted self-assessments is “closer” to the distribution of an objective measure than the unadjusted distribution. They collected self-assessments and four vignette evaluations on drinking behavior among students at a large Irish university. They use the self-reported number of drinks typically consumed per occasion as a measure of actual drinking behavior. Self-assessments are adjusted using all four vignettes simultaneously. Their results suggest that allowing and adjusting for heterogeneous reporting behavior, as well as assuming response consistency, substantially improves the fit of the model as well as the correlation between the self-assessments and objective measure.

Datta Gupta, Kristensen, and Pozzoli (2010) take a similar approach as Van Soest et al. (2011), using data from the first wave of the Survey of Health, Ageing and Retirement in Europe (SHARE). The paper focuses on work disability. Self-assessments are adjusted using all nine vignette evaluations simultaneously and grip strength is used as an objective measure. Their finding is that DIF-adjusted self-assessments are not closer to the objective measure than the unadjusted self-assessments. A potential explanation for the contradictory conclusions of Van Soest et al. (2011) and Datta Gupta et al. (2010) is that grip strength does not correspond closely enough to work disability, whereas the self-reported number of drinks measure of Van Soest et al. (2011) does to drinking behavior. Another explanation may be that the vignettes used in Van Soest et al. (2011) correspond better to the studied domain than the vignettes used in Datta Gupta et al. (2010).

Using data from two waves (2004 and 2007) of SHARE we study the validity of the vignette method for three domains not studied before: cognition, breathing and mobility. SHARE collected data on self-assessments, vignette questions and objective measures for these three domains. More details about the data are given in section 2.2.

For cognition we find that different vignettes lead to different adjusted self-assessments. One vignette brings the self-assessment closer to a measure of actual cognition, while two others do not. For breathing we find that different vignettes lead to different adjusted self-assessments. However, the vignette collected in the 2007 wave brings the self-assessment closer to a measure of actual breathing. Our findings for mobility are most encouraging. Here, all vignettes bring the self-assessment closer to a measure of actual mobility.

The remainder of this chapter is organized as follows. Section 2.2 provides more information on the SHARE data. The models as well as our two approaches to validate the vignette method are discussed in section 2.3. Results are discussed in section 2.4 and section 2.5 concludes.

2.2 Data

The chapter uses data from two waves of SHARE, a nationally representative sample of the population 50 years and older, which provides detailed information on health, socioeconomic status, and social and family networks of more than 45,000 persons. The first wave of data was collected in 2004 in eleven European countries. The second wave of data was collected in 2006-07 in the same eleven countries and an additional three countries.

For each of the three health domains we focus on, three vignette questions were collected in the first wave. By contrast, the second wave collected one vignette per domain, which was chosen out of the three from the first wave. In both waves the data on self-assessments and vignette questions are only collected in subsamples of the overall SHARE samples. In the remainder of this chapter we will refer to these subsamples as the vignette samples. Self-assessments and vignette evaluations were collected in both waves for Belgium, France, Germany, Greece, Italy, the Netherlands, Spain and Sweden, and only in the second wave for the Czech Republic, Denmark and Poland. In Greece, self-assessments and vignette evaluations were collected in both waves, but the data of the second wave were not available in the release we use. The SHARE data also contains information on objective measures for all three domains. The objective measures for cognition are available in both waves, whereas those for breathing and mobility are only available in the second wave. Moreover, only respondents younger than 75 were asked to participate in the objective measurement task for mobility.

To keep as much information as possible we use a different sample for each (domain,wave) combination. That is, we have a sample for cognition for wave 1, and another sample for cognition for wave 2. This is also the case for the other two domains. Each (domain, wave) sample is selected on the basis of the self-assessment and vignette(s) available for that combination, and on the basis of available objective measure(s) and a set of common covariates. The samples are generally distinct. Theoretically, conclusions found for the whole population can also be expected for any sample of the population. Conclusions found for a sample of the population, in particular a nonrepresentative sample, can only be seen as an indication for the conclusions to be true in the whole population. This should be kept in mind while working with different samples.

Descriptive statistics for the self-assessments, vignette evaluations and objective measures, as described below, are based on the relevant (domain,wave) sample. Descriptive statistics of the covariates are based on the vignette samples of both waves. As it turns out only for the mobility sample of the second wave the distribution of covariates is substantially different from the one for the vignette sample of the wave 2. Below we discuss on which aspects it differs.

2.2.1 Self-assessments and vignette ratings

To begin, consider as an example the self-assessment question for cognition:

> "*Overall in the last 30 days how much difficulty did you have with concentrating or remembering things?*" (none, mild, moderate, severe, extreme)

The self-assessment questions for the other domains are formulated similarly. See Appendix 2.B for the exact wording, which is the same in both waves. As noted already, the vignette collected in the second wave of SHARE was chosen out of the three vignettes collected in the first wave. Each vignette describes aspects of a hypothetical person's life relevant to the domain. To give an example, consider the cognition vignette collected in both waves:

> "*(Lisa) can concentrate while watching TV, reading a magazine or playing a game of cards or chess. Once a week she forgets where her keys or glasses are, but finds them within five minutes. Overall in the last 30 days, how much difficulty did (Lisa) have with concentrating or remembering things?"* (none, mild, moderate, severe, extreme)

The exact wording of all three anchoring vignettes collected for each of the three domains studied in this chapter is given in Appendix 2.B.

The percentage of missing observations for the self-assessments and the vignette questions is very low. It is at most 1.9 percent and also very similar across countries. Descriptive statistics for the self-assessments and vignette evaluations are given in Table 2.1 for both waves. In all cases, most respondents report to have either no or only a mild problem. Few respondents report to have a severe or extreme problem. From the vignette evaluations of wave 1 it becomes clear that, within each domain, the vignette numbered "1" is, on average, considered to be the vignette describing the mildest problem within that domain. The vignette numbered "2" describes, on average, a more severe problem than the vignette numbered "1", and the vignette numbered numbered "3" described, on average, the most extreme health problem. The vignettes labeled c1, b1, and m1 were collected in both waves.

Tables 2.2, 2.3, 2.4, and 2.5 show the answers to the self-assessment and vignette questions by country for each domain and wave.

In our empirical analyses, we always combine the two categories "severe" and "extreme" for both self-assessments and vignette evaluations, because especially in the latter category there are few observations.

2.2.2 Objective measures

One of the two validation approaches taken in this chapter studies whether DIF-adjusted self-assessments are closer to a measure of the actual situation than unadjusted self-assessments. The measures we use are discussed below for each domain. The exact wording of the questions used to collect the objective measures can be found in Appendix 2.B.

Cognition For cognition we use the following four objective measures: immediate and delayed verbal memory, verbal fluency and numerical ability. The variables used to measure cognitive functioning in SHARE are very similar to those used in other, well-known, surveys such as the English Longitudinal Study of Ageing (ELSA), the Asset and Health Dynamics Among the Oldest Old (AHEAD) study and the Health and Retirement Study (HRS). See for example Mehta et al. (2003), Llewellyn, Lang, Langa, and Huppert (2008), and Herzog and Wallace (1997). Immediate and delayed verbal memory were assessed using a 10-word learning task. Respondents were being read a list of 10 words by an interviewer. Then, they were asked to recall as many words as possible, immediately afterwards, and after a short delay during which they answered other questions that assess cognitive functioning. Verbal fluency was examined by letting respondents name as many different animals as they could think of in one minute. Numerical ability was assessed using four different questions involving simple calculations. To give an example, consider the first question asked to all respondents:

> *"If the chance of getting a disease is 10 per cent, how many people out of 1000 (one thousand) would be expected to get the disease?"*

All objective measures described above are available in both waves. Whereas immediate and delayed recall seem to be closely related to the cognition question, which asks about "concentrating and remembering things," numeracy and verbal fluency seem to be less related. Still we included them into our analyses to see whether the results would be the same. Anticipating our results, we find that the conclusions are the same irrespective of the objective measure used, except for verbal fluency for wave 1.

The percentage of missing observations for the four objective measures of cognition, as given in Table 2.6, is very low. Table 2.7 shows descriptive statistics for the objective measures for cognition for both waves. It reveals that in both waves most respondents are able to immediately recall four words or more, however few respondents recall more than seven words immediately. As expected, after a short delay, respondents recall fewer words. Most of them recall two up to five words after a short delay. The median score for numeracy is three in both waves. The number of animals respondents can mention in one minute is nineteen on average.

Table 2.8 shows descriptive statistics by country for delayed recall information collected in the second wave.

In our empirical analyses we merge, for the objective measures, categories that contain few observations. Immediate recall is coded into five different categories; the first and last four categories are each merged into one category. Delayed recall is coded into four different categories; each of the groups zero-one, two-three, four-five, six-ten is used as a single category. Numeracy is coded into four different categories by merging the first two. For verbal fluency (the number of animals mentioned) we use quintiles to divide the respondents into different groups. The quintiles are: 12, 16, 20, 24 and 67. (We also estimated the models with the respondents divided into more groups (7 and 10 groups). The results for these divisions were basically identical to the ones reported in this chapter.)

Breathing As objective measure for breathing we use the result of a so-called peak flow test, which measures the "maximum flow during an expiration delivered with maximal force starting from the level of maximal lung inflation," (Quanjer, Lebowitz, Gregg, Miller, & Pedersen, 1997). Respondents are asked to do the test twice. If two measurements are available for a respondent we take the maximum value, otherwise we take the single measurement available as value. The unit of measurement of the peak flow test is liter/minute and it ranges from 60 to 880. Note that this measure is only available for the second wave. Table 2.6 reports the percentage of missing observations for the peak flow test. The most important reason why an observation is missing is that the respondent thinks it is not safe to do the test. Descriptive statistics for the peak flow test are provided in Table 2.9. Descriptives for the peak flow test by country are given in Table 2.10.

Mobility As objective measure for mobility we use the result of a so-called stand-up test. This test asks respondents to fold their arms across their chest and keep it like this while standing up. Our measure is the time in seconds needed for five stands.

The percentage of missing observations for the stand-up test is given in Table 2.6. This percentage is computed for the group of respondents younger than 75, as only they are asked to participate. Approximately 82 percent of the respondents of the vignette sample of wave 2 is younger than 75. The most important reason that an observation is missing is either that the respondent thinks it is not safe to do a single test, or that s/he is not able to do the single test according to instructions (having their arms folded across their chest), or because the respondent thinks it is not safe to stand up five times.

Descriptive statistics for the stand up test are given in Table 2.9. On average, respondents need 11 seconds to finish the test. Table 2.10 shows descriptive statistics for the stand-up test by country.

2.2.3 Covariates

The parametric version of the anchoring vignette method models the actual level of health and reporting heterogeneity using a vector of covariates. The model will be discussed in more detail in section 2.3. In this chapter we include the following covariates, which are commonly used in applications of this model to health: country, age in groups of 5 years, gender, low/mid/high education, living alone, suffering from a long-term illness, never/sometimes/often engaged in physical activity. The education variable is based on the International Standard Classification of Education. The long-term illness variable is a binary variable indicating whether or not a respondent has a long-term health problem and for that reason does not distinguish between the number and severity of chronic illnesses. In the models for mobility we exclude physical activity as a covariate because of potential endogeneity problems. Further information regarding the “construction” of our covariates can be found in Appendix 2.C. Important to note here is that reference groups differ across our empirical analyses, but we always follow the rule that the group with most observations is taken as the reference group.

Descriptive statistics of the covariates are given for the vignette samples of both waves and not for the (domain,wave) specific samples and shown in Table 2.11. Descriptive statistics of the covariates for each of the (domain,wave) specific samples, in addition to those of the vignette samples, are given in Table 2.12. This table reveals that all (domain,wave) specific samples, except the one for mobility in wave 2, are similar to the corresponding vignette samples. The mobility sample contains fewer observations, respondents are slightly better educated, live less often alone, suffer less often from a long-term illness, are more often engaged in physical activity, and are on average younger. The different composition of this sample is likely to be due to the selection rules for the objective measure (stand-up test).

Descriptive statistics for each country separately are given in Tables 2.13 and 2.14. Note that these are for the vignette samples of both waves.

2.3 Model

One of the two approaches taken in this chapter to validate the anchoring vignette method requires the availability of an objective measure. We follow Van Soest et al. (2011), who extend the compound hierarchical ordered probit (CHOPIT) model, by also modeling the objective measure. In the model below, a subscript s denotes self-assessment, and subscripts v and o denote vignette(s) and objective measure, respectively.

2.3.1 Model for self-assessments

The self-assessment, y_{si}, of the ith person is modeled as an ordered response equation with latent variable:

$$y_{si}^{*} = \mathbf{x}_i'\beta_s + \varepsilon_{si},$$

where $\mathbf{x}_i$ is a vector of covariates including a constant term, and β_s a vector of parameters. The error term, ε_{si}, is assumed to be normally distributed with mean zero and variance σ_s^2, and independent of the covariates $\mathbf{x}_i$. The reported and observed responses, y_{si}, are generated by the following mechanism:

$$y_{si} = k \Leftrightarrow \tau_{si}^{k-1} < y_{si}^{*} \leq \tau_{si}^{k}, \qquad k = 1, \ldots, K$$

where $-\infty = \tau_{si}^{0} < \tau_{si}^{1} < \ldots < \tau_{si}^{K} = +\infty$. The thresholds are modeled as:

$$\begin{aligned} \tau_{si}^{1} &= \mathbf{x}_i'\gamma_s^1 + u_i, \\ \tau_{si}^{k} &= \tau_{si}^{k-1} + \exp\left(\mathbf{x}_i'\gamma_s^k\right), \quad k = 2, \ldots, K-1, \end{aligned}$$

where $\mathbf{x}_i$ is a vector of covariates, and γ_s^k, for $k = 1, \ldots, (K-1)$, are vectors of parameters. The random effect, u_i, is assumed to be normally distributed with mean zero and variance σ_u^2, and independent of the covariates $\mathbf{x}_i$.

The idea that reporting behavior varies across persons is formalized by modeling the thresholds to be person-specific. The latent variable, y_{si}^{*}, can be interpreted as the true level of health as perceived by the person. Note that using only self-assessments, the parameter vectors β_s and γ_s^1 are not separately identified, but the parameter vectors γ_s^k, for $k > 2$, are. That is, using only self-assessments we are not able to "decompose" the self-assessments in a part that is due to differences in "true" health (β) and a part due to heterogeneity in reporting behavior ($\gamma_s^k,\ k = 1, \ldots, K-1$).

2.3.2 Model for vignettes

The actual level of health for the hypothetical person described in vignette v is denoted ϑ_v, $v = 1, \ldots, V$. Assuming that it is the same for every person formalizes the vignette equivalence assumption. Each respondent perceives the actual level of health only with random error, i.e.:

$$y^*_{vi} = \vartheta_v + \varepsilon_{vi},$$

where the error term, ε_{vi} is assumed to be normally distributed with mean zero and variance σ^2_v and independent of the covariates $\mathbf{x}_i$. The observed vignette evaluations are generated by the following mechanism:

$$y_{vi} = k \Leftrightarrow \tau^{k-1}_{vi} < y^*_{vi} \leq \tau^k_{vi}, \qquad k = 1, \ldots, K$$

where $-\infty = \tau^0_{vi} < \tau^1_{vi} < \ldots < \tau^K_{vi} = +\infty$. The thresholds are modeled similarly as in the self-assessment model, i.e.:

$$\begin{aligned} \tau^1_{vi} &= \mathbf{x}'_i \gamma^1_v + u_i, \\ \tau^k_{vi} &= \tau^{k-1}_{vi} + \exp\left(\mathbf{x}'_i \gamma^k_v\right), \qquad k = 2, \ldots, (K-1), \end{aligned}$$

where the term u_i is assumed to be the same in the thresholds of the self-assessment and vignette model. It introduces unobserved heterogeneity and implies that for each person the vignette evaluation is correlated with the self-assessment (conditional on the covariates $\mathbf{x}_i$).

The response consistency assumption is formalized by assuming: $\tau^k_{si} = \tau^k_{vi}$, for $k = 1, \ldots, K-1$; $v = 1, \ldots, V$. In terms of the parameters this amounts to assuming $\gamma^k_s = \gamma^k_v$, for $k = 1, \ldots, K-1$; $v = 1, \ldots, V$.

2.3.3 Model for objective measure

To study whether the anchoring vignette method brings self-assessments closer to the objective situation we use measures of the objective situation. The four objective measures for cognition: immediate and delayed verbal memory, numeracy and verbal fluency, are all discrete variables. In that case we model our objective measure as follows:

$$\begin{aligned} y^*_{oi} &= \mathbf{x}'_i \beta_o + \varepsilon_{oi}, \\ y_{oi} &= l \Leftrightarrow \tau^{l-1}_o < y^*_{oi} \leq \tau^l_o, \end{aligned}$$

where $-\infty = \tau_o^0 < \tau_o^1 < \ldots < \tau_o^L = +\infty$, are unknown thresholds that are the same for all persons. The thresholds are modeled as:

$$\tau_o^1 = \exp\left(\gamma_o^1\right),$$
$$\tau_o^l = \tau_o^{l-1} + \exp\left(\gamma_o^l\right), \qquad l = 2, \ldots, (L-1).$$

The objective measures for breathing and mobility, the result of the peak flow test and the stand-up test, respectively, are continuous variables, and are modeled as follows:

$$y_{oi} = \mathbf{x}_i' \beta_o + \varepsilon_{oi}.$$

In both cases, discrete and continuous, the error term ε_{oi} is assumed to be independent of the covariates, $\mathbf{x}_i$, the unobserved heterogeneity term, u_i, and the error term of the vignette model, ε_{vi}. However, ε_{oi} is allowed to be correlated with the error term of the self-assessment model, ε_{si}, because the covariates might not capture all variation in "true" health, y_{si}^* and y_{oi}^*. The distribution of $(\varepsilon_{si}, \varepsilon_{oi})$ is assumed to be bivariate normal with mean zero, variances σ_s^2 and σ_o^2, and correlation ρ.

2.3.4 Likelihood

The likelihood contribution of the i-th person conditional on the unobserved heterogeneity, u_i, can be written as the product of a joint normal probability for the self-assessment and the objective measure, and a normal probability for the vignettes. In case of a discrete objective measure, the unconditional likelihood contribution for person i is given by

$$\int \prod_{k=1}^{K} \prod_{l=1}^{L} \prod_{v=1}^{V} \prod_{m=1}^{M} P\left(y_{si} = k, y_{oi} = l | \varphi, u_i\right)^{1(y_{si}=k, y_{oi}=l)} P\left(y_{vi} = m | \varphi, u_i\right)^{1(y_{vi}=m)} f\left(u_i\right) du_i,$$

where $f(.)$ is the normal density function with variance σ_u^2, and $1(.)$ the indicator function. The vector of parameters is $\varphi = \left(\beta', \sigma_s^2, \gamma_s', \sigma_u^2, \vartheta, \gamma_v', \sigma_{v=1}^2, \ldots, \sigma_{v=V}^2, \beta_o', \tau_o', \sigma_o^2, \rho\right)'$.

2.3.5 Identification

We use three different models to study whether the vignette method is sensitive to the domain and the choice of the vignette, and is consistent over time. Each of these three models can be considered as a "special case" of the model discussed in the previous subsections, and every model has a different set of identifying assumptions.

CHOPIT model For identification reasons the constant term of β_s equals zero and the variance of the error term in the self-assessment part is normalized to one, i.e., $\beta_{s,1} = 0, \sigma_s^2 = 1$. Because of response consistency we assume that $\gamma_s^k = \gamma_v^k \left(= \gamma^k\right)$ for $k = 1, \ldots, (K-1)\,; v = 1, \ldots, V$. When multiple vignettes are used, we assume $\sigma_v^2 = \sigma_V^2$, for $v = 1, \ldots, (V-1)$.

Model A (No DIF, No RC) Reporting behavior is assumed to be homogeneous across respondents and vignettes, therefore $\tau_{si}^k = \tau_s^k$ and $\tau_{vi}^k = \tau_V^k$, for $k = 1, \ldots, (K-1)\,; v = 1, \ldots, V$, and $\sigma_u^2 = 0$. The model does not impose response consistency. That is, it allows for the possibility that $\tau_s^k \neq \tau_V^k$ for $k = 1, \ldots, (K-1)$. However, for identification reasons $\tau_s^1 = \tau_V^1 = 1$. The variances of the error terms in both the self-assessment and vignette model are normalized to one, i.e., $\sigma_s^2 = \sigma_v^2 = 1$, for $v = 1, \ldots, V$. In case of a discrete objective measure, the first two thresholds of the objective measure are equal to one and two, i.e., $\tau_o^1 = 1, \tau_o^2 = 2$, for identification reasons.

Model B (DIF, RC) This model allows reporting behavior to be heterogeneous across persons and therefore the thresholds are allowed to be person-specific. In addition it assumes that response consistency holds, i.e., $\gamma_s^k = \gamma_v^k \, (= \gamma^k)$, for $k = 1, \ldots, (K-1)\,; v = 1, \ldots, V$. Furthermore the model normalizes the constant term in the parameter vector of the first threshold to one, i.e., $\gamma_{s,1}^1 = \gamma_{v,1}^1 = 1$, for $v = 1, \ldots, V$. The variance of the error term in the self-assessment model is normalized to one, $\sigma_s^2 = 1$. When using multiple vignettes we assume $\sigma_v^2 = \sigma_V^2$, for $v = 1, \ldots, V$. In case of a discrete objective measure, the first two thresholds of the objective measure are equal to one and two, i.e., $\tau_o^1 = 1, \tau_o^2 = 2$, for identification reasons.

2.3.6 Two validation approaches

As already discussed in the introduction, this chapter takes two approaches to validate the parametric model for anchoring vignettes. Here we explain our approaches in more detail.

As a first step in validating the vignette method, we investigate whether different vignettes lead to similar DIF-adjusted self-assessments. For each domain we estimate the CHOPIT model using one vignette at a time and compute the DIF-adjusted self-assessments. That is, we compute the predicted systematic parts: $\hat{y}_{si}^* = \mathbf{x}_i'\hat{\beta}_s$. Since three vignettes were collected for each health domain in the first wave, this gives us a set of three different DIF-adjusted self-assessments. Then we compute the correlation coefficient between any pair of DIF-adjusted self-assessments within each domain. If different vignettes lead to similar DIF-adjusted self-assessments the correlation coefficient between any pair of DIF-adjusted self-assessments (each based on a single vignette) will be close to one.

These correlations are, however, uninformative about the question whether DIF-adjusted self-assessments are "closer" to the objective situation than unadjusted self-assessments. As a second step we therefore estimate the models A and B discussed in the previous section. Model A does not allow (and adjust) for heterogeneity in reporting behavior, whereas model B does. The models are estimated for each domain separately using one vignette at a time. Each time we compute the correlation coefficients between the predicted systematic parts of (y_{si}^*, y_{oi}^*) and between the simulated values of (y_{si}^*, y_{oi}^*). The simulated values add a draw from a bivariate normal distribution, using estimates of the variances and correlation, to the predicted systematic parts. If the DIF-adjustment would bring the self-assessments closer to the actual situation, then the correlation coefficient given by model B would be higher than the corresponding one given by model A.

2.4 Results

Cognition First, in Table 2.15 we report the correlations between different DIF-adjusted self-assessments, each based on one vignette. The correlations indicate that vignettes c2 and c3 lead to similar DIF-adjusted self-assessments, whereas those based on vignette c1 are different from the other two. All this reveals is that for cognition the DIF-adjustment is sensitive to the choice of the vignette. In Table 2.18 we report parameter estimates of the CHOPIT model using a single vignette, as well as using all three vignettes simultaneously.

Second, we estimate the models A and B separately using data from wave 1 and wave 2. For wave 1, the models are estimated for each combination of one of the four objective measures and one of the three vignettes. We also estimated the models using all three vignettes simultaneously. For wave 2, the models are estimated for each of the four objective measures using the single vignette that is collected. Table 2.16 and 2.17 provide a summary of results based on data from wave 1 and wave 2, respectively.

Consider first the results for wave 1. We only discuss them for delayed recall, as they are consistent with the other objective measures and consistent over time, except for verbal fluency using data from wave 1. (In the latter case, results indicate that for all three vignettes, DIF-adjusted self-assessments are closer to the objective situation.)

When vignette c1 is used the results show that the model which corrects for reporting behavior heterogeneity (model B) gives a correlation between the predicted systematic parts of (y^*_{si}, y^*_{oi}) of 0.52 compared to 0.75 for the model that does not make this correction (model A). The correlations between the simulated values of (y^*_{si}, y^*_{oi}) are 0.22 and 0.26 for model B and A respectively. So, both correlation coefficients are lower for model B than for model A. We therefore conclude that DIF-adjusted self-assessments based on vignette c1 are more different from the objective situation than the unadjusted self-assessments.

Adjusting self-assessments using vignette c2 leads to a different conclusion. For both models the correlation between the predicted systematic parts is comparable: 0.75 for model A and 0.74 for model B. The correlation between the simulated values increases from 0.26 for model A to 0.30 for model B. On the basis of these correlations we conclude that the DIF-adjusted self-assessments based on vignette c2 are about as close to the objective situation as the unadjusted self-assessments.

Consider next the results when vignette c3 is used. The correlation between the predicted systematic parts increases from 0.75 for model A to 0.83 for model B, and the correlation between the simulated values increases from 0.26 for model A to 0.34 for model B. Here we conclude that the DIF-adjusted self-assessments based on vignette c3 are closer to the objective measure than the unadjusted self-assessments.

We also report results when using all three vignettes simultaneously (see Table 2.16). These results are similar to those based on vignette c2.

Finally, consider the results based on data from wave 2. Since the vignette collected in the second wave is chosen out of the three from the first wave, we can study whether the results are consistent over time. Vignette c1 is the vignette collected in both waves. If the results are consistent over time we expect to conclude that for wave 2 the DIF-adjusted self-assessments, adjusted using vignette c1, are more different from the objective situation than the unadjusted self-assessments. The first set of results provided in Table 2.17 are for cognition using data from wave 2 (see Table 2.20 for parameter estimates of models A and B for delayed recall). Here the results are consistent across the four objective measures and lead to the same conclusion as before: DIF-adjusted self-assessments based on vignette c1 are more different from the objective situation than the unadjusted self-assessments. So, the results for vignette c1 are found to be consistent over time.

To summarize, our results reveal that for cognition the vignette method is sensitive to the choice of the vignette.

Table 2.19 gives a selection of parameter estimates of models A and B, estimated using data from the first wave. The differences in the correlations are for an important part caused by the country dummies and gender dummy, as for these variables either of the two following cases occurs relatively often: (1) one of the two parameter estimates is significantly different from zero, while the other is not; (2) both are significantly different from zero, but with opposite signs.

Breathing First, we discuss the correlations between different DIF-adjusted self-assessments using data from wave 1. These correlations are reported in Table 2.15, and they show that the vignettes b2 and b3 lead to similar DIF-adjusted self-assessments. However, they are very different from the one based on vignette b1. Thus, we conclude that for breathing the DIF-adjustments are sensitive to the choice of the vignette.

Second, we investigate whether the DIF-adjusted self-assessments are closer to the objective variable than the unadjusted self-assessments. We do this using data from wave 2, for which an objective measure is available and vignette b1 is collected. For wave 1 there is no objective measure available. The results of the models A and B are reported in Table 2.17.

The correlation coefficient between the predicted systematic parts of (y^*_{si}, y^*_{oi}) equals 0.45 for model A and 0.53 for model B. The reason for the low correlations between the self-assessment and objective measure is that the parameter estimates of certain country and age dummies and the gender dummy show the same discrepancy as described earlier for cognition. The correlation coefficient between the simulated values of (y^*_{si}, y^*_{oi}) equals 0.25 for model A and 0.27 for model B. Although perhaps low, both correlation coefficients still increase when the self-assessments are adjusted for heterogeneity in reporting behavior. So, correcting for reporting behavior heterogeneity brings the self-assessments of wave 2 closer to the objective situation.

Table 2.21 reports parameter estimates of the CHOPIT model for wave 1 using a single vignette and using all three vignettes simultaneously. Parameter estimates of models A and B for wave 2 are given in Table 2.23.

Mobility Finally, consider the results for mobility. Table 2.15 reports the correlations between different DIF-adjusted self-assessments using one vignette at a time and data from wave 1; they are high and approximately the same. Therefore, if the method works for one of the vignettes it is likely that it will work for the other two vignettes as well.

The results of the models A and B based on data from wave 2, for which an objective measure is available and vignette m1 is collected, are given in Table 2.17. The correlation between the predicted systematic parts of (y^*_{si}, y^*_{oi}) increases from 0.53 (model A) to 0.61 (model B), and the correlation between the simulated values of (y^*_{si}, y^*_{oi}) increases from 0.18 (model A) to 0.20 (model B). So, DIF-adjusted self-assessments are closer to the objective situation than unadjusted self-assessments.

We report in Table 2.22 parameter estimates of the CHOPIT model for wave 1 using a single vignette, as well as using all three vignettes simultaneously. Parameter estimates of models A and B for wave 2 are given in Table 2.24.

2.5 Conclusions

This chapter takes two approaches to validate the parametric model for anchoring vignettes. First, we study whether different vignettes lead to similar DIF-adjusted self-assessments. Second, we study whether DIF-adjusted self-assessments are closer to a measure of the actual situation than unadjusted self-assessments. Here, we also look at the performance of a single vignette. We use SHARE data and focus on three different domains of health: cognition, breathing and mobility.

Our results show that the method is sensitive to the choice of the vignette for cognition: DIF-adjusted self-assessments based on vignette c1 are more different from the objective situation than unadjusted self-assessments; for vignette c2 we conclude that the vignette method does not bring the self-assessments closer to the objective situation; the conclusion for vignette c3 is that the self-assessments are brought closer to the objective situation. For vignette c1, which is collected in both waves of SHARE, conclusions are consistent over time. The conclusions for cognition are the same irrespective of the objective measure used, except verbal fluency for wave 1.

For the breathing vignette collected in wave 2, vignette b1, we find that DIF-adjusted self-assessments are closer to the measure for breathing than the unadjusted self-assessments. However, our results also show that there is no guarantee that it would work with one of the two other breathing vignettes collected in the first wave.

Results are most encouraging for mobility. Adjusting the self-assessments using the vignette collected in wave 2, vignette m1, brings them closer to the measure for mobility. Moreover, the vignette method is unlikely to be sensitive to the choice of the vignettes used in wave 1.

Although our results indicate that the vignette method is sensitive to the domain and choice of the vignette, this should not be taken as a reason to reject this method. Here are several ideas for future research.

First, for the cognition domain we found that different vignettes lead to different results. The vignette describing a hypothetical person with the most extreme cognitive problems (vignette c3) brings the DIF-adjusted self-assessments closer to the objective situation than the other two vignettes, describing milder problems. This suggests that the level of health of the vignette person matters, at least in this case. More research should be done to find out, not only, how the level of health of the vignette person matters, but also how to formulate vignettes in general.

Second, our results show that the vignette method is sensitive to the choice of the vignette, at least for the domains of cognition and breathing. The reason may be that the CHOPIT model is incorrectly specified, in particular the *response consistency* and *vignette equivalence* assumptions may not hold for all vignettes. More research should be done to find out whether or not these two assumptions are tenable.

Third, it would be worthwhile to develop a validation method of the nonparametric approach for anchoring vignettes. This approach has been introduced by King et al. (2004) and further developed by King and Wand (2007). It does not make any statistical assumptions, but does require the response consistency and vignette equivalence assumptions. The paper by King and Wand (2007) develops a method for evaluating and choosing anchoring vignettes, which uses entropy to measure the discriminatory power of a vignette. They recommend to use the set of vignettes that is most informative in terms of their nonparametric estimator. Although both their and our paper study individual vignettes the approaches differ. King and Wand (2007) study the amount of information in a single vignette, whereas we study whether the information of the vignette is correct.

Fourth, many other issues may be important in order to appropriately use the vignette method. For example, Buckley (2008) and Hopkins and King (2010) show several patterns of bias due to context effects. Specifically, they show that the order of the self-assessment question and the vignette questions is important.

2.A Tables

Table 2.1: Self-assessment and vignette evaluations for wave 1 and 2.

	Wave 1				Wave 2	
Cognition	Self-assessment	Vignette c1	Vignette c2	Vignette c3	Self-assessment	Vignette c1
None	44.09	22.20	5.25	2.03	41.12	26.28
Mild	35.16	48.65	27.15	8.89	38.69	53.85
Moderate	16.21	22.73	44.35	29.77	16.01	16.58
Severe	4.14	6.08	20.68	47.27	3.55	3.06
Extreme	0.39	0.35	2.58	12.04	0.62	0.23
Breathing	Self-assessment	Vignette b1	Vignette b2	Vignette b3	Self-assessment	Vignette b1
None	64.61	10.77	2.29	2.45	66.70	3.64
Mild	22.22	24.12	5.15	2.22	21.30	26.34
Moderate	9.60	38.02	19.74	8.57	8.74	40.70
Severe	3.05	24.10	52.20	44.25	2.78	26.98
Extreme	0.53	3.00	20.61	42.51	0.47	2.33
Mobility	Self-assessment	Vignette m1	Vignette m2	Vignette m3	Self-assessment	Vignette m1
None	58.37	9.62	2.31	1.55	59.06	7.21
Mild	22.18	34.75	11.83	5.89	25.61	39.68
Moderate	13.05	42.84	38.77	27.51	12.28	38.16
Severe	5.23	11.97	40.39	48.80	2.86	14.45
Extreme	1.17	0.82	6.69	16.24	0.19	0.50

The numbers in this table are percentages within each category. They are based on different samples for every domain. Each sample is selected on the relevant self-assessment and vignette(s), and on the covariates. With the exception of the breathing and mobility samples for the first wave, all samples are also selected on the relevant objective measure(s). In wave 1: for cognition $N = 4343$, for breathing $N = 4366$, and for mobility $N = 4377$. In wave 2: for cognition $N = 6895$, for breathing $N = 6393$, and for mobility $N = 4788$.

Table 2.2: Self-assessment and vignette evaluations for cognition for wave 1.

	B	F	DE	GR	IT	NL	ES	SE
Self-assessment								
None	33.88	39.15	44.33	52.59	42.12	42.69	44.13	56.60
Mild	45.17	35.91	36.08	31.47	35.06	47.95	23.91	21.83
Moderate	19.31	21.45	16.08	13.29	15.76	6.82	22.17	12.44
Severe	1.46	3.24	3.51	2.66	5.41	1.95	9.57	8.38
Extreme	0.18	0.25	0.00	0.00	1.65	0.58	0.22	0.76
Vignette c1								
None	17.85	16.08	23.30	41.40	27.53	21.83	16.74	5.58
Mild	64.30	53.37	49.48	39.44	43.76	68.81	38.26	24.11
Moderate	15.30	26.06	24.33	15.80	20.24	8.19	33.26	46.19
Severe	2.37	3.87	2.27	3.36	7.76	1.17	11.52	23.60
Extreme	0.18	0.62	0.62	0.00	0.71	0.00	0.22	0.51
Vignette c2								
None	2.19	4.74	8.25	11.47	6.35	1.36	4.35	0.51
Mild	26.59	31.17	33.40	34.41	31.53	17.35	27.83	6.09
Moderate	51.55	51.62	44.74	37.62	42.12	50.88	48.04	20.81
Severe	18.76	11.60	13.20	15.94	18.59	25.15	19.13	57.87
Extreme	0.91	0.87	0.41	0.56	1.41	5.26	0.65	14.72
Vignette c3								
None	1.28	2.37	2.89	3.08	4.00	1.17	0.43	0.25
Mild	9.84	9.23	8.66	13.57	15.29	5.26	4.35	1.78
Moderate	33.15	39.28	27.01	27.83	32.47	31.77	28.26	8.88
Severe	45.90	44.39	50.72	44.06	40.24	39.38	60.87	58.63
Extreme	9.84	4.74	10.72	11.47	8.00	22.42	6.09	30.46

The numbers in this table are percentages within each category and are based on a sample that is selected on the relevant self-assessment and vignettes, the four objective measures, and on the covariates ($N = 4343$). Country abbreviations: Belgium (B), France (F), Greece (GR), Germany (DE), Italy (IT), the Netherlands (NL), Spain (ES), Sweden (SE).

Table 2.3: Self-assessment and vignette evaluations for breathing for wave 1.

	B	F	DE	GR	IT	NL	ES	SE
Self-assessment								
None	63.93	60.86	65.32	67.56	73.82	70.29	74.34	38.85
Mild	25.14	22.09	20.16	24.58	16.04	23.50	15.57	29.32
Moderate	9.11	13.50	10.28	5.90	6.60	3.88	7.02	21.55
Severe	1.46	3.44	3.43	1.69	3.07	1.75	3.07	8.02
Extreme	0.36	0.12	0.81	0.28	0.47	0.58	0.00	2.26
Vignette b1								
None	18.94	37.06	2.42	1.26	5.66	2.14	1.10	0.75
Mild	34.43	32.52	14.52	24.58	28.54	26.60	7.02	15.54
Moderate	33.70	24.66	49.60	43.96	38.21	43.11	34.21	43.86
Severe	10.93	5.40	31.85	26.54	25.00	23.11	50.66	36.34
Extreme	2.00	0.37	1.61	3.65	2.59	5.05	7.02	3.51
Vignette b2								
None	1.82	2.94	3.83	0.70	5.19	2.72	0.66	0.75
Mild	4.92	2.21	7.66	5.34	8.96	4.27	3.51	7.02
Moderate	23.68	18.53	23.99	18.54	21.46	20.97	15.79	14.79
Severe	51.37	66.01	55.24	46.07	45.05	40.97	56.58	49.37
Extreme	18.21	10.31	9.27	29.35	19.34	31.07	23.46	28.07
Vignette b3								
None	2.00	3.31	3.63	0.56	5.66	3.30	0.66	0.75
Mild	1.64	1.72	3.43	1.26	4.25	1.75	1.32	3.76
Moderate	5.46	5.28	8.06	10.81	11.56	6.60	16.01	7.02
Severe	47.91	61.60	45.36	36.94	38.92	21.36	44.96	49.87
Extreme	42.99	28.10	39.52	50.42	39.62	66.99	37.06	38.60

The numbers in this table are percentages within each category and are based on a sample that is selected on the relevant self-assessment and vignettes, and on the covariates ($N = 4366$). Country abbreviations: Belgium (B), France (F), Greece (GR), Germany (DE), Italy (IT), the Netherlands (NL), Spain (ES), Sweden (SE).

Table 2.4: Self-assessment and vignette evaluations for mobility for wave 1.

	B	F	DE	GR	IT	NL	ES	SE
Self-assessment								
None	55.35	66.91	46.26	74.30	58.55	57.93	52.49	38.36
Mild	26.32	15.13	27.07	15.36	20.37	24.86	19.96	38.36
Moderate	12.34	13.78	18.99	5.31	11.01	11.47	17.79	17.90
Severe	4.54	3.69	7.27	3.63	7.26	4.40	8.68	4.60
Extreme	1.45	0.49	0.40	1.40	2.81	1.34	1.08	0.77
Vignette m1								
None	11.43	9.23	5.86	9.64	21.78	4.02	3.04	14.58
Mild	43.92	32.60	26.67	33.66	36.53	43.21	21.48	40.92
Moderate	36.84	47.60	49.70	44.69	30.91	38.62	54.45	34.27
Severe	7.44	9.84	16.97	11.59	9.84	11.85	20.39	9.72
Extreme	0.36	0.74	0.81	0.42	0.94	2.29	0.65	0.51
Vignette m2								
None	2.18	2.71	3.43	1.26	4.68	1.91	1.30	1.28
Mild	13.43	7.75	11.52	17.04	11.24	9.94	8.68	15.86
Moderate	41.20	39.85	35.96	38.97	30.21	33.65	44.25	46.04
Severe	35.93	45.88	43.84	36.45	43.79	39.77	40.56	35.04
Extreme	7.26	3.81	5.25	6.28	10.07	14.72	5.21	1.79
Vignette m3								
None	1.81	2.34	1.21	0.56	4.22	1.53	0.65	0.00
Mild	3.81	8.24	7.07	5.03	12.18	2.49	5.21	2.56
Moderate	35.75	37.02	26.26	22.91	20.37	29.06	24.95	14.83
Severe	43.56	47.72	56.77	41.06	51.52	39.39	59.87	59.08
Extreme	15.06	4.67	8.69	30.45	11.71	27.53	9.33	23.53

The numbers in this table are percentages within each category and are based on a sample that is selected on the relevant self-assessment and vignettes, and on the covariates ($N = 4377$). Country abbreviations: Belgium (B), France (F), Greece (GR), Germany (DE), Italy (IT), the Netherlands (NL), Spain (ES), Sweden (SE).

Table 2.5: Self-assessments and vignette evaluations for cognition, breathing and mobility for wave 2.

	Self-assessment										Vignette evaluation									
	B	CZ	DK	F	DE	IT	NL	PO	ES	SE	B	CZ	DK	F	DE	IT	NL	PO	ES	SE
Cognition																				
None	30.88	39.21	54.66	35.41	43.38	37.89	39.88	37.38	43.34	41.68	25.73	29.83	32.49	21.81	26.76	28.78	32.26	12.52	15.51	27.21
Mild	48.54	43.05	30.36	42.49	38.93	38.91	50.10	28.55	28.63	36.72	61.87	57.74	53.95	55.81	53.33	47.58	62.12	41.99	50.70	48.81
Moderate	17.54	14.24	13.06	20.11	14.40	15.86	7.62	24.86	19.88	18.36	10.18	11.19	12.85	19.26	16.80	19.68	4.81	35.73	25.05	20.52
Severe	2.69	3.05	1.92	1.70	2.93	6.31	1.80	7.00	6.56	3.02	2.11	1.02	0.71	3.12	2.67	3.82	0.80	8.84	8.75	3.02
Extreme	0.35	0.45	0.00	0.28	0.36	1.03	0.60	2.21	1.59	0.22	0.12	0.23	0.00	0.00	0.44	0.15	0.00	0.92	0.00	0.43
Breathing																				
None	60.51	56.39	78.87	56.33	66.83	76.81	66.80	64.09	71.75	63.47	6.23	1.33	3.74	3.00	3.23	6.19	2.68	4.44	0.91	4.01
Mild	27.38	27.35	14.09	25.00	20.06	16.28	27.63	16.02	17.31	24.05	5.33	28.31	26.57	23.33	28.80	32.92	25.77	11.58	12.53	24.94
Moderate	9.41	13.01	5.55	16.00	9.03	4.96	3.71	11.58	7.74	8.69	1.93	50.12	41.62	50.33	39.92	38.58	53.20	30.89	27.33	28.06
Severe	2.08	3.13	1.39	2.33	3.80	1.59	1.24	6.37	2.73	3.34	5.53	18.67	25.93	23.00	26.62	20.88	16.29	50.39	53.53	35.19
Extreme	0.61	0.12	0.11	0.33	0.29	0.35	0.62	1.93	0.46	0.45	0.98	1.57	2.13	0.33	1.43	1.42	2.06	2.70	5.69	7.80
Mobility																				
None	59.20	33.28	76.96	82.71	51.00	67.19	55.83	55.27	64.16	63.17	9.29	4.18	8.23	8.41	5.72	12.24	5.10	8.55	6.14	5.71
Mild	27.20	42.44	14.94	8.88	31.59	21.09	33.50	19.37	19.11	20.32	55.56	48.87	37.97	39.25	29.48	46.61	43.20	27.64	26.28	34.60
Moderate	11.77	19.77	6.84	8.41	14.30	10.16	7.52	18.52	12.63	11.11	28.86	39.87	39.75	40.65	43.28	33.07	39.08	39.60	45.39	30.48
Severe	1.66	4.18	1.27	0.00	3.11	1.56	2.67	6.27	3.75	5.08	6.30	7.07	13.92	11.68	20.27	7.81	11.89	23.65	21.84	27.30
Extreme	0.17	0.32	0.00	0.00	0.00	0.00	0.49	0.57	0.34	0.32	0.00	0.00	0.13	0.00	1.24	0.26	0.73	0.57	0.34	1.90

The numbers in this table are percentages within each category and are based on different samples for every domain. Each sample is selected on the relevant self-assessment and vignette, the objective measure(s), and on the covariates. For cognition $N = 6895$, for breathing $N = 6393$, and for mobility $N = 4788$. Country abbreviations: Belgium (B), Czech Republic (CZ), Denmark (DK), France (F), Germany (DE), Italy (IT), the Netherlands (NL), Poland (PO), Spain (ES), Sweden (SE).

Table 2.6: Percentage of missing observations for the objective measures.

	Wave 1	Wave 2
Cognition		
Immediate recall	0.79	0.40
Delayed recall	0.75	0.40
Numeracy	0.33	0.22
Verbal fluency	1.01	0.51
Breathing		
Peak-flow test	-	7.28
Mobility		
Stand-up test	-	18.43

The numbers in this table are based on the vignette samples of both waves, except for mobility. $N = 4544$ for wave 1, and $N = 7186$ for wave 2. Respondents are only asked to participate in the stand-up test when they are younger than 75. In wave 2, approximately 82 percent of the sample is younger than 75, or $N = 5908$. This is the basis for the percentage of missing observations for the stand-up test. For the objective measures of breathing and mobility there are differences between countries in the percentage of missing observations. For breathing, on average the percentage of missing observations is 7 percent, but in France and Italy it is 15 percent. For mobility, on average the percentage of missing observations is 18 percent, but in France it is 28 percent and in Italy it is 36 percent.

Table 2.7: Descriptive statistics for the objective measures of cognition.

	Wave 1		Wave 2	
Nr. of words	Immediate recall	Delayed recall	Immediate recall	Delayed recall
0	1.24	9.49	0.84	8.18
1	3.13	9.19	1.75	7.56
2	5.53	13.82	4.23	12.50
3	11.28	18.56	10.11	18.97
4	18.28	19.78	17.96	20.07
5	23.39	15.15	23.16	15.61
6	19.32	8.13	21.13	9.47
7	12.25	3.75	12.89	4.83
8	4.21	1.45	5.82	1.81
9	0.94	0.48	1.64	0.70
10	0.41	0.21	0.46	0.30
Median	5	3	5	4
Mean	4.86	3.40	5.11	3.62
Std.dev	1.8	1.99	1.75	2.02
Score	Numeracy		Numeracy	
1	6.65		5.24	
2	16.14		14.97	
3	31.78		31.53	
4	29.50		29.63	
5	15.93		18.64	
Median	3		3	
Mean	3.32		3.41	
Std.dev	1.12		1.11	
	Verbal fluency		Verbal fluency	
Median	18		19	
Mean	18.58		19.77	
Std.dev	7.21		7.55	

The numbers in this table are percentages within each category, or the median, mean or standard deviation. They are based on a different samples for every wave. Each sample is selected on the relevant self-assessment and vignette(s), the objective measures, and on the covariates. $N = 4343$ in wave 1, and $N = 6895$ in wave 2.

Table 2.8: Descriptive statistics for delayed recall for wave 2.

Nr. of words	B	CZ	DK	F	DE	IT	NL	PO	ES	SE	Total
0	8.54	12.77	5.06	5.38	4.53	11.45	4.61	16.21	10.74	3.24	8.18
1	9.12	7.57	3.85	9.92	4.80	11.01	3.81	11.60	15.11	3.46	7.56
2	10.64	13.67	7.69	13.88	11.20	16.89	11.02	17.50	17.89	9.50	12.50
3	20.12	20.56	15.79	20.96	19.64	18.50	13.63	21.36	21.67	18.14	18.97
4	21.99	21.58	19.64	18.70	20.09	19.38	19.04	19.34	18.89	19.87	20.07
5	14.27	13.79	20.95	16.43	20.53	9.99	15.83	8.29	8.55	21.81	15.61
6	8.89	6.33	13.97	9.07	10.67	6.46	16.23	3.68	4.77	13.39	9.47
7	4.44	2.60	7.79	3.68	5.87	3.38	9.02	1.47	2.19	6.26	4.83
8	1.29	0.79	3.14	1.70	1.87	1.47	4.01	0.55	0.20	3.24	1.81
9	0.70	0.11	1.82	0.28	0.53	0.44	2.00	0.00	0.00	0.65	0.70
10	0.00	0.23	0.30	0.00	0.27	1.03	0.80	0.00	0.00	0.43	0.30
Mean	3.51	3.17	4.32	3.53	3.96	3.16	4.40	2.66	2.79	4.25	3.62
Std.dev	1.96	1.91	2.01	1.87	1.86	2.08	2.13	1.79	1.75	1.86	2.02

The numbers in the first eleven rows are percentages withing each category. They are based on a sample that is selected on the relevant self-assessment and vignette, the four objective measures, and on the covariates ($N = 6895$). Country abbreviations: Belgium (B), Czech Republic (CZ), Denmark (DK), France (F), Germany (DE), Italy (IT), the Netherlands (NL), Poland (PO), Spain (ES), Sweden (SE).

Table 2.9: Descriptive statistics for the objective measures of breathing and mobility.

	Peak flow test	Stand-up test
Median	350	10
Mean	356.71	11.10
Std.dev	158.33	5.33

The unit of measurement for the peak flow test is litre/minute and it ranges from 60 to 880. The unit of measurement for the stand-up test is time in seconds. The numbers in this table are based on different samples for every domain. Each sample is selected on the relevant self-assessment and vignette, the objective measure, and on the covariates. For breathing $N = 6393$ and for mobility $N = 4788$.

Table 2.10: Descriptive statistics for the peak flow test (breathing) and the stand-up test (mobility) for wave 2.

	Peak flow test			Stand-up test		
Country	Median	Mean	Std.dev	Median	Mean	Std.dev
Belgium	330	345.66	149.77	10.78	11.76	5.38
Czech Rep.	320	326.21	132.10	10.09	10.95	4.29
Denmark	390	394.32	145.88	9.19	9.69	3.26
France	350	358.40	174.67	10.00	10.89	6.06
Germany	350	361.42	149.19	9.50	10.98	6.09
Italy	280	294.63	145.09	10.65	12.59	7.44
Netherlands	390	403.24	149.06	10.73	11.75	5.28
Poland	305	325.34	153.01	10.01	11.17	4.60
Spain	270	328.53	225.51	11.00	12.62	6.52
Sweden	420	434.09	141.69	9.44	9.87	3.65
Total	350	356.71	158.33	10	11.10	5.33

The unit of measurement for the peak flow test is litre/minute and it ranges from 60 to 880. The unit of measurement for the stand-up test is time in seconds. We only observe the exact time for those respondents who are able to complete the test within one minute. The numbers in this table are based on different samples for every domain. Each sample has been selected on the relevant self-assessment and vignette, the objective measure, and on the covariates. For breathing $N = 6393$, and for mobility $N = 4788$.

Table 2.11: Descriptive statistics for the vignettes samples for both waves.

	Wave 1	Wave 2
Male (%)	44.43	44.60
Education mid (%)	44.60	59.18
Education high (%)	19.86	23.20
Not alone (%)	74.42	74.43
Long-term illness (%)	46.49	49.66
Phys. act. sometimes (%)	25.08	23.53
Phys. act. often (%)	34.46	33.10
Mean age	63.06	64.56
Std.dev age	10.01	9.86
N	4544	7186

The numbers in this table are based on the vignette samples of both waves.

Table 2.12: Descriptive statistics of the covariates for the different samples for both waves.

	Wave 1			
	Cognition	Breathing	Mobility	Vignette sample
Male (%)	44.83	44.69	44.71	44.43
Education mid (%)	45.08	45.03	44.92	44.60
Education high (%)	20.22	20.22	20.20	19.86
Not alone (%)	74.76	74.94	74.80	74.42
Long-term illness (%)	45.84	46.11	46.13	46.49
Phys. act. sometimes (%)	25.33	25.24	25.22	25.08
Phys. act. often (%)	34.58	34.63	34.48	34.46
Mean age	62.92	62.91	62.94	63.06
Std.dev age	9.95	9.94	9.95	10.01
N	4343	4366	4377	4544

	Wave 2			
	Cognition	Breathing	Mobility	Vignette sample
Male (%)	44.71	45.13	45.76	44.60
Education mid (%)	59.29	59.61	59.54	59.18
Education high (%)	23.36	24.25	27.46	23.20
Not alone (%)	74.95	75.44	79.45	74.43
Long-term illness (%)	49.31	48.22	44.13	49.66
Phys. act. sometimes (%)	23.61	24.15	25.77	23.53
Phys. act. often (%)	33.50	34.24	39.81	33.10
Mean age	64.37	64.09	61.11	64.56
Std.dev age	9.74	9.57	7.16	9.86
N	6895	6393	4788	7186

The numbers in this table are based on different samples for every domain. Each sample, except the vignette sample, is selected on the relevant self-assessment, vignette(s) and objective measure(s), and on the covariates.

Table 2.13: Descriptive statistics of the covariates for wave 1 per country.

Country	N	Male (%)	Education mid (%)	Education high (%)	Not alone (%)	Long-term illness (%)	Phys. act. sometimes (%)	Phys. act. often (%)	Mean age	Std.dev age
Belgium	567	43.74	50.71	25.18	73.90	47.80	24.16	27.34	62.57	9.73
France	885	42.71	36.27	19.45	67.99	48.75	20.98	31.97	63.91	10.39
Germany	508	43.50	74.70	24.51	77.56	57.09	27.36	41.73	63.44	9.15
Greece	720	45.83	37.19	20.06	73.75	30.28	44.03	34.17	60.83	10.62
Italy	445	43.82	36.40	6.52	73.42	45.62	21.35	26.74	63.53	9.35
Netherlands	538	47.58	64.10	21.62	84.20	38.85	16.73	49.81	62.25	9.39
Spain	464	41.81	26.78	10.15	71.98	55.17	12.93	26.51	64.64	10.54
Sweden	417	47.24	33.57	30.43	77.22	56.12	27.82	38.37	64.15	9.64
Total	4544	44.43	44.60	19.86	74.42	46.49	25.08	34.46	63.06	10.01

The numbers in this table are based on the vignette sample for wave 1.

Table 2.14: Descriptive statistics of the covariates for wave 2 per country.

Country	N	Male (%)	Education mid (%)	Education high (%)	Not alone (%)	Long-term illness (%)	Phys. act. sometimes (%)	Phys. act. often (%)	Mean age	Std.dev age
Belgium	896	45.65	51.74	22.22	72.63	43.53	21.57	27.42	65.41	10.04
CR	923	40.20	88.86	11.03	62.95	53.30	28.01	25.16	64.61	9.99
Denmark	1029	44.12	60.98	38.93	74.73	45.97	19.12	37.84	64.34	9.94
France	388	44.07	37.43	23.56	64.34	50.52	26.87	26.36	65.42	10.36
Germany	1164	45.70	69.69	29.61	76.19	54.47	28.68	36.52	65.05	9.30
Italy	697	45.77	41.24	9.05	83.05	37.59	24.39	24.10	64.83	8.80
Netherlands	527	47.63	59.22	27.18	82.31	38.78	18.97	55.56	61.73	9.89
Poland	566	43.29	81.28	17.29	73.85	69.79	21.38	25.62	62.99	9.81
Spain	518	45.37	29.71	13.20	79.26	51.93	18.96	34.82	64.29	10.41
Sweden	478	45.61	36.21	33.05	77.57	52.93	23.99	41.61	66.36	10.10
Total	7186	44.60	59.18	23.20	74.43	49.66	23.53	33.10	64.56	9.86

The numbers in this table are based on the vignette sample for wave 2.

Table 2.15: Correlations between two DIF-adjusted self-assessments for wave 1.

Cognition		$\hat{y}^*_{c1}$	$\hat{y}^*_{c2}$	$\hat{y}^*_{c3}$
	$\hat{y}^*_{c1}$	1.00	0.76	0.75
	$\hat{y}^*_{c2}$		1.00	0.92
	$\hat{y}^*_{c3}$			1.00
Breathing		$\hat{y}^*_{b1}$	$\hat{y}^*_{b2}$	$\hat{y}^*_{b3}$
	$\hat{y}^*_{b1}$	1.00	0.56	0.44
	$\hat{y}^*_{b2}$		1.00	0.94
	$\hat{y}^*_{b3}$			1.00
Mobility		$\hat{y}^*_{m1}$	$\hat{y}^*_{m2}$	$\hat{y}^*_{m3}$
	$\hat{y}^*_{m1}$	1.00	0.96	0.89
	$\hat{y}^*_{m2}$		1.00	0.94
	$\hat{y}^*_{m3}$			1.00

The numbers are correlation coefficients between predicted values of two DIF-adjusted self-assessments, each computed by estimating the CHOPIT-model using one vignette at a time. The vignette used is denoted in the subscript. By predicted values we mean: $\hat{y}^*_{si} = \mathbf{x}'_i\hat{\beta}_s$. Estimations are done separately for each domain and the samples are selected on the relevant self-assessment and vignettes, and on the covariates. For cognition the sample is also selected on the four objective measures. For cognition $N = 4343$, for breathing $N = 4366$, and for mobility $N = 4377$.

Table 2.16: Summary of results for cognition for wave 1.

Model	No. of parameters	Loglikelihood	AIC	Corr($\hat{y}^*_{si}, \hat{y}^*_{oi}$)	Corr($\varepsilon_{si}, \varepsilon_{oi}$)	Corr(y^*_{si}, y^*_{oi})
Immediate recall						
Vignette c1						
A	53	-16,209.12	32,524.24	0.70	0.12	0.24
B	116	-15,689.02	31,610.04	0.45	0.12	0.19
Vignette c2						
A	53	-16,255.96	32,617.92	0.70	0.12	0.24
B	116	-15,770.65	31,773.30	0.68	0.13	0.28
Vignette c3						
A	53	-15,197.01	30,500.02	0.70	0.12	0.24
B	116	-14,894.04	30,020.08	0.78	0.13	0.32
All 3 vignettes						
A	55	-25,719.59	51,549.18	0.70	0.12	0.24
B	118	-24,578,40	49,392.80	0.64	0.13	0.24
Delayed recall						
Vignette c1						
A	52	-15,137.61	30,379.22	0.75	0.15	0.26
B	115	-14,618.11	29,466.22	0.52	0.15	0.22
Vignette c2						
A	52	-15,184.44	30,472.88	0.75	0.15	0.26
B	115	-14,698.00	29,626.00	0.74	0.15	0.30
Vignette c3						
A	52	-14,125.50	28,355.00	0.75	0.15	0.26
B	115	-13,821.86	27,873.72	0.83	0.16	0.34
All 3 vignettes						
A	54	-24,648.08	49,404.16	0.75	0.15	0.26
B	117	-23,506.10	47,246.20	0.71	0.15	0.27
Numeracy						
Vignette c1						
A	52	-15,129.56	30,363.12	0.75	0.08	0.21
B	115	-14,608.73	29,447.46	0.54	0.07	0.17
Vignette c2						
A	52	-15,176.40	30,456.80	0.75	0.08	0.21
B	115	-14,692.70	29,615.40	0.72	0.07	0.25
Vignette c3						
A	52	-14,117.46	28,338.92	0.75	0.08	0.21
B	115	-13,815.77	27,861.54	0.78	0.07	0.29
All 3 vignettes						
A	54	-24,640.04	49,388.08	0.75	0.08	0.21
B	117	-23,499.67	47,233.34	0.69	0.07	0.21
Verbal fluency						
Vignette c1						
A	53	-16,022.07	32,150.14	0.50	0.12	0.20
B	116	-15,502.38	31,236.76	0.62	0.11	0.22
Vignette c2						
A	53	-16,070.72	32,247.44	0.50	0.12	0.20
B	116	-15,587.28	31,406.56	0.73	0.12	0.30
Vignette c3						
A	53	-15,009.33	30,124.66	0.50	0.12	0.20
B	116	-14,709.81	29,651.62	0.74	0.12	0.32
All 3 vignettes						
A	55	-25,530.93	51,117.86	0.50	0.12	0.20
B	118	-24,393,78	49,023.56	0.73	0.11	0.26

Model A assumes that there is no heterogeneity in reporting behaviour and that response consistency does not hold (No DIF, No RC). Model B assumes that there is heterogeneity in reporting behaviour and that response consistency holds (DIF, RC). Corr($\hat{y}^*_{si}, \hat{y}^*_{oi}$) is the correlation coefficient between the predicted systematic values, Corr(y^*_{si}, y^*_{oi}) is the correlation coefficient between the simulated values. All estimates are based on a sample that is selected on the relevant self-assessment and vignettes, the four objective measures, and on the covariates. $N = 4343$.

Table 2.17: Summary of results for cognition, breathing and mobility for wave 2.

Domain	No. of parameters	Loglikelihood	AIC	Corr($\hat{y}^*_{si}, \hat{y}^*_{oi}$)	Corr($\varepsilon_{si}, \varepsilon_{oi}$)	Corr(y^*_{si}, y^*_{oi})
Cognition						
Immediate recall						
A	57	-25,010.81	50,135.62	0.78	0.21	0.32
B	126	-24,558.24	49,368.48	0.40	0.21	0.24
Delayed recall						
A	56	-23,223.06	46,558.12	0.79	0.21	0.32
B	125	-22,770.11	45,790.22	0.43	0.21	0.24
Numeracy						
A	56	-23,538.05	47,188.10	0.71	0.14	0.24
B	125	-23,082.89	46,415.78	0.40	0.13	0.18
Verbal fluency						
A	57	-24,880.49	49,874.98	0.73	0.11	0.24
B	126	-24,421.47	49,094.94	0.37	0.10	0.16
Breathing						
A	57	-20,646.37	41,406.74	0.45	0.18	0.25
B	129	-20,224.53	40,707.06	0.53	0.18	0.27
Mobility						
A	47	-24,815.41	49,724.82	0.53	0.14	0.18
B	104	-24,555.99	49,319.98	0.61	0.14	0.20

Model A assumes that there is no heterogeneity in reporting behaviour, and that response consistency does not hold (No DIF, No RC). Model B assumes that there is heterogeneity in reporting behaviour and that response consistency holds (DIF, RC). Corr($\hat{y}^*_{si}, \hat{y}^*_{oi}$) is the correlation coefficient between the predicted systematic values, Corr(y^*_{si}, y^*_{oi}) is the correlation coefficient between the simulated values. Estimates are based on different samples for every domain. Each sample is selected on the relevant self-assessment and vignette, the objective measure(s), and on the covariates. For cognition $N = 6895$, for breathing $N = 6393$, and for mobility $N = 4788$.

Table 2.18: Selection of parameter estimates of the CHOPIT models for cognition for wave 1.

	Vignette c1				Vignette c2				Vignette c3				All 3 vignettes			
	β_s		γ^1		β_s		γ^1		β_s		γ^1		β_s		γ^1	
Covariates	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values
Constant	0.00	-	-2.23	-18.03	0.00	-	-2.39	-22.79	0.00	-	-2.45	-20.23	0.00	-	-2.25	-22.84
Belgium	-0.23	-2.82	-0.28	-2.83	0.19	2.14	0.23	3.54	0.08	0.67	0.11	1.43	-0.04	-0.57	0.10	2.21
Germany	-0.04	-0.49	0.06	0.59	-0.15	-1.57	0.08	1.08	-0.12	-0.95	0.25	3.06	-0.08	-1.05	0.14	2.76
Greece	-0.26	-3.39	-0.09	-1.13	-0.05	-0.61	0.15	2.26	0.00	0.01	0.18	2.54	-0.15	-2.29	0.10	2.39
Italy	-0.11	-1.23	0.29	3.60	-0.01	-0.07	0.28	3.98	-0.27	-2.34	0.04	0.52	-0.12	-1.63	0.15	3.55
Netherlands	-0.15	-1.71	-0.22	-1.91	0.74	7.53	0.60	8.97	0.51	4.05	0.36	4.61	0.19	2.46	0.29	6.08
Spain	0.15	1.78	0.48	6.03	0.02	0.17	0.31	4.41	0.30	2.37	0.60	7.11	0.15	2.04	0.38	8.42
Sweden	0.81	8.76	1.06	12.73	0.94	8.43	1.51	18.70	0.78	5.33	1.36	12.00	0.88	10.85	1.19	21.72
Male	0.13	2.90	-0.01	-0.14	0.02	0.43	-0.02	-0.46	-0.01	-0.14	-0.08	-1.92	0.07	1.81	-0.02	-0.96
Age ¡ 50	-0.02	-0.17	-0.01	-0.05	-0.13	-1.18	-0.04	-0.53	-0.14	-0.91	-0.10	-1.06	-0.03	-0.32	-0.06	-1.20
Age 55 to 60	0.00	0.05	-0.20	-2.53	0.06	0.72	0.00	0.03	0.07	0.68	-0.02	-0.36	0.04	0.57	-0.03	-0.87
Age 60 to 65	-0.05	-0.72	-0.13	-1.64	-0.01	-0.11	-0.16	-2.59	-0.05	-0.47	-0.06	-0.84	-0.03	-0.46	-0.09	-2.17
Age 65 to 70	-0.11	-1.43	-0.07	-0.84	-0.13	-1.54	-0.03	-0.43	-0.16	-1.44	-0.05	-0.67	-0.11	-1.65	-0.02	-0.60
Age 70 to 75	-0.17	-1.96	-0.20	-2.35	-0.23	-2.47	-0.18	-2.76	-0.19	-1.60	-0.22	-2.78	-0.19	-2.53	-0.15	-3.32
Age 70 to 80	-0.28	-2.79	-0.12	-1.16	-0.18	-1.74	-0.02	-0.26	-0.36	-2.75	-0.12	-1.27	-0.25	-2.97	-0.06	-1.10
Age ¿ 80	-0.46	-3.95	-0.07	-0.77	-0.53	-4.43	-0.12	-1.39	-0.61	-4.08	-0.20	-1.90	-0.53	-5.46	-0.08	-1.46
Education low	-0.07	-1.31	0.25	4.28	-0.29	-4.77	0.04	1.07	-0.54	-6.96	-0.12	-2.34	-0.22	-4.57	0.02	0.65
Education high	0.13	2.09	0.11	1.64	0.19	2.64	0.01	0.28	0.34	3.57	0.26	4.47	0.16	2.78	0.11	3.18
Phys. act. sometimes	0.11	1.88	-0.12	-1.83	0.10	1.56	-0.07	-1.57	0.24	2.77	0.00	0.05	0.14	2.60	-0.03	-1.09
Phys. act. often	0.07	1.19	-0.07	-1.31	0.07	1.10	-0.07	-1.70	0.01	0.10	-0.12	-2.34	0.05	1.02	-0.06	-1.96
Alone	-0.12	-2.26	0.13	2.56	-0.20	-3.52	0.05	1.26	-0.28	-3.90	-0.03	-0.70	-0.15	-3.33	0.02	0.79
Illness long	-0.36	-7.68	0.05	1.06	-0.39	-7.39	0.05	1.41	-0.35	-5.23	0.14	3.09	-0.36	-8.68	0.06	2.50

The sample is selected on the relevant self-assessment and vignettes, the four objective measures, and on the covariates. $N = 4343$.

Table 2.19: Selection of parameter estimates for cognition for wave 1 with delayed recall as the objective measure.

	Model A				Model B, vignette c1				Model B, vignette c2				Model B, vignette c3				Model B, all 3 vignettes			
	β_o		β_s		β_s		γ^1		β_s		γ^1		β_s		γ^1		β_s		γ^1	
Covariates	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values
Constant	2.40	39.32	3.00	40.66	3.23	25.93	1.00	-	3.39	29.98	1.00	-	3.46	28.77	1.00	-	3.25	81.31	1.00	
Belgium	0.02	0.39	-0.08	-1.30	-0.23	-2.77	-0.29	-2.83	0.20	2.19	0.23	3.47	0.09	0.76	0.11	1.43	-0.04	-0.55	0.10	2.29
Germany	0.11	1.86	0.05	0.71	-0.05	-0.52	0.06	0.61	-0.16	-1.62	0.08	1.08	-0.12	-1.02	0.25	3.10	-0.09	-1.40	0.14	2.92
Greece	0.11	2.16	0.19	3.19	-0.25	-3.28	-0.10	-1.16	-0.04	-0.50	0.14	2.22	0.01	0.14	0.18	2.55	-0.15	-2.43	0.10	2.49
Italy	-0.35	-5.72	0.04	0.61	-0.11	-1.23	0.29	3.51	-0.01	-0.07	0.28	3.95	-0.27	-2.38	0.04	0.50	-0.12	-1.62	0.15	4.70
Netherlands	0.23	4.04	0.09	1.38	-0.15	-1.68	-0.22	-1.93	0.75	7.58	0.60	8.73	0.52	4.06	0.36	4.65	0.19	2.68	0.29	7.58
Spain	-0.40	-6.69	0.03	0.44	0.15	1.77	0.48	5.91	0.02	0.19	0.31	4.40	0.30	2.58	0.60	7.27	0.15	2.00	0.38	9.29
Sweden	0.35	5.68	0.20	2.83	0.80	8.64	1.06	12.53	0.94	8.36	1.52	18.61	0.78	5.43	1.37	12.21	0.88	9.95	1.19	39.39
Male	-0.20	-6.36	0.11	3.03	0.13	2.86	-0.00	-0.10	0.02	0.40	-0.02	-0.42	-0.01	-0.18	-0.08	-1.91	0.07	2.01	-0.02	-1.06
Age ¡ 50	0.09	1.31	-0.04	-0.48	-0.02	-0.18	-0.00	-0.03	-0.14	-1.20	-0.04	-0.50	-0.14	-1.00	-0.10	-1.08	-0.03	-0.36	-0.06	-1.21
Age 55 to 60	-0.12	-2.48	0.03	0.45	0.01	0.07	-0.20	-2.49	0.06	0.75	0.00	0.05	0.07	0.70	-0.02	-0.35	0.04	0.70	-0.03	-1.10
Age 60 to 65	-0.21	-4.09	-0.08	-1.32	-0.06	-0.72	-0.13	-1.59	-0.01	-0.14	-0.16	-2.53	-0.06	-0.52	-0.06	-0.81	-0.03	-0.55	-0.09	-2.37
Age 65 to 70	-0.38	-7.05	-0.19	-3.18	-0.11	-1.41	-0.07	-0.82	-0.13	-1.54	-0.03	-0.42	-0.16	-1.50	-0.05	-0.67	-0.11	-1.72	-0.02	-0.62
Age 70 to 75	-0.45	-7.65	-0.24	-3.67	-0.17	-1.93	-0.19	-2.33	-0.23	-2.45	-0.18	-2.73	-0.19	-1.61	-0.23	-2.80	-0.19	-2.62	-0.15	-3.31
Age 75 to 80	-0.63	-8.82	-0.32	-4.14	-0.28	-2.76	-0.12	-1.17	-0.18	-1.72	-0.02	-0.27	-0.36	-2.68	-0.12	-1.26	-0.25	-2.94	-0.06	-1.11
Age ¿ 80	-1.04	-12.19	-0.51	-5.89	-0.46	-3.91	-0.08	-0.84	-0.53	-4.40	-0.13	-1.47	-0.60	-4.14	-0.20	-1.91	-0.52	-5.52	-0.08	-1.34
Education low	-0.41	-10.59	-0.18	-4.27	-0.08	-1.32	0.25	4.23	-0.30	-4.77	0.04	0.86	-0.54	-7.07	-0.12	-2.38	-0.23	-4.51	0.02	0.69
Education high	0.29	6.94	0.14	2.94	0.13	2.06	0.11	1.65	0.19	2.64	0.01	0.27	0.34	3.63	0.26	4.48	0.16	2.97	0.11	3.88
Phys. act. sometimes	0.12	2.96	0.14	3.05	0.11	1.85	-0.12	-1.84	0.10	1.54	-0.07	-1.56	0.24	2.82	0.00	0.04	0.14	3.09	-0.03	-1.23
Phys. act. often	0.05	1.35	0.09	2.21	0.07	1.16	-0.07	-1.30	0.07	1.05	-0.07	-1.68	0.00	0.04	-0.12	-2.37	0.05	1.06	-0.06	-2.14
Alone	-0.12	-3.13	-0.14	-3.39	-0.12	-2.23	0.13	2.58	-0.20	-3.51	0.05	1.22	-0.28	-3.94	-0.03	-0.68	-0.15	-3.71	0.02	0.91
Illness long	-0.11	-3.28	-0.38	-10.64	-0.36	-7.58	0.05	1.05	-0.39	-7.32	0.05	1.26	-0.35	-5.30	0.14	3.13	-0.36	-9.09	0.06	2.85

Model A assumes that there is no heterogeneity in reporting behaviour and that response consistency does not hold (No DIF, No RC). Model B assumes that there is heterogeneity in reporting behaviour and that response consistency holds (DIF, RC). All estimates are based on a sample that is selected on the relevant self-assessment and vignettes, the four objective measures, and on the covariates. $N = 4343$.

Table 2.20: Parameter estimates for cognition for wave 2 with delayed recall as the objective measure and vignette c1.

	Model A				Model B							
	β_o		β_s		β_s		γ^1		γ^2		γ^3	
Covariates	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values
Constant	2.44	49.35	3.19	52.25	3.16	31.56	1.00	-	-0.16	-1.94	0.20	4.62
Belgium	-0.11	-2.33	-0.21	-4.10	-0.37	-5.58	-0.34	-4.10	0.05	0.64	0.22	5.74
Czech Rep.	-0.34	-7.51	-0.03	-0.66	-0.21	-3.17	-0.32	-3.72	0.05	0.57	0.13	3.57
Denmark	0.14	3.13	0.20	3.90	0.03	0.52	-0.38	-4.06	0.22	2.80	-0.02	-0.47
France	-0.08	-1.35	-0.13	-1.84	-0.09	-0.92	-0.25	-2.25	0.24	2.52	0.09	1.73
Italy	-0.19	-3.73	-0.19	-3.24	-0.27	-3.45	0.03	0.37	-0.07	-0.90	-0.06	-1.39
Netherlands	0.18	3.38	-0.07	-1.18	-0.37	-4.68	-0.42	-3.41	-0.21	-1.59	0.31	7.28
Poland	-0.66	-12.47	-0.29	-4.97	0.20	2.61	0.55	7.11	0.09	1.30	-0.22	-4.48
Spain	-0.40	-6.92	-0.12	-1.87	0.14	1.66	0.37	4.13	-0.06	-0.72	-0.10	-1.89
Sweden	0.29	5.21	-0.03	-0.53	-0.04	-0.53	-0.06	-0.59	0.13	1.50	-0.10	-1.97
Male	-0.26	-10.59	0.05	1.84	0.15	4.09	0.05	1.17	0.08	2.08	-0.04	-1.78
Age ¡ 50	0.23	2.91	0.02	0.27	-0.01	-0.06	0.22	1.73	-0.19	-1.30	-0.13	-1.90
Age 50 to 55	0.12	2.89	0.00	0.07	-0.06	-1.03	-0.09	-1.17	0.01	0.09	0.03	0.78
Age 60 to 65	-0.08	-1.96	-0.04	-0.92	0.03	0.48	-0.02	-0.32	0.07	1.06	0.04	1.08
Age 65 to 70	-0.24	-5.75	-0.21	-4.54	-0.15	-2.45	-0.06	-0.85	0.11	1.71	0.05	1.47
Age 70 to 75	-0.40	-9.06	-0.30	-6.00	-0.19	-2.90	0.12	1.63	-0.04	-0.59	0.05	1.52
Age 75 to 80	-0.73	-13.85	-0.46	-7.96	-0.38	-5.03	0.04	0.54	0.08	1.02	-0.04	-0.83
Age ¿ 80	-0.90	-15.72	-0.69	-11.26	-0.56	-7.21	0.07	0.75	0.06	0.80	0.06	1.35
Education low	-0.34	-8.61	-0.08	-1.87	0.03	0.56	0.19	3.08	-0.05	-0.86	-0.06	-1.70
Education high	0.34	11.12	0.14	3.92	0.02	0.40	-0.13	-2.03	-0.03	-0.52	0.04	1.47
Phys. act. sometimes	0.14	4.51	0.13	3.74	0.16	3.58	-0.09	-1.65	0.08	1.64	0.06	2.42
Phys. act. often	0.15	5.07	0.14	4.30	0.15	3.36	-0.13	-2.52	0.08	1.71	0.07	2.77
Alone	-0.09	-3.18	-0.03	-0.93	-0.04	-1.06	0.05	0.97	-0.03	-0.74	-0.03	-1.43
Illness long	-0.08	-3.21	-0.40	-14.07	-0.44	-11.86	0.01	0.29	0.02	0.44	-0.06	-2.70

	Model A		Model B	
Threshold parameters	Coeff	T-values	Coeff	T-values
γ_s^1	1.00	-		
γ_s^2	-0.01	-0.49		
γ_s^3	0.13	7.91		
γ_v^1	1.00	-		
γ_v^2	-0.01	-0.24		
γ_v^3	0.39	29.29		
γ_o^1	0.00	-	0.00	-
γ_o^2	0.00	-	0.00	-
γ_o^3	0.07	2.56	0.07	2.57
Variances				
σ_s^2	1.00	-	1.00	-
σ_u^2	0.00	-	0.38	19.16
σ_v^2	1.00	-	0.75	43.30
σ_o^2	0.91	53.33	0.91	52.71
ρ	0.21	14.59	0.21	13.52
Vignette dummy	Coeff	Std.error	Coeff	Std.error
ϑ	2.84	0.03	2.60	0.02

Model A assumes that there is no heterogeneity in reporting behaviour and that response consistency does not hold (No DIF, No RC). Model B assumes that there is heterogeneity in reporting behaviour and that response consistency holds (DIF, RC). All estimates are based on a sample that is selected on the relevant self-assessment and vignette, the four objective measures, and on the covariates. $N = 6895$.

Table 2.21: Selection of parameter estimates of the CHOPIT model for breathing for wave 1.

	Vignette b1				Vignette b2				Vignette b3				All 3 vignettes			
	β_s		γ^1		β_s		γ^1		β_s		γ^1		β_s		γ^1	
Covariates	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values
Constant	0.00	-	0.52	6.75	0.00	-	0.45	5.89	0.00	-	0.46	6.06	0.00	-	0.60	7.24
Belgium	-0.51	-5.80	-0.42	-7.94	-0.07	-0.53	0.00	0.01	-0.62	-3.20	-0.55	-2.96	-0.35	-4.13	-0.31	-6.19
Germany	-1.15	-11.64	-1.10	-14.25	0.49	3.60	0.60	4.82	0.38	1.92	0.49	2.55	-0.58	-6.53	-0.56	-10.07
Greece	-1.22	-13.18	-1.25	-16.34	-0.22	-1.67	-0.21	-1.67	-0.60	-3.00	-0.61	-3.04	-0.98	-11.84	-1.08	-16.39
Italy	-1.11	-10.99	-0.79	-10.86	0.17	1.28	0.54	4.60	0.27	1.44	0.64	3.55	-0.70	-7.82	-0.42	-8.38
Netherlands	-1.29	-12.65	-1.16	-14.42	-0.08	-0.63	0.11	0.88	-0.20	-1.07	-0.01	-0.04	-0.85	-9.70	-0.75	-13.67
Spain	-1.75	-16.58	-1.32	-14.80	-0.56	-4.09	-0.13	-0.92	-0.50	-2.48	-0.09	-0.41	-1.39	-15.45	-1.05	-14.62
Sweden	-0.61	-6.35	-1.27	-14.39	0.62	4.54	-0.00	-0.01	0.37	1.78	-0.25	-1.20	-0.20	-2.27	-0.96	-13.27
Male	-0.09	-1.83	-0.09	-2.37	0.06	0.84	0.08	1.21	0.06	0.54	0.08	0.78	-0.01	-0.32	-0.01	-0.23
Age ¡ 50	0.04	0.31	0.03	0.38	-0.04	-0.21	-0.07	-0.36	-0.57	-1.96	-0.60	-2.02	0.02	0.17	0.01	0.18
Age 55 to 60	-0.08	-0.96	-0.08	-1.45	0.11	0.99	0.13	1.14	0.08	0.48	0.09	0.53	-0.01	-0.08	0.00	0.07
Age 60 to 65	0.02	0.29	-0.07	-1.14	0.17	1.41	0.11	0.94	0.24	1.36	0.16	0.92	0.10	1.28	0.03	0.58
Age 65 to 70	0.12	1.37	-0.09	-1.43	0.39	3.25	0.19	1.66	0.47	2.64	0.26	1.50	0.24	3.04	0.05	1.03
Age 70 to 75	0.08	0.86	-0.22	-2.99	0.33	2.58	0.08	0.60	0.38	1.93	0.10	0.53	0.20	2.46	-0.06	-1.11
Age 70 to 80	0.22	2.12	-0.10	-1.23	0.54	3.76	0.28	2.00	0.40	1.75	0.11	0.49	0.32	3.42	0.02	0.34
Age ¿ 80	-0.06	-0.52	-0.23	-2.56	0.32	1.94	0.21	1.42	0.65	2.71	0.51	2.14	0.18	1.57	0.05	0.76
Education low	0.01	0.10	-0.00	-0.11	0.29	3.42	0.31	3.96	0.43	3.41	0.45	3.66	0.10	1.77	0.10	2.97
Education high	-0.13	-1.86	-0.01	-0.22	-0.25	-2.55	-0.14	-1.45	-0.35	-2.14	-0.24	-1.46	-0.20	-3.23	-0.09	-1.90
Phys. act. sometimes	-0.22	-3.39	-0.03	-0.57	-0.25	-2.71	-0.06	-0.66	-0.24	-1.71	-0.06	-0.39	-0.21	-3.55	-0.02	-0.49
Phys. act. often	-0.24	-3.95	-0.01	-0.28	-0.33	-3.68	-0.09	-1.14	-0.18	-1.40	0.04	0.31	-0.22	-3.92	0.03	0.78
Alone	0.05	0.77	0.02	0.58	0.09	1.12	0.08	1.09	0.21	1.74	0.19	1.61	0.06	1.17	0.04	1.26
Illness long	0.50	9.48	0.08	1.98	0.51	6.88	0.09	1.25	0.54	5.02	0.14	1.31	0.46	9.69	0.03	0.91

The sample is selected on the relevant self-assessment and vignettes, and on the covariates. $N = 4366$.

Table 2.22: Selection of parameter estimates of the CHOPIT model for mobility for wave 1.

	Vignette m1				Vignette m2				Vignette m3				All 3 vignettes			
	β_s		γ^1		β_s		γ^1		β_s		γ^1		β_s		γ^1	
Covariates	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values
Constant	0.00	-	1.07	14.44	0.00	-	1.07	14.27	0.00	-	1.04	13.81	0.00	-	1.10	14.17
Belgium	0.43	5.36	0.04	0.67	0.40	4.12	-0.03	-0.33	0.09	0.87	-0.33	-3.65	0.38	5.12	-0.01	-0.34
Germany	0.32	3.80	-0.24	-3.97	0.65	6.41	0.08	0.95	0.30	2.69	-0.29	-2.86	0.44	5.64	-0.11	-2.41
Greece	-0.04	-0.56	-0.01	-0.14	0.03	0.35	0.01	0.08	-0.35	-3.38	-0.34	-3.53	-0.06	-0.83	-0.04	-1.00
Italy	0.52	5.81	0.30	5.39	0.38	3.71	0.12	1.50	0.35	3.30	0.11	1.24	0.47	5.79	0.26	6.12
Netherlands	0.24	2.73	-0.23	-3.46	0.35	3.48	-0.10	-1.21	0.04	0.41	-0.40	-4.16	0.24	3.03	-0.23	-4.90
Spain	0.02	0.24	-0.32	-4.66	0.25	2.48	-0.12	-1.37	-0.01	-0.06	-0.34	-3.38	0.09	1.15	-0.28	-5.36
Sweden	0.71	8.14	0.04	0.70	0.58	5.56	-0.23	-2.45	-0.28	-2.18	-1.06	-8.34	0.61	7.41	-0.05	-0.96
Male	-0.16	-3.76	-0.02	-1.26	-0.07	-1.42	0.11	2.35	-0.03	-0.60	0.15	2.86	-0.12	-2.97	0.03	1.37
Age ¡ 50	0.04	0.36	0.00	0.00	0.04	0.31	0.03	0.20	0.13	0.95	0.15	1.14	0.03	0.31	-0.00	-0.06
Age 55 to 60	0.02	0.33	-0.03	-0.56	0.09	1.09	0.07	0.99	0.06	0.67	0.04	0.44	0.03	0.47	-0.02	-0.58
Age 60 to 65	0.10	1.31	-0.03	-0.60	0.05	0.56	-0.08	-1.13	0.03	0.34	-0.10	-1.09	0.09	1.19	-0.04	-0.92
Age 65 to 70	0.34	4.50	0.02	0.35	0.38	4.41	0.10	1.45	0.43	4.37	0.18	1.99	0.36	5.12	0.05	1.08
Age70 to 75	0.39	4.82	0.00	0.06	0.37	3.95	-0.02	-0.26	0.29	2.64	-0.11	-1.06	0.37	4.86	-0.03	-0.62
Age 75 to 80	0.66	6.96	0.03	0.48	0.67	6.27	-0.01	-0.12	0.66	5.71	-0.01	-0.12	0.69	7.88	0.04	0.82
Age ¿ 80	0.80	7.40	0.00	0.05	0.94	8.06	0.18	1.84	0.90	6.80	0.14	1.16	0.87	9.17	0.07	1.20
Education low	0.10	1.73	-0.01	-0.30	0.26	4.05	0.17	3.08	0.18	2.60	0.07	1.17	0.16	3.16	0.07	2.32
Education high	-0.22	-3.51	-0.05	-1.24	-0.34	-4.47	-0.20	-2.87	-0.34	-4.03	-0.21	-2.51	-0.26	-4.41	-0.10	-2.70
Alone	0.09	1.72	-0.03	-0.82	0.20	3.50	0.09	1.86	0.20	3.08	0.08	1.43	0.12	2.57	-0.00	-0.10
Illness long	0.77	16.64	0.00	0.03	0.78	14.46	0.01	0.28	0.84	14.10	0.09	1.65	0.79	18.51	0.01	0.40

The sample is selected on the relevant self-assessment and vignettes, and on the covariates. $N = 4377$.

Table 2.23: Parameter estimates for breathing for wave 2 with vignette b1.

	Model A				Model B							
	β_o		β_s		β_s		γ^1		γ^2		γ^3	
Covariates	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values
Constant	-3.37	-12.71	2.59	6.21	2.52	4.41	1.00	-	-0.44	-1.11	-0.02	-0.06
Belgium	0.01	0.14	-0.08	-1.31	-0.16	-2.06	-0.28	-6.05	0.15	2.60	0.13	2.25
Czech Rep.	-0.13	-3.34	-0.17	-2.99	-0.03	-0.39	-0.16	-3.56	0.27	4.94	0.17	2.73
Denmark	0.16	4.57	0.36	5.95	0.38	4.71	-0.01	-0.36	0.06	1.11	-0.02	-0.38
France	0.11	2.09	-0.20	-2.42	-0.04	-0.38	-0.11	-1.80	0.34	4.59	0.01	0.15
Italy	-0.28	-6.27	0.34	4.58	0.29	3.01	-0.10	-1.91	0.02	0.30	0.02	0.24
Netherlands	0.11	2.41	-0.07	-0.96	0.14	1.40	-0.23	-4.27	0.34	5.27	0.23	3.24
Poland	-0.13	-3.10	-0.03	-0.43	0.14	1.64	0.45	9.34	-0.14	-1.89	-0.50	-5.48
Spain	-0.06	-1.15	0.23	2.83	0.70	6.23	0.58	10.18	-0.11	-1.28	-0.08	-0.80
Sweden	0.48	10.46	-0.05	-0.69	0.04	0.47	0.25	5.01	-0.29	-3.65	0.01	0.13
Male	0.62	22.57	-0.05	-1.24	0.02	0.30	0.16	4.88	-0.07	-1.58	-0.04	-0.92
Age ¡ 50	0.17	2.66	0.14	1.21	0.10	0.66	-0.04	-0.49	-0.06	-0.60	0.03	0.28
Age 50 to 55	0.12	3.61	0.04	0.67	0.03	0.47	0.04	1.14	-0.08	-1.58	0.01	0.26
Age 60 to 65	-0.08	-2.34	-0.03	-0.62	0.01	0.08	0.02	0.46	0.01	0.12	0.02	0.45
Age 65 to 70	-0.21	-5.95	-0.14	-2.13	-0.04	-0.47	0.02	0.50	0.06	1.24	0.03	0.53
Age 70 to 75	-0.42	-11.32	-0.19	-3.37	-0.11	-1.41	0.02	0.51	0.05	0.99	0.04	0.61
Age 75 to 80	-0.57	-13.06	-0.37	-5.56	-0.33	-3.78	-0.05	-0.91	0.14	2.25	-0.03	-0.38
Age ¿ 80	-0.70	-14.75	-0.39	-5.60	-0.32	-3.44	-0.02	-0.39	0.18	2.73	-0.07	-0.96
Education low	-0.07	-2.04	-0.06	-1.16	-0.15	-2.29	-0.01	-0.33	-0.07	-1.32	-0.04	-0.85
Education high	0.17	6.67	0.14	3.32	0.17	3.03	0.03	1.12	-0.01	-0.25	0.01	0.32
Phys. act. sometimes	0.10	3.95	0.23	5.55	0.22	3.94	-0.08	-2.43	0.03	0.83	0.06	1.39
Phys. act. often	0.19	7.79	0.35	8.82	0.37	6.90	0.00	0.11	-0.04	-0.99	0.06	1.57
Alone	-0.08	-3.16	-0.03	-0.75	-0.04	-0.84	-0.00	-0.05	-0.00	-0.01	-0.01	-0.25
Illness long	-0.07	-3.51	-0.61	-18.42	-0.59	-12.70	0.05	2.05	0.02	0.71	-0.03	-0.83
Height/100	1.88	11.80	0.38	1.53	0.68	2.08	0.24	1.35	0.07	0.28	-0.08	-0.31

	Model A		Model B	
Threshold parameters	Coeff	T-values	Coeff	T-values
γ_s^1	1.00	-		
γ_s^2	-0.28	-7.04		
γ_s^3	-0.18	-7.55		
γ_v^1	1.00	-		
γ_v^2	0.07	4.02		
γ_v^3	0.24	10.39		
Variances				
σ_s^2	1.00	-	1.00	-
σ_u^2	0.00	-	0.27	10.70
σ_v^2	1.00	-	0.63	29.70
σ_o^2	0.79	113.04	0.79	155.22
ρ	0.18	11.47	0.18	12.52
Vignette dummy	Coeff	Std.error	Coeff	Std.error
ϑ	1.54	0.02	1.90	0.30

Model A assumes that there is no heterogeneity in reporting behaviour and that response consistency does not hold (No DIF, No RC). Model B assumes that there is heterogeneity in reporting behaviour and that response consistency holds (DIF, RC). Since the objective measure is a continuous variable there are no thresholds in the objective part of the model. All estimates are based on a sample that is selected on the relevant self-assessment and vignette, the objective measure, and on the covariates. $N = 6393$.

Table 2.24: Parameter estimates for mobility for wave 2 with vignette m1.

	Model A				Model B							
	β_o		β_s		β_s		γ^1		γ^2		γ^3	
Covariates	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values	Coeff	T-values
Constant	10.69	41.06	0.61	10.09	0.59	9.08	1.00	-	-0.08	-1.30	-0.20	-3.01
Belgium	0.82	2.88	-0.13	-1.99	0.25	3.01	0.31	4.99	0.17	3.09	0.04	0.51
Czech Rep.	-0.13	-0.46	0.36	5.80	0.49	6.25	-0.06	-0.87	0.31	5.68	0.19	3.01
Denmark	-1.08	-4.18	-0.58	-9.13	-0.34	-4.33	0.26	4.46	-0.09	-1.50	0.02	0.28
France	0.01	0.03	-0.83	-7.53	-0.56	-4.18	0.30	3.44	-0.15	-1.51	0.12	1.30
Italy	1.49	4.37	-0.32	-3.91	0.07	0.69	0.38	5.32	-0.02	-0.28	0.02	0.25
Netherlands	1.03	3.27	-0.04	-0.52	0.06	0.67	0.03	0.36	0.21	3.32	-0.01	-0.11
Poland	0.17	0.50	-0.06	-0.81	0.07	0.71	0.26	3.70	-0.34	-4.26	-0.10	-1.37
Spain	1.70	4.55	-0.22	-2.45	-0.13	-1.21	0.15	1.79	-0.27	-2.84	0.01	0.09
Sweden	-1.20	-3.44	-0.22	-2.70	-0.14	-1.37	0.15	1.92	-0.10	-1.22	-0.36	-4.10
Male	-0.83	-5.47	-0.11	-3.14	-0.20	-4.54	-0.08	-2.54	0.00	0.08	-0.04	-1.16
Age ¡ 50	-1.33	-3.01	-0.19	-1.72	-0.06	-0.48	0.20	2.23	-0.17	-1.73	-0.22	-1.91
Age 50 to 55	-0.40	-1.81	-0.03	-0.58	-0.04	-0.61	-0.00	-0.03	-0.04	-0.81	-0.03	-0.48
Age 60 to 65	0.58	2.61	0.02	0.39	0.06	0.86	0.01	0.22	0.06	1.37	0.03	0.51
Age 65 to 70	0.64	2.68	0.08	1.38	0.11	1.54	0.01	0.17	0.02	0.51	0.04	0.66
Age ¿ 70	1.87	7.13	0.14	2.32	0.15	2.02	-0.05	-0.79	0.11	2.20	0.12	1.94
Education low	0.18	0.68	0.07	1.11	0.06	0.76	0.01	0.15	-0.02	-0.37	-0.00	-0.06
Education high	-0.59	-3.34	-0.20	-4.62	-0.30	-5.50	-0.09	-2.36	0.00	0.10	-0.03	-0.61
Alone	0.14	0.74	-0.02	-0.43	0.01	0.18	0.04	1.01	-0.03	-0.82	-0.08	-1.741
Illness long	0.92	6.04	0.83	22.83	0.84	18.37	-0.02	-0.53	-0.01	-0.32	0.11	2.88

	Model A		Model B	
Threshold parameters	Coeff	T-values	Coeff	T-values
γ_s^1	1.00	-		
γ_s^2	-0.09	-3.49		
γ_s^3	-0.05	-1.31		
γ_v^1	1.00	-		
γ_v^2	0.32	16.07		
γ_v^3	0.11	5.49		
Variances				
σ_s^2	1.00	-	1.00	-
σ_u^2	0.00	-	0.18	4.57
σ_v^2	1.00	-	0.67	35.94
σ_o^2	5.14	97.83	5.14	97.83
ρ	0.14	8.55	0.14	8.45
Vignette dummy	Coeff	Std.error	Coeff	Std.error
ϑ	2.46	0.03	2.13	0.06

Model A assumes that there is no heterogeneity in reporting behaviour and that response consistency does not hold (No DIF, No RC). Model B assumes that there is heterogeneity in reporting behaviour and that response consistency holds (DIF, RC). Since the objective measure is a continuous variable there are no thresholds in the objective part of the model. All estimates are based on a sample that is selected on the relevant self-assessment and vignette, the objective measure, and on the covariates. $N = 4788$.

2.B Self-assessment and vignette questions

Here the self-assessment and vignette questions of both SHARE waves are given. The vignette asked in wave 2 is chosen out the three asked in the first wave and is always the first vignette given here (c1 for cognition, b1 for breathing, and m1 for mobility). The vignette of wave 2 is slightly different from the one of wave 1 since the last sentence differs. In all cases the possible answer categories are 'none', 'mild', 'moderate', 'severe', and 'extreme'.

Cognition

Self-assessment question

- Overall in the last 30 days how much difficulty did you have with concentrating or remembering things?

Vignettes

c1: (Lisa) can concentrate while watching TV, reading a magazine or playing a game of cards or chess. Once a week she forgets where her keys or glasses are, but finds them within five minutes. Overall in the last 30 days, how much difficulty did (Lisa) have with concentrating or remembering things? (Wave 2: In your opinion, how much difficulty does (Lisa) have with concentrating or remembering things?)

c2: (Sue) is keen to learn new recipes but finds that she often makes mistakes and has to reread several times before she is able to do them properly. Overall in the last 30 days, how much difficulty did (Sue) have with concentrating and remembering things?

c3: (Eve) cannot concentrate for more than 15 minutes and has difficulty paying attention to what is being said to her. Whenever she starts a task, she never manages to finish it and often forgets what she was doing. She is able to learn the names of people she meets. Overall in the last 30 days, how much difficulty did (Eve) have with concentrating or remembering things?

Objective measures

- Immediate recall: Now, I am going to read a list of words from my computer screen. We have purposely made the list long so it will be difficult for anyone to recall all the words. Most people recall just a few. Please listen carefully, as the set of words cannot be repeated. When I have finished, I will ask you to recall aloud as many of the words as you can, in any order. Is this clear?

- Delayed recall: A little while ago, I read you a list of words and you repeated the ones you could remember. Please tell me any of the words you can remember now? (This question is directly asked after the final question assessing numerical ability.)

- Verbal fluency: Now, I would like you to name as many different animals as you can think of. You have one minute to do this.
- Numeracy: Next, I would like to ask you some questions which assess how people use numbers in everyday life.

  1) If the chance of getting a disease is 10 per cent, how many people out of 1,000 would be expected to get the disease?
  2) If the respondent's answer to question 1 was incorrect, the next question is: In a sale, a shop is selling all items at half price. Before the sale, a sofa costs 300 (local currency). How much will its cost in the sale?
  3) If the respondent's answer to question 1 was correct, the next question is: A second hand car dealer is selling a car for 6,000 (local currency). This is two-thirds of what it costs new. How much did the car costs new?
  4) If the answer to question 3 is correct, the next question will be: Let us say you have 2,000 (local currency) in a savings account. The accounts earns ten per cent interest each year. How much will you have in the account at the end of two years?

  The algorithm SHARE uses to compute the score for numeracy is as follows. Every respondents is asked the first question. If s/he gives the correct answer the score for numeracy equals 3 and the next question is question 3. If question 3 is answered correctly, the score for numeracy becomes 4 and the next question is question 4. If the respondent answers question 4 correctly, then the score for numeracy becomes 5. If the first question is answered incorrectly, then the score for numeracy equals 1 and the next question is question 2. If this question is answered correctly the score for numeracy will become 2.

Breathing

Self-assessment question

- In the last 30 days, how much of a problem did you have because of shortness of breath?

Vignettes

b1: (Mark) has no problems with walking slowly. He gets out of breath easily when climbing 20 meters uphill or a flight of stairs. In the last 30 days, how much of a problem did Mark have because of shortness of breath? (Wave 2: How much of a problem does (Mark) have because of shortness of breath?)

b2: (Paul) suffers from respiratory infections about once every year. He is short of breath 3 or 4 times a week and had to be admitted in hospital twice in the past month with a bad cough that required treatment with antibiotics. In the last 30 days, how much of a problem did (Paul) have because of shortness of breath?

b3: (Henri) has been a heavy smoker for 30 years and wakes up with a cough every morning. He gets short of breath even while resting and does not leave the house anymore. He often needs to be put on oxygen. In the last 30 days, how much of a problem did (Henri) have because of shortness of breath?

Objective measure (peak-flow test)

- The next test that I am going to ask you to perform will measure how fast you can expel air from your lungs. It is important that you blow as hard and as fast as you can. I would like you to perform the test two times. When we are ready to begin, I will ask you to stand up. Take as deep a breath as possible. Open your mouth and close your lips firmly around the outside of the mouthpiece, and then blow as hard and as fast as you can into the mouthpiece

Mobility

Self-assessment question

- Overall in the last 30 days, how much of a problem did you have with moving around?

Vignettes

m1: (Rob) is able to walk distances of up to 200 meters without any problems but feels tired after walking one kilometer or climbing more than one flight of stairs. He has no problems with day-to-day activities, such as carrying food from the market. Overall in the last 30 days, how much of a problem did (Rob) have with moving around? (Wave 2:In your opinion, how much of a problem does Rob have with moving around?)

m2: (Kevin) does not exercise. He cannot climb stairs or do other physical activities because he is obese. He is able to carry groceries and do some light household work. Overall in the last 30 days, how much of a problem did (Kevin) have moving around?

m3: (Tom) has a lot of swelling in his legs due to his health condition. He has to make an effort to walk around his home as his legs feel heavy. Overall in the last 30 days, how much of a problem did (Tom) have moving around?

Objective measure

- The next test measures the strength and endurance in your legs. I would like you to fold your arms across your chest and sit so that your feet are on the floor; then stand up keeping your arms folded across your chest. The respondent is then asked whether s/he thinks it is safe to stand up five from a chair five times without using their arms. When the answer is affirmative the respondent is asked to do the test and the interviewer records the time (in seconds) used for five stands.

2.C Description of covariates

Covariate	Description
Age	Depending on the wave, each respondent's age is calculated as 2004 or 2007 minus the year or birth, which information is provided by SHARE. We include dummies for age groups.
Education	Based on the International Standard Classification of Education (ISCED 97) we categorize education in three dummies: low, middle and high. We define ISCED levels 0 and 1 as low education, 2 and 3 as middle and 4, 5 and 6 as high education.
Gender	We include a dummy for being male.
Not alone	SHARE contains information on the marital status of its respondents. Possible answers are: (1) married and living together with spouse, (2) registered partnership, (3) married, living separated from spouse, (4) never married, (5) divorced, (6) widowed. We include a dummy for whether a respondent is living alone or not, where we define "not living alone" if marital status is reported as (1) or (2).
Physical activity	SHARE contains information on the frequency of physical activity, such as sports or heavy housework, of its respondents. Possible answers are (1) more than once a week, (2) once a week, (3) one to three times a month, (4) hardly ever, or never. We define that a respondents is engaged *often* in physical activity if the answer is (1), *sometimes* if the answer is (2) or (3) and *never* if the answer is (4).
Illness long	The SHARE questionnaire contains the following question: "Some people suffer from chronic or long-term health problems. By long-term we mean it has troubled you over a period of time or is likely to affect you over a period of time. Do you have any long-term health problems, illness, disability or infirmity?" The answer can be yes or no.

Chapter 3

The effect of private health insurance on medical care utilization and self-assessed health in Germany

3.1 Introduction

In Germany, employees are generally obliged to participate in the public health insurance system, where coverage is universal, co-payments and deductibles are moderate, and premia are based on income. However, they may buy private insurance instead, if their income exceeds the so-called compulsory insurance threshold.[1] Here, premia are based on age and health, persons may choose to what extent they are covered, and deductibles and co-payments are common.[2] These differences in the incentive structure may affect both health behavior and the demand for medical care. In particular, because of the higher co-payments and deductibles, privately insured patients have stronger incentives to invest in prevention to decrease the likelihood of occurrence of an illness. Therefore, even when the treatment provided to privately and publicly insured patients is exactly the same, we would expect privately insured patients to be less inclined to demand medical services.

An important difference affecting the supply of services is that for the same treatment the compensation doctors receive for privately insured patients is, on average, 2.3 times as high as the compensation for publicly insured patients (Walendzik, Gress, Manouguian, & Wasem, 2008). Therefore, doctors have an incentive to treat privately insured patients first, and more intensely, possibly providing better treatment (Jürges, 2009). This is reflected, for example, in lower waiting times for privately insured patients (Lungen, Stollenwerk, Messner, Lauterbach, & Gerber, 2008). These differences affecting the supply of services may also affect the demand for medical care. For example, persons may

[1] About 90 percent of the German population is insured in the public health insurance system. Most remaining persons buy private insurance (Colombo & Tapay, 2004).

[2] In our data (the sample also used for Tables 3.2 and 3.4 below), 70 percent of the privately insured persons who answered the respective question have insurance contracts that involve deductibles or co-payments.

be more inclined to see a doctor because the quality of the treatment is higher or waiting times are shorter.

The combination of demand and supply side incentives determines whether the amount of services consumed is higher or lower for privately insured persons, and which effect insurance type has on health. Ultimately, it is an empirical question whether more or less services are consumed and how health depends on insurance status.

In this chapter, we study the effect of being privately insured on the number of doctor visits, the number of nights spent in a hospital and self-assessed health. We do not estimate the effects of specific insurance characteristics but interpret the results in light of the fact that deductibles and co-payments are common features of private insurance contracts. An unusual feature of the German health insurance system allows us to control for selection into private insurance: as soon as income in the last year exceeds the so-called compulsory insurance threshold, persons become eligible to opt out of the public health insurance system and may buy private insurance instead. Random variation in income around this compulsory insurance threshold generates a natural experiment that allows us to conduct a regression discontinuity (RD) analysis.[3] This yields estimates of local average treatment effects for those persons who buy private insurance once becoming eligible. These effects are interesting to policymakers considering to increase the compulsory insurance threshold. Such an increase would force exactly those persons for whom we estimate the effect to be publicly insured, and hence our estimates are informative about the effects of such a policy change.

We use survey data from the German Socio Economic Panel (GSOEP) for our analysis because German administrative data, that contain accurate income measures, do not contain health related information. In the data, we find direct evidence for measurement error in income. Moreover, we find that there is a sizable number of persons who, according to their reported income, are not eligible to buy private insurance but at the same time report to be privately insured. The methodological contribution in this chapter is to model the measurement error in the so-called forcing variable, income in our case, within the RD framework. This then allows us to estimate the effects of interest.

Controlling for selection into private insurance we find a significant negative effect of being privately insured on the number of doctor visits for those persons who visit the doctor at least once in a three month period. At the same time, we find no significant effect on the number of nights spent in a hospital, which can arguably be influenced less by a person, and a positive effect on self-assessed health. This suggests that privately insured patients receive better or more intense treatment each time they see a doctor, or that they invest more in prevention.

The remainder of this chapter is organized as follows. Sections 3.2 and 3.3 discuss related results and the institutional details, respectively. In section 3.4, we provide information on the data and

[3] The RD approach has been suggested by Thistlethwaite and Campbell (1960) and has recently been developed by Hahn, Todd, and Van der Klaauw (2001). They show that under relatively mild assumptions the RD method can be interpreted as a local randomized experiment. This gives the results a strong internal validity. However, in general, a drawback is that the effect is only estimated for a small subset of the population of interest/the population that a social planner is concerned with. See also Imbens and Lemieux (2008), Lee and Lemieux (2010) and Van der Klaauw (2009) for recent discussions. Our setup is the same as in Battistin, Brugiavini, Rettore, and Weber (2009).

document that there is measurement error in income. Section 3.5 discusses the econometric approach, emphasizing our approach to modeling measurement error. Results are presented in section 3.6, and a sensitivity analysis is performed in section 3.7. Finally, section 3.8 concludes.

3.2 Related Literature

The empirical literature on demand for health services dates back at least to the 1970s when the RAND Health Insurance Experiment (HIE) was conducted. One important finding is that the use of medical services responds negatively to changes in cost sharing, with a stronger effect for outpatient care than for inpatient care (Newhouse, 1974; Manning et al., 1987). Additionally, insurance coverage is found to have an insignificant impact on health status of the overall population (Newhouse, 1993).

Conducting such a randomized experiment is, however, typically not feasible because of financial constraints, ethical considerations, or other reasons. The literature on health care utilization therefore mostly uses observational data. Levy and Meltzer (2004) discuss the related endogeneity of health insurance. They survey the literature on the effect of insurance coverage on health and conclude that there is a measurable effect only for the groups most likely to be the targets of public coverage expansions, namely infants, the elderly, and the poor. Cutler and Zeckhauser (2000) survey the literature on the effects of insurance characteristics on health and conclude that there is no clear evidence yet on which policies are most effective in achieving this. Buchmüller, Grumbach, Kronick, and Kahn (2005) survey the literature on the effect of health care utilization and conclude that the literature consistently finds significant effects of insurance on all types of utilization.

There are at least four studies for Germany that relate demand for medical services to insurance type. They all use the GSOEP data. Geil, Million, Rotte, and Zimmerman (1997) estimate a count data model for hospital visits on data from 1984-1989, 1992, and 1994. They find no relationship between insurance coverage and the hospitalization decision. Riphahn, Wambach, and Million (2003) estimate a bivariate count data model for physician and hospital visits. They use data from 1984 through 1995 and find that neither hospital nights nor doctor visits depend on the insurance type of a person. Pohlmeier and Ulrich (1995) and Jürges (2009) both estimate a negative binomial hurdle model. Pohlmeier and Ulrich (1995) use data from 1985 and find that privately insured persons are less likely to contact a general practitioner but the number of visits once they do so is not significantly different from the one for publicly insured patients. Jürges (2009) uses data from 2002 and finds that privately insured persons are less likely to visit a doctor at all, but given that they do the number of doctor visits is significantly larger than that of patients covered by public health insurance. What all four papers have in common is that they do not control for selection into private insurance.

An approach that has become popular to control for selection into health insurance is to utilize *natural experiments*, a setting in which a certain group of persons experiences an unexpected and exogenous change in the incentive structure, and to perform a *regression discontinuity (RD) analysis*. In the context of health insurance, Chiappori, Durand, and Geoffard (1998) analyze data from a French natural experiment and find no evidence of moral hazard for general practitioner or specialists office visits. Two recent applications of the RD approach to health insurance are Card, Dobkin, and Maestas (2008) and Card, Dobkin, and Maestas (2009), who exploit the fact that in the United States persons of age 65 and above are generally insured because they are eligible for Medicare. They find that this leads to an increase in doctor visits, with the highest increase for groups that previously lacked coverage. There is a sharp increase in hospitalization, but effects differ across groups and by type of admission. Moreover, they find an increase in the number of procedures performed, treatment intensity, as well as a significant and large reduction in mortality.

3.3 Institutional details

In Germany, about 90 percent of the population is publicly insured (Colombo & Tapay, 2004). Buying public insurance is mandatory for dependent employees as long as their income does not exceed the so-called compulsory insurance threshold. The public insurance premium is a certain percentage (currently around 15 percent that is equally divided between the employer and the employee) of gross income up to the so-called contribution ceiling, and equal to it thereafter.[4]

Table 3.1 shows the contribution ceilings and the compulsory insurance thresholds by the year in which the income was earned. To see how the system works consider a person whose income, including all extra payments, in 2000 was $40,000$ Euros. Then, he is eligible to buy private insurance in 2001 because his income exceeded $39,574$ Euros, the compulsory insurance threshold. If his income stays the same or decreases in 2001, then he will have to join the public insurance system again in 2002 because the compulsory insurance threshold is $40,034$ Euros for income earned in 2001. He can apply for an exemption if he loses eligibility *solely* due to the increase in the compulsory insurance threshold, i.e. if his income in 2001 is at least $39,574$ Euros. This applies to very few persons in our data, about a tenth of a percent of all person-year observations, and we therefore abstract from this exemption in the remainder.[5]

In general, once a person has bought private health insurance he can only get back into the public system when he becomes unemployed (provided that he is younger than 55) or when his income falls below the compulsory insurance threshold (Colombo & Tapay, 2004).

Because of a reform the compulsory insurance threshold increased substantially for income earned in 2003 and later. A special rule applied to persons who actually bought private insurance in 2003, but

[4]See Jürges (2009) and the references therein for more details on this and the following discussion.

[5]This is because income typically increases faster than the compulsory insurance threshold. Sozialgesetzbuch V §8 also defines two more situations in which exemptions are granted, namely a temporary reduction in working hours and a combination of part-time employment and paternity leave. We abstract also from these possibilities because they are not widely used. If, to the contrary, the number of persons who would be granted an exemption was large, say 1 percent of all persons, then we would overestimate the discontinuity in the probability to be privately insured at the compulsory insurance threshold by 1 percentage point because we assume that this probability is zero for persons earning less than the compulsory insurance threshold, and we would therefore underestimate the local average treatment effects. It follows from (3.1) below that we would overestimate the effect by about 5 percent if the discontinuity in the denominator was, e.g., 19 instead of 20 percent.

Table 3.1: Contribution ceiling and compulsory insurance threshold

Year	Contribution ceiling	Compulsory insurance threshold	Mean income
1994	34,968	34,968	24,633
1995	35,892	35,892	25,126
1996	36,816	36,816	25,905
1997	37,728	37,728	26,423
1998	38,652	38,652	26,660
1999	39,108	39,108	27,060
2000	39,574	39,574	27,358
2001	40,034	40,034	27,741
2002	40,500	40,500	28,231
2003	41,400	45,900	28,626
2004	41,850	46,350	28,938
2005	42,300	46,800	29,060

Reported for West Germany by year in which the income was earned. Amounts are nominal amounts per year and in Euros. The contribution ceilings and the compulsory insurance thresholds are based on Sozialgesetzbuch V and own calculations. Mean income is taken from Sozialgesetzbuch VI, Anlage 1.

who were not eligible for this anymore according to the new thresholds. They could still buy private insurance provided that their income is at least equal to the contribution ceiling, which increased only moderately.[6]

Contributions for private health insurance are mainly based on health and age, so buying private insurance is especially attractive for young persons. As a consequence of this, and because of the fact that private insurers are allowed to reject persons, the risk pool of the private insurers is much better than in the public system.

Coverage is universal in the public system. Deductibles and co-payments are limited. Privately insured persons can buy better care, e.g. treatment by the head doctor in a hospital or a single room in a hospital, but this comes at a higher price. Deductibles and co-payments are much more common, and many insurers offer a rebate if a person did not use medical services in the past calendar year. Unfortunately, specific characteristics of private insurance are not recorded in our data.

At this point it is worth noticing that there is a feature called family insurance in the German public health insurance system. A spouse is automatically insured if a person is insured. For this it is mandatory that the spouse is not full time self-employed and that the spouse does not earn more than a rather low specified amount. If a married man is working then this system generates incentives against working for his wife because then she would have to pay contributions which amount to about 7.5 percent of her gross wage (the employer matches this and pays about the same amount to the system). The family insurance feature does not exist for private health insurance and therefore,

[6]We excluded these persons from the empirical analysis.

individual insurance has to be purchased for each family member.

As already pointed out before, insurance status has important consequences for the compensation of doctors. For a given treatment the compensation doctors receive for privately insured patients is, on average, 2.3 times as high as the compensation for publicly insured patients (Walendzik et al., 2008). Furthermore, there is indirect evidence that doctors face strong time constraints when treating patients. The consultation length for the average (publicly insured) person is very low in Germany.[7] Deveugele, Derese, Van den Brink-Muinen, Bensing, and De Maeseneer (2002, Table 4) compare the average consultation length for general practitioners in six countries and find that with 7.6 minutes it is lowest in Germany. It is highest in Switzerland, where it is equal to 15.6 minutes. Together with the differences in the compensation this suggests that doctors dedicate more time to privately insured patients.

3.4 Data

The GSOEP we use in this study contain information at the individual level on medical care utilization, self-assessed health, and background variables. We analyze data from West Germany for the period from 1995 to 2006.[8]

Our sample is constructed such that eligibility to opt out of the public insurance system is exclusively determined by income. Unemployed persons who receive unemployment benefits are required to be in the public health insurance system. For them there is no way to opt out and therefore they are excluded. Self-employed, civil servants, soldiers, teachers in private schools and students are not required to be in the public system, even if their income is below the compulsory insurance threshold. Hence, eligibility does not depend on income and therefore they are excluded from the sample as well. Retired persons, who receive a public pension, are required to have public health insurance. They may opt out if insurance was not mandatory in at least five years after the age of 55 and most of the time before that. Hence eligibility is only weakly related to income and therefore they are excluded. Persons of age 55 and older are excluded for two reasons. First, because for them various ways to opt for (early) retirement exist. Second, because for them it is difficult to get back into the public health insurance system. Persons under the age of 25 are excluded because a large fraction of them is covered by their parents' insurance.

To summarize, our study population consists of West German persons, aged 25 to 55, with a regular employment contract for whom eligibility to opt out of the public health insurance system is exclusively determined by income.

Table 3.2 contains descriptive statistics for the variables we use in the analysis. The first set of

[7] Recall that about 90 percent of the persons are publicly insured. See footnote 1.

[8] We do not use data before 1995 because the question on the number of doctor visits was phrased differently. We use data only up to 2006 because from 2007 onwards persons had to earn more than the compulsory insurance threshold in three consecutive years in order to be eligible to buy private insurance. East German persons have been excluded because it turned out that for them, even when we control for measurement error in income, there is no jump in the probability to be privately insured when income is equal to the compulsory insurance threshold.

Table 3.2: Descriptive statistics

	(1) Ineligible	(2) Eligible	(3) Public insurance	(4) Private insurance	(5) Total
At least 1 doctor visit	0.619	0.594	0.611	0.521	0.613
	-	-	-	-	-
Doctor visits given at least 1 visit	3.304	2.920	3.243	2.904	3.223
	(4.212)	(3.365)	(4.094)	(3.287)	(4.052)
Doctor visits	2.045	1.733	1.999	1.651	1.977
	(3.682)	(2.963)	(3.581)	(2.866)	(3.541)
At least 1 night in hospital	0.079	0.065	0.078	0.057	0.076
	-	-	-	-	-
Nights in hospital	0.862	0.655	0.833	0.572	0.613
	(5.027)	(4.033)	(4.844)	(4.603)	(4.830)
Self-assessed health	3.585	3.696	3.596	3.799	3.609
	(0.850)	(0.790)	(0.841)	(0.777)	(0.838)
Gross income	23,914.80	61,249.00	29,879.10	63,515.70	31,998.60
	(9,693.90)	(27,755.60)	(18,005.30)	(41,082.40)	(21,837.50)
Years of education	11.533	13.971	11.881	14.785	12.065
	(2.228)	(2.929)	(2.471)	(2.945)	(2.601)
Married	0.654	0.746	0.676	0.649	0.674
	-	-	-	-	-
Male	0.500	0.848	0.562	0.784	0.576
	-	-	-	-	-
Age	39.393	42.161	39.872	41.775	39.992
	(8.338)	(7.206)	(8.229)	(7.274)	(8.186)
N	35,822	9,900	42,841	2,881	45,722

Means and standard deviations (in parentheses). For binary variables only proportions are shown. Sample consists of dependent employees for whom eligibility to opt out of the public health insurance system is exclusively determined by income. t-tests show that for all variables the difference in the mean between ineligible and eligible persons and between publicly insured and privately insured persons significantly different from zero.

rows contains the outcome variables.[9] Persons who are are eligible (to buy private insurance) and privately insured visit the doctor slightly less often, and report to be in slightly better health. They report to be less likely to stay in a hospital and to spend less nights in a hospital on average.

The second set of rows contains summary statistics for individual characteristics. Gross income is, by construction, on average higher for eligibles. In light of this it is not surprising that it is higher for privately insured (because only those with high enough incomes are eligible to buy private insurance). The remaining rows are informative about selection into private insurance. Given the characteristics of public and private insurance it is relatively more attractive to buy private insurance for persons who are not married. This is because spouses whose income is relatively low are automatically covered by the insurance of the person. This is reflected by the fact that privately insured persons are less likely to be married. Also, they are older and better educated.

[9]For the self-assessed health question, ‘bad’ is re-coded as a 1, ‘poor’ as 2, and so on, up to ‘very good’ as 5. Hence, a positive association between health and private insurance would be reflected in a positive coefficient on an indicator for private insurance in an ordinary least squares regression.

Table 3.3: Ordinary least squares estimates

	(1) At least one doctor visits	(2) Doctor visits for subsample	(3) Doctor visits	(4) At least one night in hospital	(5) Nights in hospital	(6) Self-assessed health
Baseline outcome	0.622***	3.469***	2.197***	0.074***	0.397*	4.177***
	(0.033)	(0.340)	(0.241)	(0.016)	(0.222)	(0.068)
Private health insurance	-0.021	0.170	0.045	-0.009	-0.046	0.083**
	(0.017)	(0.163)	(0.117)	(0.007)	(0.112)	(0.035)
Years of education	0.002	-0.076***	-0.041***	-0.003***	-0.044***	0.028***
	(0.002)	(0.016)	(0.011)	(0.001)	(0.010)	(0.004)
Married	0.011	-0.141	-0.065	0.003	-0.098	0.005
	(0.009)	(0.094)	(0.068)	(0.004)	(0.068)	(0.019)
Gender (male)	-0.182***	-0.448***	-0.873***	-0.019***	-0.138**	0.046**
	(0.009)	(0.099)	(0.078)	(0.005)	(0.069)	(0.022)
Age	0.002***	0.038***	0.031***	0.001***	0.032***	-0.025***
	(0.001)	(0.006)	(0.004)	(0.000)	(0.004)	(0.001)
N	23,820	14,360	23,820	23,820	23,820	23,820

Standard errors are clustered at the individual level and shown in parentheses. $*, **, ***$ denote significance at the 10, 5, and 1 percent level, respectively. Sample consists of dependent employees for whom eligibility to opt out of the public health insurance system is exclusively determined by income and who earn between $15,000$ Euros less and $25,000$ Euros more than the compulsory insurance threshold.

Table 3.3 contains ordinary least squares estimates of the outcome variables in the respective columns on insurance status, years of education, marital status and age.[10] The respective coefficient on insurance status is the observed difference in the average outcome variable between privately and publicly insured persons once we control for years of education, marital status and age. We find that there is only a significant difference in self-reported health.[11] This is in line with Geil et al.'s (1997) finding of no relationship between insurance coverage and the hospitalization decision and Riphahn et al.'s (2003) finding that neither hospital nights nor doctor visits depend on the insurance type of the person. It is also in line with Pohlmeier and Ulrich's (1995) finding that once privately insured persons see a doctor, the number of visits is not significantly different from the number of visits for publicly insured patients. In contrast to our results, however, they and Jürges (2009) find that privately insured patients are significantly less likely to contact a doctor. This is because we have restricted the estimation sample to contain only persons who earn between $15,000$ Euros less and $25,000$ Euros more than the compulsory insurance threshold. Once we loosen this restriction, we also find a significant negative relationship. Finally, Jürges (2009) finds that given that persons visit the doctor at least once, the number of doctor visits is significantly larger for privately insured patients.

[10]To make the results comparable to our main results that are below the sample only contains persons who earn between $15,000$ Euros less and $25,000$ Euros more than the compulsory insurance threshold. The main difference when we use the unrestricted sample is that the coefficient on private insurance status for at least one doctor visit is negative and significant at the 5 percent level, with a point estimate of -0.029, and the coefficient on at least one hospital night is -0.003 and significant at the 1 percent level. One explanation for this is that the unrestricted sample contains more persons with high incomes, and a fraction of them is privately insured, and more persons with low incomes who are publicly insured, and that persons with higher incomes are of better health.

[11]Following Pohlmeier and Ulrich (1995) and Jürges (2009) we also estimated hurdle models, which yielded the same qualitative results.

We get the same result once we additionally condition on health, which arguably is highly endogenous as health is directly related to the number of doctor visits and the number of hospital nights. We shall stress once more at this point that all of the above results, including the ones reported in Table 3.3, are purely descriptive as insurance status is very likely to be endogenous.

As described above, we control for selection into private insurance by using random variation in income around the compulsory insurance threshold. A key variable in our analysis is thus gross yearly income. This is not reported by the GSOEP respondents but constructed from their reports on their average gross monthly income in the previous year and their reports on supplementary income such as 13th month salary, 14th month salary, Christmas bonus, vacation pay, profit share, premia, and bonuses. Using self-reported income and Table 3.1 we can compute the eligibility status for every person.

Table 3.4: Eligibility and health insurance type

	Public insurance	Private insurance	N
Ineligible	35,245	577	35,822
Eligible	7,596	2,304	9,900
N	42,841	2,881	45,722

Sample consists of dependent employees for whom eligibility to opt out of the public health insurance system is exclusively determined by income.

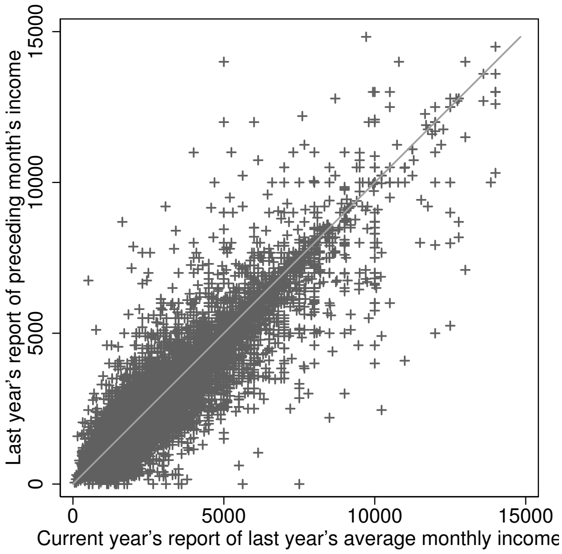
Table 3.4 shows that there is a sizable number of persons, 577, who, according to their reported income, are not eligible to buy private insurance, but at the same time report to have done so. These 577 persons constitute 20 percent of the persons with private health insurance. Misreporting insurance status or measurement error in income may both be valid explanations for this.[12]

We consider it to be more plausible that income is measured with error because income is a real number, and may thus be recalled with errors, whereas insurance status is more easily known because it is typically either public or private insurance. Moreover, there is direct evidence for measurement error in income because the GSOEP questionnaire asks respondents twice about their monthly income in a given year.[13] In particular, respondents are asked about the income they received in the preceding month (without extra payments) and about their average monthly income in the previous year. This provides us with two measures of monthly income for the same year. If both income reports would be reported without any error, and if the within year variance in monthly income is low, then both measures should be close to one another. That is, the data points in a scatter plot should be close to

[12]There is an extensive literature on measurement error in income, see for example Bound, Brown, and Mathiowetz (2001) for a survey. In order to study the accuracy of survey reports, they are typically compared with either employers' or administrative records. Some studies find that survey reports are highly correlated with record values, while others find much lower correlations. The mean of survey reports is found to be close to the mean of the record values. That is, under- or over-reporting, if present, is found to be moderate on average.

[13]This is not the case for the total yearly income that we use to determine eligibility. Yearly income includes extra payments such as holiday pay. The fact that it is yearly income and not monthly income that determines eligibility is the reason that we do not exploit the availability of two monthly income measures in the main analysis.

Figure 3.1: Joint distribution of the two income measures



Sample consists of dependent employees for whom eligibility to opt out of the public health insurance system is exclusively determined by income. For this figure we use only income reports below $15,000$ Euros per month.

the 45 degree line. Such a scatter plot is shown in Figure 3.1. The deviations from the 45 degree line are substantial. This strongly suggests that there is measurement error in income.[14]

3.5 Econometric approach

Let $(y_i(0), y_i(1))$ be the pair of potential outcomes for each member i of the study population. In our case $y_i(0)$ denotes the health outcome the i-th person would experience in case public health insurance was assigned to him and $y_i(1)$ denotes the health outcome the i-th person would experience if private health insurance was assigned. That is, we consider private health insurance to be the "treatment."

Let z_i^* denote the difference between income earned in the previous year and the corresponding compulsory insurance threshold. Then, a person is eligible to buy private health insurance when $z_i^* \geq 0$. Buying private insurance is voluntary for eligible persons so that some will buy it while others will not.

Following Hahn et al. (2001) we make three assumptions. First, we assume that the effect of private insurance is independent of z_i^* around $z_i^* = 0$.[15]

[14] This is robust to controlling for working hours and job changes by means of a regression. The R^2 in this regression is 0.846, meaning that 15.4 percent of the variation in the report of previous year's average income remains unexplained.

[15] In the empirical analysis we make the stronger assumption that the effect of private insurance is not only locally independent of income, but over the whole range of incomes reported in our sample. We make this assumption because it implies that the local average treatment effect is the same in all years. Otherwise, it would not be because the threshold income level, at which the effect is defined, increases over time. Under the stronger independence-of-income assumption, we can obtain more efficient estimates of the effect, because we do not have to estimate a different effect for each year. We

This assumption is plausible as long as variation in income is independent of the effect of private insurance. It could be violated if persons were able to manipulate their income such that they become eligible to buy private insurance and the effect of private insurance was different for those persons. To the best of our knowledge there is no evidence for such manipulations in Germany. Second, we assume that the mean value of $y_i(0)$ conditional on z_i^* is a continuous function of z_i^* at $z_i^* = 0$. This assumption holds if the mean health outcome would be a smooth function in income around the compulsory insurance threshold once public insurance was exogenously assigned to everybody. This is highly plausible. Third, we assume that the decision to buy private insurance is monotone in eligibility. This is the monotonicity condition of Imbens and Angrist (1994). It holds by construction because ineligibles cannot buy private insurance. Under these assumptions the average treatment effect for those persons that would buy private health insurance when becoming eligible is given by

$$\Delta^{LATE} \equiv \mathrm{E}\left(y_i\left(1\right) - y_i\left(0\right)|p_i = 1, z_i^* = 0\right) = \frac{\mathrm{E}\left(y_i|z_i^* = 0^+\right) - \mathrm{E}\left(y_i|z_i^* = 0^-\right)}{\mathrm{E}\left(p_i|z_i^* = 0^+\right)}, \tag{3.1}$$

where y_i is the *observed* health outcome, p_i is an indicator of private insurance, $\mathrm{E}\left(\cdot|z_i^* = 0^+\right) \equiv \lim_{\delta\downarrow 0} \mathrm{E}\left(\cdot|z_i^* = \delta\right)$, and $\mathrm{E}\left(\cdot|z_i^* = 0^-\right) \equiv \lim_{\delta\uparrow 0} \mathrm{E}\left(\cdot|z_i^* = \delta\right)$. This effect is of particular interest because it is directly related to the question what the effect of requiring all persons with incomes slightly above the compulsory insurance threshold to buy public insurance would be, namely the opposite of the effect we estimate.

Measurement error in income leads to misclassification of eligibility. Importantly, this misclassification is not independent of the true underlying income because if the true underlying income is below (above) the compulsory insurance threshold the classification error can only be that the person is (not) eligible to buy private insurance. This precludes the use of an instrumental variables approach to estimating the unknown quantities in the numerator and denominator in (3.1).

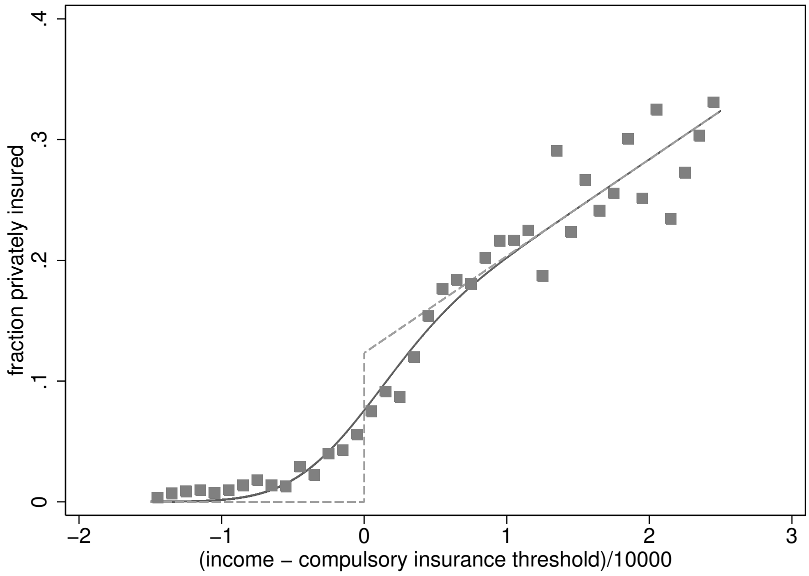
The effect of the measurement error in income on estimates of these quantities is that no discontinuity in reported income is observed at the threshold (Battistin et al., 2009). In Figure 3.2, the dots are fractions of privately insured persons, which we plot against the difference in income and the compulsory insurance threshold. The figure shows that these fractions are not zero if reported income is below the compulsory insurance threshold, i.e. if the value of the difference on the horizontal axis is negative, and that indeed there is no discontinuity in the fraction of privately insured at the threshold.

Towards estimating the local average treatment effect in the presence of measurement error we now develop an expression for the probability to be privately insured, which is equal to the conditional expectation of the indicator for being privately insured. Our approach is parametric and our main assumption is that $z_i^* = z_i - u_i$, where u_i is normally distributed independent of z_i and has mean zero and variance σ_u^2.[16] Furthermore, u_i is assumed to be independent of private insurance status and the

have also estimated models in which the effect was allowed to differ by year. A test for the equality of the different estimates was not rejected.

[16] Notably, this is not classical measurement error. For classical measurement error we have $z_i = z_i^* + u_i$, which is equivalent, but we assume that u_i is independent of z_i^*, and not of z_i. See Wooldridge (2002) for a discussion.

Figure 3.2: Probability to be privately insured



Sample consists of dependent employees for whom eligibility to opt out of the public health insurance system is exclusively determined by income and who earn between $15,000$ Euros less and $25,000$ Euros more than the compulsory insurance threshold ($N = 24,203$). The specification here imposes, for illustration, that the size of the discontinuity is the same in all years.

potential outcomes. We specify the (piecewise) linear probability model

$$\mathrm{E}\left(p_i|z_i^*\right) = \begin{cases} 0 & \text{if } z_i^* < 0 \\ \alpha + \beta z_i^* & \text{if } z_i^* \geq 0. \end{cases}$$

Recall that when true income is below the compulsory insurance threshold, i.e. when $z_i^* < 0$, then the probability of being privately insured is zero because ineligibles may not buy private insurance. Conversely, when true income exceeds the compulsory insurance threshold, i.e. when $z_i^* \geq 0$, persons may buy private insurance.

We show in Appendix A that under these assumptions

$$\mathrm{E}\left(p_i|z_i\right) = \Phi\left(\frac{z_i}{\sigma_u}\right) \cdot \left(\alpha + \beta z_i + \beta \sigma_u \frac{\phi\left(\frac{z_i}{\sigma_u}\right)}{\Phi\left(\frac{z_i}{\sigma_u}\right)}\right), \tag{3.2}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function and $\phi(\cdot)$ is the standard normal probability density function. Notably, this is the prediction for the relationship between the probability to be privately insured and the difference between reported income and the compulsory insurance threshold, z_i. The solid line in Figure 3.2 shows the estimated relationship for our data when we pool data across all years. The dots are sample fractions of privately insured. Comparing them to the solid line shows that the fit is reasonably good. Finally, the dashed line in this figure is the underlying relationship between the probability to be privately insured and the difference between actual (measured without error) yearly income and the compulsory insurance threshold, z_i^*.

A similar expression can be obtained for $\mathrm{E}\left(y_i|z_i\right)$. This involves specifying different linear functions to the left and right of the discontinuity,

$$\mathrm{E}\left(y_i|z_i^*\right) = \begin{cases} \alpha_0 + \beta_0 z_i^* & \text{if } z_i^* < 0 \\ \alpha_1 + \beta_1 z_i^* & \text{if } z_i^* \geq 0, \end{cases}$$

so that, under our assumptions,

$$\begin{aligned} \mathrm{E}\left(y_i|z_i\right) = & \left(1 - \Phi\left(\frac{z_i}{\sigma_u}\right)\right)\left(\alpha_0 + \beta_0 z_i - \beta_0 \sigma_u \frac{\phi\left(\frac{z_i}{\sigma_u}\right)}{1-\Phi\left(\frac{z_i}{\sigma_u}\right)}\right) \\ & + \Phi\left(\frac{z_i}{\sigma_u}\right)\left(\alpha_1 + \beta_1 z_i + \beta_1 \sigma_u \frac{\phi\left(\frac{z_i}{\sigma_u}\right)}{\Phi\left(\frac{z_i}{\sigma_u}\right)}\right). \end{aligned} \tag{3.3}$$

The parameters for both $\mathrm{E}\left(p_i|z_i\right)$ and $\mathrm{E}\left(y_i|z_i\right)$ are jointly estimated using the feasible generalized nonlinear least squares estimator for nonlinear systems of equations. From these parameter estimates we then calculate the local average treatment effect. For this observe that α, α_0, and α_1 are equal to $\mathrm{E}\left(p_i|z_i^* = 0^+\right)$, $\mathrm{E}\left(y_i|z_i^* = 0^-\right)$, and $\mathrm{E}\left(y_i|z_i^* = 0^+\right)$, respectively. Hence, it follows from (3.1) that the local average treatment effect is given by

$$\Delta^{LATE} = \frac{\alpha_1 - \alpha_0}{\alpha}. \tag{3.4}$$

3.6 Results

We jointly estimate the equation for the probability to be privately insured conditional on reported income, (3.2), and the equation for medical care utilization conditional on reported income, (3.3). Throughout, we allow the probability to be privately insured to have year specific jumps at the compulsory insurance threshold. This is reasonable since the compulsory insurance threshold changed over time (see Table 3.1). We impose that the local average treatment effect is the same in all years, i.e. we impose that Δ^{LATE}, our parameter of main interest, is not only locally independent of z_i^*, but over a whole range of values.[17] Then, it follows from (3.4) that we can replace α_1 by $\alpha_0 + \Delta^{LATE} \cdot \alpha$. Notice that the size of both the numerator and the denominator in (3.1) is still allowed to vary across years, but we impose that the relative change in both is the same. Finally, we impose that expected health outcomes do not depend on income, i.e. $\beta_0 = \beta_1 = 0$.[18]

[17] The assumption that Δ^{LATE} is the same in all years is not necessary. In fact, we have also estimated models in which Δ^{LATE} was allowed to differ by year. A test for the equality of the different Δ^{LATE} estimates could not be rejected.

[18] We conducted several robustness checks. By jointly estimating more general models (involving non-zero slopes that were allowed to differ across years, e.g.) and our baseline specification we could check, respectively, whether treatment effect estimates were significantly different from the ones obtained using the baseline specification, and in general they were

We first estimate (3.2) alone. Results are reported in Table 3.5.[19] Coefficient estimates are marginal effects because the underlying model is a linear probability model. The probability is zero for negative z_i^* and for positive z_i^* it is linear in it. The results indicate that for all years there is a discontinuous jump in the probability to buy private insurance at $z_i^* = 0$. In 1995, the size of the jump is 9 percentage points, in 1996 it is 6 percentage points. From 1997 to 2001 the jump is about 10 percentage points. In 2002 and 2003, the jump increases slightly, and between 2004 and 2006 the jump substantially increases to approximately 18 percentage points. Supposedly, this is due to the increase in the compulsory insurance threshold for income earned in 2003, which affects the jump in the probability to be privately insured in 2004. For all persons in our estimation sample the predicted value for the probability to be privately insured is between 0 and 1.

Table 3.6 presents the estimates of Δ^{LATE} for doctor visits in the past three months, the number of nights spent in a hospital, and self-assessed health. The respective baseline outcome is the average outcome for publicly insured persons for whom true income is equal to the compulsory insurance threshold.

In specification (1), we use an indicator for at least one doctor visit as the dependent variable. This is a linear probability model since the expected outcome is a probability. 60.6 percent of the publicly insured persons see a doctor at least once within a three month period. We find no significant effect of private insurance on this. In specification (2), we estimate the effect of private insurance on the number of doctor visits for those persons who visit a doctor at least once. The baseline outcome is 3.329 doctor visits. The effect of private insurance on this is estimated to be negative and significant at the 1 percent level. The estimated magnitude of the effect, however, seems to be too big. Specification (3) is for the number of doctor visits in the entire sample. This is a combination of the two effects we discussed above. The mean baseline outcome is estimated to be 2.013. The estimated effect is negative and significant at the 1 percent level, but again the magnitude of the point estimate is too big as it exceeds the baseline in terms of the magnitude.

Manning, Morris, and Newhouse (1981) argue that the decision to visit a doctor at all, the so-called contact decision, is made by the person, whereas the number of visits is mainly determined by the doctor. However, it could also be that the patient and the doctor jointly determine the number of visits, or that fewer visits are needed for privately insured patients because they have invested in prevention. Furthermore, it could be that privately insured patients are treated more intensely so that less doctor visits are necessary. This is sensible because doctors are paid based on the number of treatments, not on the number of visits itself, and receive a higher compensation when they treat privately insured patients. They are time constrained and may thus focus on treating privately insured patients first (Lungen et al., 2008; Jürges, 2009), while spending relatively little time on publicly insured patients (Deveugele et al., 2002).

In specification (4) we use an indicator for at least one night spent in a hospital as the dependent variable. This is also a linear probability model. 7.4 percent of the publicly insured spend at least

not.

[19] Estimates are very similar when we estimate (3.2) and (3.3) together.

Table 3.5: Probability to be privately insured

(Gross income - threshold)/10000	0.075***
	(0.005)
Discontinuity 1995	0.089***
	(0.013)
Discontinuity 1996	0.064***
	(0.013)
Discontinuity 1997	0.099**
	(0.041)
Discontinuity 1998	0.098***
	(0.014)
Discontinuity 1999	0.107***
	(0.013)
Discontinuity 2000	0.101***
	(0.010)
Discontinuity 2001	0.109***
	(0.011)
Discontinuity 2002	0.132***
	(0.010)
Discontinuity 2003	0.114***
	(0.010)
Discontinuity 2004	0.193***
	(0.011)
Discontinuity 2005	0.191***
	(0.012)
Discontinuity 2006	0.178***
	(0.011)
σ_u	0.463***
	(0.034)
R^2	0.184
N	24,203

Standard errors are clustered at the individual level and shown in parentheses. $*, **, ***$ denote significance at the 10, 5, and 1 percent level, respectively. Sample consists of dependent employees for whom eligibility to opt out of the public health insurance system is exclusively determined by income and who earn between $15,000$ Euros less and $25,000$ Euros more than the compulsory insurance threshold.

Table 3.6: Baseline specification

	(1) At least one doctor visits	(2) Doctor visits for subsample	(3) Doctor visits	(4) At least one night in hospital	(5) Nights in hospital	(6) Self-assessed health
Δ^{LATE}	-0.079	-3.746***	-2.137***	-0.063*	-1.084*	0.449***
	(0.076)	(0.945)	(0.546)	(0.035)	(0.572)	(0.160)
Baseline outcome	0.606***	3.329***	2.013***	0.074***	0.783***	3.614***
	(0.005)	(0.054)	(0.039)	(0.002)	(0.039)	(0.011)
N	24,203	14,579	24,203	24,203	24,203	24,203

Standard errors are clustered at the individual level and shown in parentheses. $*, **, ***$ denote significance at the 10, 5, and 1 percent level, respectively. Sample consists of dependent employees for whom eligibility to opt out of the public health insurance system is exclusively determined by income and who earn between $15,000$ Euros less and $25,000$ Euros more than the compulsory insurance threshold.

one night in a hospital. The results indicate that there is no significant effect of private insurance on this (at the 5 percent level). Specification (5) is for the number of nights spent in a hospital and also here we find no significant effect of private insurance (also at the 5 percent level). These findings for hospital nights are in line with those of Geil et al. (1997) and Riphahn et al. (2003), and is intuitively plausible as the number of nights spent in a hospital can be influenced less by the person than the the number of doctor visits is.

Finally, we find that private insurance has a positive effect on health. Although again, the size of the effect seems to be too big, it is plausible that privately insured patients report to be of better health because they either invest more in prevention or because they receive better treatment once they visit a doctor.

3.7 Sensitivity analysis

Generally, we do not have to control for covariates when performing an RD analysis unless the distribution of the covariates changes when we move from the left to the right of the discontinuity (Imbens & Lemieux, 2008). The measurement error in the forcing variable, however, prevents us from performing the usual tests. However, it is still feasible to perform the analysis incorporating a dependence of the baseline outcome and the probability to be privately insured on additional covariates. Table 3.7 reports the results. They are similar to our baseline results.

Some of the studies that use GSOEP data additionally condition on health when estimating the relationship between private insurance coverage and the health outcomes (Jürges, 2009, e.g.). For two reasons we consider it reasonable to condition on previous period's health instead of current health. First, one of the outcomes in this study is current period's health so that conditioning on current health is not sensible, at least for this outcome. Second, current period's health is likely to be endogenous. We condition on previous period's health by re-estimating the model for persons who report in the previous period that their health is "satisfactory." The results were obtained using a two-

Table 3.7: Specification with covariates

	(1) At least one doctor visits	(2) Doctor visits for subsample	(3) Doctor visits	(4) At least one night in hospital	(5) Nights in hospital	(6) Self-assessed health
Δ^{LATE}	0.104	-2.480***	-0.819	-0.020	-0.687	0.499***
	(0.083)	(0.875)	(0.533)	(0.039)	(0.621)	(0.171)
Baseline outcome	0.598***	3.263***	1.964***	0.073***	0.770***	3.612***
	(0.005)	(0.053)	(0.039)	(0.002)	(0.040)	(0.011)
Years of education	0.000	-0.075***	-0.048***	-0.003***	-0.048***	0.026***
	(0.002)	(0.016)	(0.012)	(0.001)	(0.011)	(0.004)
Married	0.011	-0.119	-0.053	0.003	-0.084	0.011
	(0.009)	(0.095)	(0.068)	(0.004)	(0.068)	(0.019)
Gender (male)	-0.188***	-0.395***	-0.861***	-0.018***	-0.126*	0.032
	(0.009)	(0.098)	(0.078)	(0.005)	(0.069)	(0.022)
Age	0.002***	0.036***	0.029***	0.001***	0.031***	-0.025***
	(0.001)	(0.006)	(0.004)	(0.000)	(0.004)	(0.001)
N	23,830	14,360	23,830	23,830	23,830	23,830

Standard errors are clustered at the individual level and shown in parentheses. $*, **, ***$ denote significance at the 10, 5, and 1 percent level, respectively. Sample consists of dependent employees for whom eligibility to opt out of the public health insurance system is exclusively determined by income and who earn between $15,000$ Euros less and $25,000$ Euros more than the compulsory insurance threshold.

Table 3.8: Baseline specification for subsample of persons whose health in the previous period was "satisfactory"

	(1) At least one doctor visits	(2) Doctor visits for subsample	(3) Doctor visits	(4) At least one night in hospital	(5) Nights in hospital	(6) Self-assessed health
Δ^{LATE}	-0.257	-4.421**	-4.284***	-0.083	-3.220***	0.053
	(0.174)	(1.869)	(1.447)	(0.093)	(1.113)	(0.234)
Baseline outcome	0.682***	3.552***	2.467***	0.089***	0.927***	3.200***
	(0.010)	(0.107)	(0.089)	(0.006)	(0.092)	(0.014)
N	4,071	2,742	4,071	4,071	4,071	4.071

Standard errors are clustered at the individual level and shown in parentheses. $*, **, ***$ denote significance at the 10, 5, and 1 percent level, respectively.

stage procedure to achieve convergence. We first estimated (3.2) alone, obtaining estimates of σ_u, and then imposed this in the second stage when estimating the remaining parameters. This was necessary because otherwise our estimator did not converge. The reported standard errors do not account for the first stage estimation error. Table 3.8 contains the results.

As a further robustness check, it is interesting to estimate the difference in the respective expected outcome between persons with reported values of z_i slightly above zero and slightly below zero. Battistin et al. (2009) show that under the assumption that at least some persons report their income accurately these estimates are lower bounds on the magnitude of the numerator in (3.1). Moreover, and more importantly, they also show that then the sign is equal to the sign of the local average treatment effect. For this we perform separate local linear regressions to the left and to the right of 0, using Silverman's rule-of-thumb (ROT) bandwidth, of the respective outcome on z_i. Table 3.9 reports the results. In line with our baseline estimates it shows that private insurance has a big effect on the number of doctor visits.[20]

Table 3.9: Local linear regression estimates of the discontinuity at the threshold

	ROT	half ROT	twice ROT
At least 1 doctor visit	0.021	0.021	0.006
	(0.027)	(0.034)	(0.018)
Doctor visits given at least 1 visit	-0.120	-0.230	-0.162
	(0.214)	(0.318)	(0.147)
Doctor visits	-0.033	-0.105	-0.059
	(0.154)	(0.235)	(0.101)
At least 1 night in hosp.	-0.002	0.007	-0.001
	(0.010)	(0.014)	(0.007)
Nights in hospital	0.048	0.073	0.171
	(0.140)	(0.232)	(0.115)
Self-assessed health	-0.002	-0.025	0.017
	(0.035)	(0.055)	(0.029)

Estimates of the discontinuity that were obtained using Silverman's rule-of-thumb (ROT) bandwidth, as well as half and twice that bandwidth. We use a different bandwidth to the left and to the right of the discontinuity, respectively. Standard errors are bootstrapped, clustered at the individual level, and shown in parentheses. Sample consists of dependent employees for whom eligibility to opt out of the public health insurance system is exclusively determined by income and who earn between $15,000$ Euros less and $25,000$ Euros more than the compulsory insurance threshold.

Our baseline specification in this chapter imposes that the health outcome does not depend on income, i.e., $\beta_0 = \beta_1 = 0$. Table 3.10 contains results for two more general specifications. In the upper panel, it is still imposed that $\beta_0 = \beta_1$, but we allow the expected outcomes to depend on income. In the lower panel, we additionally allow for $\beta_0 \neq \beta_1$.[21] Here, we do not find a significant effect of private health insurance on self-assessed health when we use the more general specifications, and the effects on hospital nights are found to be positive rather than negative. However, in the other robustness checks we report below the results are very similar to the results presented in Table 3.6.

We have mentioned that there are two measures of monthly income for the same year. This allows us to conduct two additional robustness checks. First, the availability of the two measures allows us, under additional assumptions, to directly estimate the variance of the measurement error in (yearly) income. This result is due to Kotlarski's Theorem.[22] The additional assumptions are that (1) the measurement errors of the two measurements are mutually independent and independent of true income, and that (2) there is only measurement error in monthly income and not in the additional

[20] Notably, the size of the discontinuity that is estimated here is not the local average treatment effect.

[21] The results in that table were obtained using the two-stage procedure described before.

[22] It follows from Kotlarski's Theorem (Kotlarski, 1967) that if $x_1 = x^* + u_1$, and $x_2 = x^* + u_2$, with (x^*, x_1, x_2) being statistically independent of each other, and either $\mathrm{E}(x_1) = \mathrm{E}(x_2) = 0$ or $\mathrm{E}(x^*) = 0$, then the distributions of x^*, u_1, and u_2 can be identified from the joint distribution of (x_1, x_2).

payments. In particular, the variance of the measurement error in the measure for monthly income is then given by the variance of the used monthly income measure minus the covariance between the two measures, and the variance in yearly income is then $12^2 = 144$ times the variance in monthly income. This gives an estimate of $\sigma_u = 0.155$, which is considerably lower than the estimate we obtain in our main analysis, suggesting that actually extra payments are also prone to measurement error. Nevertheless, using this estimate, we obtain results that are still very similar to our baseline results. They are presented in Table 3.11.

The second way in which we exploit the availability of the two income measures is as follows. We only include individuals in our estimation sample for whom the difference between both monthly income measures is not bigger than 1,500, 1,000 and 500 Euros, respectively. Results are reported in Table 3.12. Again, the results are very similar to those in the main analysis.

The last robustness check concerns the appropriateness of the functional form assumptions we make. In our main analysis we have included individuals whose income is between $15,000$ Euros less and $25,000$ Euros more than the compulsory insurance threshold. We now re-estimate our model for two narrower "windows." Results are reported in Table 3.13. For doctor visits the results remain the same, for nights spent in a hospital and self-assessed health the magnitude changes, but the signs remain the same so that the results are still qualitatively the similar.[23]

3.8 Conclusions

In this chapter we estimate the effect of private health insurance on the number of doctor visits, the number of nights spent in a hospital, and self-assessed health in Germany. Variation in income around the compulsory insurance threshold generates a natural experiment which allows us to control for selection into private insurance and estimate respective average treatment effects for persons who buy private insurance once they become eligible by earning enough.

We show that it is important to account for measurement error in income and suggest a way to do so. We find a significant negative effect of private insurance on the number of doctor visits for those persons who see the doctor at least once. At the same time, we find no effect of private health insurance on the number of nights spent in a hospital, and a positive effect on self-assessed health. This suggests that private health insurance either has a positive effect on investment in prevention, because of the monetary incentives provided to the insured, or that privately insured patients receive more intense or better treatment each time they visit a doctor.

[23] Again, these results were obtained using a two-stage procedure in which we estimate σ_u in a first stage and then impose it in the second.

Table 3.10: More general specifications

	(1) At least one doctor visits	(2) Doctor visits for subsample	(3) Doctor visits	(4) At least one night in hospital	(5) Nights in hospital	(6) Self-assessed health
Δ^{LATE}	0.011	-5.115***	-2.691***	0.013	0.708	-0.345
	(0.144)	(1.550)	(0.956)	(0.073)	(1.082)	(0.258)
Baseline outcome	0.600***	3.407***	2.047***	0.070***	0.673***	3.663***
	(0.010)	(0.094)	(0.066)	(0.004)	(0.062)	(0.018)
(Gross inc. - thresh.)/10000	-0.006	0.086	0.035	-0.005	-0.111**	0.049***
	(0.008)	(0.080)	(0.056)	(0.004)	(0.055)	(0.015)
N	24,203	14,579	24,203	24,203	24,203	24,203
Δ^{LATE}	0.003	-5.064***	-2.630***	0.007	0.168**	-0.339
	(0.142)	(1.551)	(0.957)	(0.073)	(0.067)	(0.256)
Baseline outcome	0.602***	3.363***	2.014***	0.068***	0.922***	3.661***
	(0.009)	(0.112)	(0.077)	(0.006)	(0.194)	(0.017)
(Gross inc. - thresh.)/10000	-0.005	0.030	-0.008	-0.005	0.074	0.046***
if < 0	(0.007)	(0.112)	(0.077)	(0.005)	(0.165)	(0.017)
(Gross inc. - thresh.)/10000	-0.022	0.009	-0.037	0.017	-1.534	0.031
if ≥ 0	(0.062)	(0.135)	(0.109)	(0.028)	(1.036)	(0.061)
N	24,203	14,579	24,203	24,203	24,203	24,203

Standard errors are clustered at the individual level and shown in parentheses. $*, **, ***$ denote significance at the 10, 5, and 1 percent level, respectively. Sample consists of dependent employees for whom eligibility to opt out of the public health insurance system is exclusively determined by income and who earn between $15,000$ Euros less and $25,000$ Euros more than the compulsory insurance threshold.

Table 3.11: Baseline specification imposing the Kotlarski estimate of $\sigma_u = 0.155$

	(1) At least one doctor visits	(2) Doctor visits for subsample	(3) Doctor visits	(4) At least one night in hospital	(5) Nights in hospital	(6) Self-assessed health
Δ^{LATE}	-0.079	-3.237***	-1.979***	-0.064*	-0.861	0.454***
	(0.076)	(0.803)	(0.514)	(0.035)	(0.556)	(0.158)
Baseline outcome	0.605***	3.294***	1.993***	0.074***	0.768***	3.616***
	(0.005)	(0.051)	(0.037)	(0.002)	(0.036)	(0.011)

Standard errors are clustered at the individual level and shown in parentheses. $*, **, ***$ denote significance at the 10, 5, and 1 percent level, respectively. Sample consists of dependent employees for whom eligibility to opt out of the public health insurance system is exclusively determined by income and who earn between $15,000$ Euros less and $25,000$ Euros more than the compulsory insurance threshold.

Table 3.12: Baseline specification for subsamples selected according to difference between the two income reports

	(1) At least one doctor visits	(2) Doctor visits for subsample	(3) Doctor visits	(4) At least one night in hospital	(5) Nights in hospital	(6) Self-assessed health
			difference at most 1,500			
Δ^{LATE}	-0.152*	-2.884***	-2.038***	-0.066*	-1.178**	0.358**
	(0.085)	(0.859)	(0.582)	(0.039)	(0.598)	(0.172)
Baseline outcome	0.615***	3.278***	2.015***	0.075***	0.788***	3.603***
	(0.006)	(0.056)	(0.041)	(0.003)	(0.041)	(0.012)
N	20,276	12,341	20,276	20,276	20,276	20,276
			difference at most 1,000			
Δ^{LATE}	-0.117	-2.706***	-1.847***	-0.046	-0.844	0.341**
	(0.085)	(0.837)	(0.575)	(0.039)	(0.600)	(0.173)
Baseline outcome	0.616***	3.282***	2.021***	0.075***	0.778***	3.601***
	(0.006)	(0.056)	(0.042)	(0.003)	(0.041)	(0.012)
N	19,637	12,004	19,637	19,637	19,637	19,637
			difference at most 500			
Δ^{LATE}	-0.097	-2.133***	-1.497***	-0.049	-0.502	0.318*
	(0.081)	(0.787)	(0.543)	(0.037)	(0.604)	(0.168)
Baseline outcome	0.618***	3.278***	2.026***	0.073***	0.755***	3.597***
	(0.006)	(0.060)	(0.044)	(0.003)	(0.042)	(0.013)
N	17,141	10,527	17,141	17,141	17,141	17,141

Standard errors are clustered at the individual level and shown in parentheses. $*, **, ***$ denote significance at the 10, 5, and 1 percent level, respectively. Sample consists of dependent employees for whom eligibility to opt out of the public health insurance system is exclusively determined by income and who earn between $15,000$ Euros less and $25,000$ Euros more than the compulsory insurance threshold.

Table 3.13: Baseline specification for narrower samples according to reported income

	(1) At least one doctor visits	(2) Doctor visits for subsample	(3) Doctor visits	(4) At least one night in hospital	(5) Nights in hospital	(6) Self-assessed health
	income between 10,000 below and 15,000 above threshold					
Δ^{LATE}	-0.042	-3.218***	-1.930***	-0.049	-0.424	0.312
	(0.103)	(1.118)	(0.686)	(0.050)	(0.863)	(0.200)
Baseline outcome	0.602***	3.245***	1.960***	0.074***	0.747***	3.623***
	(0.007)	(0.068)	(0.049)	(0.003)	(0.050)	(0.014)
N	15,614	9,373	15,614	15,614	15,614	15,614
	income between 8,000 below 10,000 above threshold					
Δ^{LATE}	-0.068	-3.713**	-2.878***	-0.008	-0.115	0.139
	(0.152)	(1.453)	(0.986)	(0.076)	(1.211)	(0.266)
Baseline outcome	0.601***	3.216***	1.943***	0.073***	0.693***	3.636***
	(0.007)	(0.077)	(0.054)	(0.003)	(0.049)	(0.014)
N	11,746	7,041	11,746	11,746	11,746	11,746

Standard errors are clustered at the individual level and shown in parentheses. $*, **, ***$ denote significance at the 10, 5, and 1 percent level, respectively. Sample consists of dependent employees for whom eligibility to opt out of the public health insurance system is exclusively determined by income and who earn between $15,000$ Euros less and $25,000$ Euros more than the compulsory insurance threshold.

3.A Derivations

In this appendix we derive an expression for $\mathrm{E}(p_i|z_i^*) = \Pr(p_i = 1|z_i^*)$. Recall that $z_i^* = z_i - u_i$, where u_i is normally distributed with mean 0 and variance σ_u^2, statistically independent of z_i, p_i and the potential outcomes. For $z_i^* < 0$ we have that $\mathrm{E}(p_i|z_i^*) = 0$ by definition. For $z_i^* \geq 0$ we specify $\mathrm{E}(p_i|z_i^*)$ to be a linear function in z_i^*, a linear probability model. That is,

$$\mathrm{E}(p_i|z_i^*) = \begin{cases} 0 & \text{if } z_i^* < 0 \\ \alpha + \beta z_i^* & \text{if } z_i^* \geq 0. \end{cases}$$

By the law of total probability,

$$\mathrm{E}(p_i|z_i) = \Pr(z_i^* < 0|z_i) \cdot 0 + \Pr(z_i^* \geq 0|z_i) \cdot \mathrm{E}(p_i|z_i, z_i^* \geq 0).$$

The assumptions about the measurement error imply that this is equivalent to

$$\mathrm{E}(p_i|z_i) = \Pr(u_i \leq z_i) \cdot (\alpha + \beta\, \mathrm{E}(z_i - u_i|z_i, u_i \leq z_i)). \tag{3.5}$$

Recall that if v is standard normally distributed then $\mathrm{E}(v|v < c) = -\phi(c)/\Phi(c)$, which is known as the inverse Mills ratio, where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the standard normal cumulative distribution function and the probability density function, respectively. Using this, (3.5) can be rewritten as

$$\mathrm{E}(p_i|z_i) = \Phi\left(\frac{z_i}{\sigma_u}\right) \cdot \left(\alpha + \beta z_i + \beta\sigma_u \frac{\phi\left(\frac{z_i}{\sigma_u}\right)}{\Phi\left(\frac{z_i}{\sigma_u}\right)}\right).$$

Chapter 4

Is there empirical evidence for decreasing returns to scale in a health capital model?

4.1 Introduction

The canonical model of the demand for health and health investment (e.g., medical care) arises from Grossman (1972a, 1972b, 2000). In Grossman's human capital framework persons invest in health (e.g., invest time and consume medical goods and services) for the consumption benefits (health provides utility) as well as production benefits (healthy persons have higher earnings) that good health provides. The model provides a conceptual framework for interpretation of the demand for health and medical care in relation to a person's resource constraints, preferences, and consumption needs over the life cycle.

While Grossman's model has great theoretical and intuitive appeal and has led to a rich body of literature and many useful insights in health economics, it also has several limitations. For example, (i) in empirical work it is generally found that health and the demand for medical care are negatively related, whereas Grossman's model appears to predict a positive relationship (Wagstaff, 1986a; Zweifel & Breyer, 1997; Galama & Kapteyn, 2011); (ii) empirically, health declines faster for people with lower socio-economic status, and the model does not predict this (Case & Deaton, 2005); and (iii) even though the model assumes forward looking rational agents, the model solutions do not depend on past or future values of exogenous variables or initial health and wealth (e.g., Usher, 1976); for example, structural and reduced form equations for health depend only on present time conditions, such as the current wage rate and current prices, see (42), (45), and (47) in Grossman (2000). The lack of history in the model's solution implies that, whatever the initial health state, health immediately jumps to its equilibrium path.

A particularly devastating criticism appears to be the claim of Ehrlich and Chuma (1990) and Galama (2011) that the model does not have a unique solution, and that in particular the equilibrium path typically used as the model's solution is in fact not a correct solution. This would invalidate many theoretical and empirical analyses based on the model. However, this claim has been disputed by Ried (1998) and Grossman (1998, 2000). Grossman (2000) also provides a review and rebuttal of some of the other limitations mentioned. It is not the aim of this chapter to revisit this theoretical debate and we refer the interested reader to these authors.

Despite the limitations, theoretical extensions and competing economic models are still relatively few. Promising adaptations of the model are the models of Ehrlich and Chuma (1990) and Galama (2011), who have extended the Grossman model to include a health production process that is characterized by decreasing returns to scale (DRTS), whereas the standard model assumes a linear health production function with constant returns to scale (CRTS). Substantively, DRTS is appealing, because great improvements in health can be made with low levels of health investment (e.g., improving sanitation), whereas at high levels of health investment, very expensive treatments often provide only a relatively small improvement in health (e.g., Perlroth, Goldman, & Garber, 2010), although this may still be economically valuable (e.g., Goldman et al., 2010). It is generally understood that health production is subject to the law of diminishing returns (e.g., Wagstaff, 1986b). A model incorporating decreasing returns may thus provide a more realistic representation of real-world health production processes. For decreasing returns, the marginal cost of investment is found to be an increasing function of investment: the marginal cost of investment is higher for higher levels of investment, because the higher the level of investment, the smaller the health gain. As a result, healthy and high socioeconomic status persons face different effective health costs.

In addition to these direct substantive advantages, introducing decreasing returns also removes the limitations of the Grossman model mentioned above (Galama, 2011): The model with DRTS predicts (i) a negative correlation between health investment and health; (ii) that the wealthy and educated live longer and experience slower declines in health; and (iii) that current health status is a function of the initial level of health and the histories of prior health investments made. Furthermore, the DRTS model predicts the empirical stylized fact that health investment rapidly increases near the end of life and that length of life is finite. Finally, because the Hamiltonian of the DRTS model is strictly concave and the first-order condition for health investment is monotonically increasing in health investment, it is uncontroversial that the model solution for DRTS is finite and unique.

Empirical tests of the health production literature have thus far been based on the equilibrium equation derived under the assumption of a linear health production process (e.g., Grossman, 1972b; Wagstaff, 1986a). In this chapter we test the predictions of a theory of health capital with decreasing returns to scale in health production. To this end, we employ an equation for health investment that was derived by Galama (2011). We estimate this equation using the Panel Study of Income Dynamics (PSID) and contrast our findings with those of a relatively small existing empirical literature.

Our contribution is as follows. First, to the best of our knowledge, no prior attempts have been made to empirically confront the predictions of a theory of health capital with decreasing returns to scale. Second, we carefully account for the endogenous nature of health in the demand for medical care. Only a few papers in the empirical literature have estimated direct relationships between medical care and health, and none of the papers that have tested the predictions of health capital theory have attempted to address the inherent endogeneity of health.

We proceed as follows. In section 4.2 we give a concise overview of prior empirical work. In section 4.3, we describe the structural model on which our estimates are based. Section 4.4 describes our empirical strategy. Section 4.5 describes the data we use and some measurement issues. Section 4.6 presents the estimation results and section 4.7 concludes.

4.2 Prior empirical tests of health capital theory

In this section, we describe previous empirical results for the Grossman model and its extensions. Reduced form models have been estimated by Grossman (1972a), Van de Ven and Van der Gaag (1982), Wagstaff (1986a, 1993), Leu and Doppmann (1986), Leu and Gerfin (1992), Van Doorslaer (1987), Erbsland, Ried, and Ulrich (1995), Nocera and Zweifel (1998), Gerdtham, Johannesson, Lundberg, and Isacson (1999), and Gerdtham and Johannesson (1999). These papers use a large variety of methodologies and data from diverse cultural and institutional environments. Despite this, the studies are broadly in agreement with one another and in line with the predictions of the Grossman model. Health is found to increase with income (wages, lifetime earnings) and education, and decrease with age, the price of medical goods and services, and with environmental factors such as physically and mentally demanding work environments, manual labor, and psychological stress factors. In addition, these studies find that health is better among persons who participate in sports and have healthy eating and sleeping habits, and is lower for persons who are overweight and who smoke. Singles have worse health than married persons and females have worse health than males. Furthermore, moderate alcohol consumption is found to have either a positive or a negligible association with health.

In contrast to the relative abundance of estimates of reduced form equations, few studies have estimated structural equations.[1] The only ones we are aware of are Van de Ven and Van der Gaag (1982), Wagstaff (1986a), Erbsland et al. (1995), and Nocera and Zweifel (1998). As is the case with the relation we estimate, the structural equations for health investment contain health as an explanatory variable, whereas the reduced form equations do not, because the endogenous variable health is substituted out of the equation. The structural relations also provide a prediction for the sign of the relationship between health investment and health.

While the reduced form estimates are generally in agreement with the predictions of the Grossman model, this is not true for estimates of structural equations. Van de Ven and Van der Gaag (1982), Wagstaff (1986a), and Erbsland et al. (1995) find statistically significant negative relations between health investment and health. This contrasts sharply with the theory, which predicts that this relationship is positive under the assumption of a linear health production process as is widely utilized in the

[1] The health capital literature uses the terminology *structural equation* (or relation) for an equation that contains some endogenous explanatory variables. While we follow this terminology, it must be noted that these structural equations are often derived equations and not the original model equations that define the structural model. Furthermore, the coefficients in these structural equations are generally functions of the structural parameters, but not structural parameters themselves.

health production literature; see, for example, (13) in Wagstaff (1986a). A partial exception is Nocera and Zweifel (1998), who report a negative but insignificant relation between health investment and health for their sample I, but a positive and statistically significant relation for their sample II. However, the significance of the latter is likely inflated by the use of deterministically interpolated data in sample II. Thus, taken together, the evidence from these studies seems to be in the direction of a negative relation, as predicted by the DRTS model, as opposed to the prediction of a positive relation in the equilibrium equation of the CRTS model.

An important limitation of all of these studies is that they do not account for the endogeneity of health. Health is an endogenous variable, because health and health investment are joint outcomes of the lifetime utility maximization problem. Therefore, in addition to estimating regression models that take health as exogenous, we will estimate equations that account for the endogeneity of health through instrumental variables methods.

4.3 A health capital model

Our approach is based on the theoretical model presented in Galama (2011), which is an extension of the Grossman model in that it allows for a decreasing returns to scale health production process. Here we highlight the main features of the model. For consistency and comparability with prior work we adopt the usual assumptions made in this literature, mostly those made by Grossman (1972a, 1972b, 2000); Cropper (1977) and Wagstaff (1986a). Our review of the literature shows that the assumptions made by these authors are commonly made in the Grossman literature. Whenever we deviate from the literature, we do so in order to relax assumptions rather than impose additional restrictions. We discuss the implications of the important assumptions in some detail. Many assumptions in the literature are made for convenience, generally to obtain linear expressions for empirical testing. Thus, the major assumption that the extant literature and this work rely on is that such linear expressions are reasonable approximations of the model. The assumptions we adopt from the literature in this section mainly serve to contrast predictions between our theoretical results and those of the larger literature. However, in our empirical specification, we do not impose the sign restrictions that are predicted by the theoretical model.

Persons are assumed to derive utility from consumption C_t and health H_t. The *period utility* or *instantaneous utility* is denoted by $U(C_t, H_t)$. Persons live for T periods, starting with period 0, and they maximize their lifetime utility V_0, which is additively separable, and discount future utility at a per-period rate of β. Thus, lifetime utility takes the form

$$V_0 = \sum_{t=0}^{T-1} \left(\frac{1}{1+\beta}\right)^t U(C_t, H_t).$$

We use the isoelastic utility function

$$U(C_t, H_t) = \frac{1}{1-\rho}\left(C_t^{\zeta} H_t^{1-\zeta}\right)^{1-\rho},$$

where ζ is the relative "share" of consumption versus health in the utility function ($0 \le \zeta \le 1$). This utility function is also known as the *power utility* function and as the *constant relative risk aversion* (CRRA) utility function, where $\rho > 0$ is the *coefficient of relative risk aversion*. However, because our model is deterministic, a person does not face any risk. The parameter ρ thus reflects preference for utility smoothing across time, rather than risk aversion. The isoelastic utility function is arguably the most widely used utility function in intertemporal utility maximization problems (e.g., Hansen & Singleton, 1982; Attanazio & Weber, 1989; Rust & Phelan, 1997; Keane, 2011). It is increasing and strictly concave, and the parameter ρ allows a wide range of curvature, which translates into flexibility regarding the amount of utility smoothing across time. See Wakker (2008) for an extensive discussion of its properties and Chiappori and Paiella (2011) for empirical support for this functional form.

When utility depends on multiple components (typically consumption and leisure in labor supply models, consumption and health in our case), the period utility can be specified as additive or multiplicative in these components (see, e.g., MaCurdy, 1981, pp. 1064–1065). Our multiplicative specification, which most closely follows French (2005), is able to account for the observation that the marginal utility of consumption declines as health deteriorates (e.g., Finkelstein, Luttmer, & Notowidigdo, 2008). This would rule out the strongly separable (additive) functional form for the utility function typically employed in the literature (e.g., Wagstaff, 1986a), where the marginal utility of consumption is independent of health. However, it turns out that this is not essential for the empirical equation we estimate, in which the marginal utility of consumption does not enter explicitly.

Initial health H_0 is given, and health evolves according to the dynamic transition equation

$$H_{t+1} - H_t = I_t^{\alpha} - d_t H_t, \tag{4.1}$$

where d_t is an age-specific deterioration rate (biological aging rate) and I_t is investment in health. The parameter α reflects the returns to scale. The literature, starting with Grossman (1972a, 1972b), typically assumes that $\alpha = 1$, which, combined with the investment function (4.2) below, amounts to constant returns to scale. By introducing the parameter α, we relax this restriction, allowing for decreasing returns to scale ($0 < \alpha < 1$). Note, however, that our solution does not apply to $\alpha = 1$, which illustrates the special nature of constant returns to scale.

We adopt the next several assumptions from the Grossman literature to allow for comparability of results and to arrive at relations that can eventually be log-linearized. Investment in health I_t is assumed to be produced by combining time inputs $\tau_{I,t}$ with goods and services purchased in the market m_t (e.g., medical care), according to a Cobb-Douglas production function

$$I_t = \mu_{I,t}\, m_t^{1-k_I}\, \tau_{I,t}^{k_I}, \tag{4.2}$$

where $\mu_{I,t}$ is an efficiency factor and $1 - k_I$ and k_I are the elasticities of investment in health with respect to goods and services purchased in the market and with respect to time inputs, respectively. Effectively this assumption allows one to log-linearize investment I_t into its components, goods and services m_t and time inputs $\tau_{I,t}$.

We assume that the more educated are more efficient consumers and producers of health investment. This is based on the interpretation of education as a productivity factor in time inputs and in identifying and seeking effective care (Grossman, 1972a, 2000). We adopt the usual relation

$$\mu_{I,t} = \mu_I^* \exp(\rho_I E), \tag{4.3}$$

where E is the level of education (e.g., years of schooling) and ρ_I is the parameter indicating the efficiency effect of education. Note that we make the standard assumption that education is given and constant across time, and thus in this specification, $\mu_{I,t}$ does not depend on t. This assumption is most applicable to people who have completed their schooling (adults).

We follow Cropper (1977) and Wagstaff (1986a) and assume that the biological aging rate to be of the form

$$d_t = \exp(\lambda_0 + \lambda_1 t + \lambda_2 \mathrm{EC}_t), \tag{4.4}$$

where EC_t is a vector of environmental variables (working and living conditions, hazardous environment, etc.) that affect the biological aging rate.

For consumption, we specify a similar production function as for health investment. Thus, consumption C_t is assumed to be produced by combining time inputs $\tau_{C,t}$ (leisure) with goods and services purchased in the market X_t, according to a Cobb-Douglas production function

$$C_t = \mu_{C,t} X_t^{1-k_C} \tau_{C,t}^{k_C}, \tag{4.5}$$

where $\mu_{C,t}$ is an efficiency factor and $1 - k_C$ and k_C are the elasticities of consumption with respect to goods and services purchased in the market and with respect to time inputs, respectively. For the efficiency factor, we use an analogous specification as for the health investment efficiency factor

$$\mu_{C,t} = \mu_C^* \exp(\rho_C E), \tag{4.6}$$

with ρ_C capturing the effect of education on consumption efficiency.

In addition to the *consumption benefit* of health through its inclusion in the utility function, health also has a *production benefit* through its reduction of sick time. Sick time is assumed to be a power law in health

$$s_t = \tau_{\mathrm{tot}} \left(\frac{H_t}{H_{\min}} \right)^{-\gamma}, \tag{4.7}$$

where τ_{tot} is the total time budget (the total number of hours in a period). It is assumed that the parameter γ is positive, so that sick time decreases with health. This choice of functional form for sick time has the desirable properties $\lim_{H_t \uparrow \infty} s_t = 0$ and $\lim_{H_t \downarrow H_{\min}} s_t = \tau_{\text{tot}}$, that is, with infinite health, a person is never sick, and near the minimum possible level of health, a person is always sick. The specifications (4.2) – (4.7) are employed in the literature in order to arrive at expressions that can eventually be log-linearized, which we will also do in section 4.3.1.

In the Grossman model, sick time governs the relation between income and health. Persons work $\tau_{w,t}$ hours, which is limited by the time budget constraint

$$\tau_{w,t} + \tau_{I,t} + \tau_{C,t} + s_t = \tau_{\text{tot}}.$$

Thus, increased sick time due to worse health must be offset by reduced work time, time spent investing in health, or time devoted to consumption (leisure). Income in period t equals the hourly wage rate w_t times the number of work hours $\tau_{w,t}$

$$Y_t = w_t\, \tau_{w,t}.$$

We assume wage rates are exogenously given and do not depend on work hours. However, they may depend on other exogenous variables, such as education. Note that the current version of our model focuses on workers and it limits their income to earnings.

The final component of the structural model is assets. Initial assets A_0 are given, and assets evolve according to the dynamic transition equation

$$A_{t+1} - A_t = \delta\, A_t + Y_t - p_{X,t}\, X_t - p_{m,t}\, m_t, \tag{4.8}$$

where δ is the rate of return on capital and $p_{X,t}$ and $p_{m,t}$ are the prices of consumption goods and services, and medical goods and services, respectively. Persons do not face a borrowing constraint in our model, but they are subject to the lifetime budget constraint $A_T = A^*$, where A^* is a given constant.

Persons die when health reaches the minimum health level $H_t = H_{\min}$. Length of life T is endogenous and is determined by maximizing lifetime utility with respect to T. For the purpose of this chapter, endogenous length of life T affects the shadow price of initial wealth q_0^A, which acts as a person-specific constant.

4.3.1 Structural relation between medical care and health

In this chapter, we study the relationship between the demand for medical goods and services m_t and health H_t. Galama (2011) derived a structural equation for this relationship for the model presented above

$$b_{1t}\, m_t^{1-\alpha} - (1-\alpha) m_t^{1-\alpha} \tilde{m}_t = b_{2t}\, H_t^{-\rho\phi} + b_{3t}\, H_t^{-(1+\gamma)}, \tag{4.9}$$

with the notation $\tilde{x}_t \equiv 1 - x_{t-1}/x_t$ for a variable x_t, and the following functions

$$b_{1t} \equiv \exp(\lambda_0 + \lambda_1 t + \lambda_2 \mathrm{EC}_t) + \delta - (1 - \alpha k_I)\tilde{p}_{m,t} - \alpha k_I \tilde{w}_t;$$

$$b_{2t} \equiv b_2^* \left(q_0^A\right)^{-\phi} e^{[(\phi-1)\rho_C + \alpha\rho_I]E} p_{m,t}^{-(1-\alpha k_I)} w_t^{-[k_C(\phi-1)+\alpha k_I]} p_{X,t}^{-(1-k_C)(\phi-1)} \left(\frac{1+\beta}{1+\delta}\right)^{-\phi t};$$

$$b_{3t} \equiv b_3^* e^{\alpha\rho_I E} p_{m,t}^{-(1-\alpha k_I)} w_t^{1-\alpha k_I};$$

with the following constants

$$b_2^* \equiv (1-\zeta)\zeta^{(1-\rho)(\rho\phi-1)/\rho} \alpha k_I^{\alpha k_I} (1-k_I)^{1-\alpha k_I} (\mu_I^*)^\alpha \left[\mu_C^* k_C^{k_C} (1-k_C)^{1-k_C}\right]^{\phi-1};$$

$$b_3^* \equiv \gamma \tau_{\mathrm{tot}} H_{\mathrm{min}}^\gamma \alpha k_I^{\alpha k_I} (1-k_I)^{1-\alpha k_I} (\mu_I^*)^\alpha,$$

$$\phi \equiv \frac{1}{1-\zeta+\rho\zeta}.$$

As mentioned above, the constant q_0^A is the shadow price of initial wealth. It is part of the *adjoint function* (analogous to a Lagrange multiplier; see, e.g., Varaiya, 1998, pp. 78–80) of the assets transition equation and it emerges in the optimization process.

4.4 From theoretical to empirical model

Ideally, we would like to estimate the structural relation (4.9) between health investment and health directly. However, the equation is highly nonlinear in both variables and parameters, which implies that it is difficult to establish whether the structural parameters are identified from this equation. Given the results for our linearized equation below, we suspect that they may not be, but we have been unable to prove this. Unlike for the linearized equation below, it is not clear whether and how it would be possible to obtain useful partial results if the parameters are not identified. Furthermore, assuming identification, estimating the parameters from this equation is computationally difficult, because it may lead to convergence problems and local optima. Another drawback of estimating the nonlinear equation is that it is less straightforward to compare our results with empirical results found in the literature, which are based on linearized equations. Therefore, we leave estimating the nonlinear equation for future research and follow the literature by deriving linearized approximations to the equation of interest and estimating these resulting linear equations.

4.4.1 Pure investment and pure consumption

Analytical solutions for the Grossman model are usually based on two sub-models: the *pure investment* model and the *pure consumption* model. In the pure investment model, health does not provide utility and hence $\zeta = 1$ and $b_{2t} = 0$, whereas in the pure consumption model, health does not provide a production benefit and $b_{3t} = 0$.

We proceed by assuming that $(1-\alpha)\tilde{m}_t$ is negligible compared to b_{1t}. This is the case if changes in wages and prices are small and changes in health investment are smaller than the biological aging rate plus the rate of return to capital. Lifecycle models predict considerable consumption smoothing, and this carries over to health investment, so this is likely consistent with the predictions of the structural model. It is, however, difficult to assess the validity of this assumption empirically, because the biological aging rate depends on the scale of health, which is not well-defined. Some tentative computations suggest that it is likely satisfied for out-of-pocket (OOP) and total medical expenditures, but maybe not for hospital nights. Under this assumption, the structural relation simplifies and we can obtain an approximately linear structural relation for the demand for health investment goods and services m_t in the pure investment and pure consumption models. This also facilitates comparisons with the results in the literature.

For the pure investment model, $\zeta = 1$ and $b_{2t} = 0$, and with $(1-\alpha)\tilde{m}_t$ negligible compared to b_{1t}, (4.9) reduces to

$$\begin{aligned}\log m_t \approx\ & \frac{\log b_3^*}{1-\alpha} - \frac{1+\gamma}{1-\alpha}\log H_t + \frac{\alpha\rho_I}{1-\alpha}E - \frac{1-\alpha k_I}{1-\alpha}\log p_{m,t} + \frac{1-\alpha k_I}{1-\alpha}\log w_t \\ & - \frac{\lambda_0}{1-\alpha} - \frac{\lambda_1}{1-\alpha}t - \frac{\lambda_2}{1-\alpha}\mathrm{EC}_t + R_t \end{aligned} \tag{4.10}$$

where

$$R_t = -\frac{1}{1-\alpha}\log\left(1 + \frac{\delta - (1-\alpha k_I)\tilde{p}_{m,t} - \alpha k_I \tilde{w}_t}{\exp(\lambda_0 + \lambda_1 t + \lambda_2 \mathrm{EC}_t)}\right).$$

Analogously, for the pure consumption model, $b_{3t} = 0$ and with $(1-\alpha)\tilde{m}_t$ negligible compared to b_{1t}, (4.9) reduces to

$$\begin{aligned}\log m_t \approx\ & \frac{\log b_2^*}{1-\alpha} - \frac{\rho\phi}{1-\alpha}\log H_t - \frac{\phi\log q_0^A}{1-\alpha} + \frac{(\phi-1)\rho_C + \alpha\rho_I}{1-\alpha}E - \frac{1-\alpha k_I}{1-\alpha}\log p_{m,t} \\ & - \frac{k_C(\phi-1) + \alpha k_I}{1-\alpha}\log w_t - \frac{(1-k_C)(\phi-1)}{1-\alpha}\log p_{X,t} \\ & - \frac{\lambda_0}{1-\alpha} - \frac{\lambda_1}{1-\alpha}t - \frac{\lambda_2}{1-\alpha}\mathrm{EC}_t - \left[\frac{\phi}{1-\alpha}\log\left(\frac{1+\beta}{1+\delta}\right)\right]t + R_t. \end{aligned} \tag{4.11}$$

It is customary to view λ_0 as an error term (see, e.g., Wagstaff, 1986a, and Grossman, 1972a, 1972b, 2000). Furthermore, it is often assumed that R_t is either small or constant (Grossman, 1972a, 2000). This is the case if the rate of return on capital δ and changes in the wage rate w_t and the price $p_{m,t}$ are much smaller than the health deterioration rate d_t or if the rate of return to capital δ and changes in the wage rate w_t and the price $p_{m,t}$ follow the same pattern as d_t, so that their ratio is approximately constant. Alternatively, it has been assumed that R_t is approximately a linear function of age t (e.g., Wagstaff, 1986a). Neither of these assumptions about R_t is likely to hold across the lifecycle. Health deterioration is assumed to accelerate across the lifecycle (e.g., Grossman, 2000) and our functional form is consistent with this. Asset returns do not show a consistent time trend so are best thought of as constant, and similarly for price changes. Wage tends to increase at younger ages, then stay relatively constant, with a slight drop off before retirement. Thus, R_t is not constant or small across the entire lifecycle: at young ages it is approximately constant but nonzero, whereas at older ages it is closer to zero but not constant. Over short time spans, it is approximately linear or even constant, so in a panel data analysis with a few closely spaced waves the linearity assumption should work fine. In cross sections or longer panels, it is better to either include the term R_t in the model and estimate it as a nonlinear regression model, or approximate it by including a flexible function of age. We do the latter and include a quadratic in age below.

4.4.2 Model predictions

In the previous section, we derived two approximations, (4.10) and (4.11) of the structural relation between health investment and health, depending on different simplifying assumptions. Both equations are of the form

$$\log m_t = \theta_1 + \theta_2 \log H_t + \theta_3 E + \theta_4 \log w_t + \theta_5 \log p_{m,t} + \theta_6 \log p_{X,t} + \theta_7 t + \theta_8 \mathrm{EC}_t + u + R_t, \quad (4.12)$$

where u is a person-specific error term and R_t is a nonlinear residual term. As mentioned above, we approximate R_t by a quadratic in age, which subsumes the linear age term already in the model.

The coefficients in (4.12) are functions of the structural parameters, and the exact form of these functions depends on whether we assume the pure investment or the pure consumption model. For example, in the pure investment model, $\theta_6 = 0$, whereas in the pure consumption model, it is generally nonzero. We cannot fully recover the structural parameters from our econometric model. However, the structural model does provide predictions about the signs of most of the coefficients in the econometric model. In particular, according to both the pure investment and pure consumption models, the sign of the coefficient of log health is negative if there are decreasing returns to scale: persons in better health invest less in health, *ceteris paribus*. This contrasts with the literature that assumes constant returns to scale, which predicts a positive coefficient of log health. Note that we cannot simply set $\alpha = 1$ in our equation to arrive at this constant returns to scale result.

In the pure investment model, higher education leads to a higher demand for medical goods and services, whereas in the pure consumption model, the sign is ambiguous and depends on the relative efficiency gains from education for consumption and health investment. This contrasts with the usual prediction of the Grossman model (e.g., Feldstein, 1993, p. 78; Zweifel, Breyer, & Kifmann, 2009, pp. 83–84) that education unambiguously reduces this demand. Similarly, the sign of the coefficient of log wage is unambiguously positive in the pure investment model, but because generally $\phi - 1 < 0$ (assuming a coefficient of relative risk aversion larger than 1, as is typically found in the literature), the coefficient of log wage is not guaranteed to be positive in the pure consumption model. The coefficients of age and environmental factors can be either positive or negative, because these partially pick up the effect of the nonlinear residual term R_t. In the pure investment model (4.10), if we disregard R_t, the coefficient of age is negative: at older ages, health investment is lower (*ceteris paribus*, in particular, controlling for health). However, $\partial R_t/\partial t$ may be positive, so after approximating R_t with a quadratic in age, the model does not provide a clear prediction for the sign of the coefficient of age. This is also true for the pure consumption model, where the equation has an additional term in age and its sign depends on whether the discount rate β is larger than the asset return rate δ.

4.5 Data

Our analysis is based on the Panel Study of Income Dynamics, a longitudinal survey of a representative sample of persons in the U.S. and the households in which they reside. The PSID started in 1968 and was conducted annually up to 1997, after which it has been conducted biennially.

The PSID covers nearly the entire life cycle of persons; from childhood through old age. In addition, the PSID collects rich data on income, wealth, demographic characteristics, labor force participation, education level, consumption behavior (e.g., alcohol, smoking), medical expenditures, and detailed information on health (e.g., self-reported health, childhood health).

The most detailed information is collected for the household heads. For married couples, the PSID typically assigns the head status to the male. As a result female heads of household in the survey are not representative of the U.S. female population (i.e., typically not married/partnered, in a same-sex couple, or married to an incapacitated male). Therefore, we limit our analyses to male heads. Furthermore, detailed health data are available for all ages only since the 2003 wave. Therefore, the estimates presented below are based on the pooled 2003, 2005, and 2007 waves. The PSID interviewed 7,822 families in 2003, 8,002 in 2005, and 8,289 in 2007. Of all the households interviewed, 5,483 are headed by males in 2003, 5,594 in 2005 and 5,761 in 2007.

It follows from (4.1) and (4.8) that stock variables such as health are measured at the beginning of the period. However, in the data, health variables refer to the situation at the time of the interview, whereas flow variables such as medical expenditures refer to the past. Hence, to make the data consistent with the concepts in the theoretical model, we need to use medical expenditures and other flow variables measured in 2005 and 2007 in combination with stock variables measured a wave earlier (2003 and 2005, respectively). This implies that we effectively have two waves of data for the regression analyses reported below.

4.5.1 Measurement of health investment

We consider three different dependent variables. The first is out-of-pocket medical costs, which is the part of total medical expenditures that is paid for by the respondent, and this is the total of three different categories: (a) nursing home and hospital bills; (b) doctor, outpatient surgery, and dental bills; and (c) prescriptions, in-home medical care, special facilities, and other goods and services. The insurance premium is excluded from the out-of-pocket costs because this in itself does not indicate any health investment. The second dependent variable is total medical expenditures. This includes out-of-pocket medical costs and what has been covered by the insurer. Our third dependent variable is nights spent in a hospital.

Total medical expenditures conceptually comes closest to our theoretical variable, but respondents often have little information about what the insurer pays. Out-of-pocket costs may be a good proxy for true total medical expenditures. Especially if a person makes active decisions about health care utilization based on out-of-pocket costs, this may be an even better reflection of (the goods and services part of) true health investment as conceptualized in the model. A drawback of these two expenditure-based measures is that they are only measured at the household level in the PSID, whereas our model is based on a single person. Section 4.5.1 below describes the imputation method we employ to deal with this issue.

The number of hospital nights is unlikely to be an accurate indicator of our theoretical concept. It may be a measure of bad health rather than investment in health. Also, it does not capture a potentially large part of health investment that does not involve spending a night in a hospital. However, the measure is not meaningless. The model predicts that in response to bad health, persons will invest more in health, and in principle, anything that improves health is health investment. Arguably people spend nights in a hospital because that is better for their health than not doing so. The measure has a few additional advantages: (1) it is likely to be measured better (less noisy) than medical expenditures; (2) it is measured at the individual level in the PSID and thus does not require imputation; (3) it can serve as a test for the sensitivity of the regression results to different operationalizations of the dependent variable; and (4) it is a measure that is commonly used measure of health investment (e.g., Van de Ven & Van der Gaag, 1982; Wagstaff, 1986a, 1993; Erbsland et al., 1995), thus allowing us to compare our results with prior estimates in the literature.

Imputation of individual medical expenditures

As noted above, the PSID records total medical expenditures and medical out-of-pocket costs at the *household* level instead of at the individual level. However, the model is for *individuals*. Therefore, we use data from the Medical Expenditure Panel Survey (MEPS) to impute individual-level total expenditures and out-of-pocket costs for individuals in the PSID. The MEPS is a series of short panels (up to two years per household) that are nationally representative samples of the U.S. population. In addition to, among others, demographics, employment, income, and health, it contains detailed high quality information about health insurance, medical care utilization, and medical expenditures at the individual level. Part of this information is obtained from medical providers or employers. MEPS collected data on 32,681 individuals in 2003, 32,320 in 2005 and 29,370 in 2007. We use the following imputation procedure to obtain individual-level medical expenditure data for (male) heads in the PSID:

1. Define cells based on age, sex, and individual health insurance status (note that individuals within the household may have different health insurance status).

2. Based on these cells, randomly match a MEPS person to each PSID household member (hot deck imputation).

3. Use the MEPS individual's total medical expenditures and out-of-pocket costs as imputations for the individual medical expenditures and out-of-pocket costs in the PSID.

4. Rescale so that individual total medical expenditures and out-of-pocket costs sum to the observed household total medical expenditures and out-of-pocket costs in the PSID. Thus, we effectively only impute *allocation percentages*.

We repeat this procedure five times. The econometric models, discussed below, are estimated using one imputed variable at a time. Since we have five imputed variables for total medical expenditures and out-of-pocket medical expenditures, we obtain five sets of regression results, which are then combined according to the "Rubin rules" (Rubin, 1987). This takes imputation uncertainty into account in computing the standard errors and thus better reflects the sampling variability of the estimators than single imputation. This results in more random noise in the dependent variable than if the individual amounts were reported, but it does not affect the unbiasedness of the estimators and only leads to larger standard errors.

It is difficult to check the quality of the imputations, because we only impute allocation factors. We computed the within-household sums of the intermediate imputations in step 3 and compared their distributions with the distributions of the reported household-level variables. There is some variation between the five imputation samples, but taking the average of the five samples, the means of the imputed amounts are slightly (2–4 percent) higher than the reported ones in 2005 and a bit lower (7–12 percent) in 2007. For total medical expenditures, the average of the standard deviations was lower in both years (17 percent in 2005 and 33 percent in 2007), whereas for out-of-pocket costs, it was 12 percent higher in 2005 and 6 percent lower in 2007. Note, however, that after rescaling in step 4 these differences become zero by construction.

Imputation was not necessary in two instances: for single households, as the household coincides with the individual, and for individuals in households that report zero household medical expenditures (or out-of-pocket costs, respectively). There is no imputation uncertainty for these households. For 13 percent of the sample for individual total medical expenditures and for 15 percent of the sample for individual OOP, imputation was not required. (Note that the samples differ because of differences in missing values in family medical expenditures and out-of-pocket costs.)

For a small number of households—a maximum of 3 percent across waves for both medical expenditures and out-of-pocket costs—the imputed values in step 3 above were zero for all household members, while the reported family medical expenditures or out-of-pocket costs in the PSID were positive. In these cases, the procedure above does not lead to a valid allocation of the family expenditures or out-of-pocket costs, and we randomly assigned the reported family value (with equal probabilities) to one of the household members.

4.5.2 Measurement and endogeneity of health

We use self-reported health to measure health, which is most often used health variable in empirical studies and administered in virtually all surveys. It is a categorical variable, in which respondents assess their health using five categories: (1) poor; (2) fair; (3) good; (4) very good; (5) excellent. This variable is a good measure of health as judged by its predictive power for medical care utilization (e.g., Van Doorslaer, Koolman, & Jones, 2004), labor force participation (Bound, 1991), mortality (e.g., Dowd & Zajacova, 2007; Huisman, Van Lenthe, & Mackenbach, 2007; Idler & Benyamini, 1997).

In the econometric models we estimate, health is an endogenous variable, because according to our theoretical model, health and health investment are joint outcomes of the lifetime utility maximization problem. Therefore, in addition to ordinary regressions we also perform instrumental variables analyses. We use as instruments self-reported childhood (age 0 to 16) health, which uses the same five-point scale as adult self-reported health (i.e., poor, fair, good, very good, excellent), and a binary indicator for whether or not one's parents smoked during childhood.

The PSID contains more information concerning childhood health, but the two measures used can be considered strong instruments according to the first stage F-statistic and the remaining variables do not add predictive power. A large literature has established a strong correlation between early childhood conditions and later-life health, see for example Van den Berg et al. (2006, 2010). Self-reported childhood health was asked retrospectively in 2003 and 2007. We take the average if it was reported in both years. The more detailed information was collected only in the 2007 wave.

Concerns regarding whether respondents can accurately recall events that happened during childhood are understandable; answers may also be biased towards current health status. Analyses presented in Smith (2009), however, suggest that these concerns are too pessimistic and that this type of information can be usefully exploited. Another concern is that health during childhood may affect health investments through additional channels beyond health (and other variables in the model, particularly education). This could occur, for example, through higher health investments during childhood and habit formation or path dependence thereafter, possibly violating the exclusion restriction. We believe that health and education are the main channels through which childhood health is related to medical investments and therefore, we believe our IV results are valid. Moreover, even if we cannot control for all potential pathways, exploring differences between IV results and standard OLS results is still useful for an assessment of the potential severity of the endogeneity problem.

4.5.3 Exogenous covariates

Exogenous explanatory variables in the regression models include age, education, race, an indicator for marital status (treating cohabitors as married), household size, urbanicity, the wage rate, and an indicator for whether the household head has private health insurance (either provided by an employer or purchased directly).

Based on the reported years of education (YoE), persons are stratified into three different classes:

primary education (YoE $<$ 12), secondary education (12 $\leq$ YoE $<$ 16), and tertiary education (YoE $\geq$ 16). We generated three categories for race: white, black, and other.

The three dummy variables indicating the degree of urbanicity are derived from the Beale's code of urbanicity which we have aggregated as follows:

"Urban": Central or fringe counties of metropolitan areas of 1 million population or more, and counties in metropolitan areas of 250,000 to 1 million population;

"Suburban": Counties in metropolitan areas of less than 250,000, urban populations of 20,000 or more, and urban populations of less than 20,000 that are adjacent to metropolitan areas; and

"Rural": Urban populations of less than 20,000 that are not adjacent to metropolitan areas and rural populations.

PSID respondents are asked whether they are salaried (and paid by week, two weeks, month, or year) or paid by the hour. To create a measure of (hourly) wage, we assume that, for those who are salaried, they work full-time (i.e., 40 hours per week).

Feldstein (1993) discusses three measures for the price of medical goods and services: the medical care price index (MCPI), the Scitovsky index, and health insurance premiums. It is, however, doubtful whether these accurately reflect the relevant prices for our purposes. Because of these considerations and because we only have a limited number of time periods, we proxy for the price variables by including a private health insurance (including employer-provided, as opposed to public insurance such as Medicare and Medicaid) status dummy that captures price differences for medical goods and services as faced by a person. In our empirical analyses we do not control for the possible endogeneity of private insurance.

Our regression models also include the education of the household head's wife, which is measured in years. We also estimate models with year dummies to capture any cross-year price differences that are common to all persons.

4.5.4 Sample selection

If age, race, or education was missing for a respondent in one of the waves, we checked whether this information could be derived from other waves. After this "logical imputation", observations for which age or race were still missing were excluded from the sample. There were 8 such cases in 2003, 6 in 2005, and 7 in 2007. Further, the sample selection is based on four additional criteria resulting in a sample of male household heads who are younger than 65, have health insurance (public or private), are working, and have an hourly wage of at least \$4. Male heads with an hourly wage less than \$4 are dropped because these appear to be extreme cases. We also limit our sample to male heads who are working since our structural model does not contain retirement (or unemployment). Persons face very different prices for medical goods and services depending on whether or not they have health insurance. In the absence of detailed data about how those prices actually differ, we have chosen to

Table 4.1: Characteristics of the regression sample[a]

	2003	2005
Age, mean	41.1	41.2
Race:		
White (%)	71.3	72.2
Black (%)	20.9	20.5
Other (%)	7.7	7.3
Married (%)	82.0	82.8
Household size, mean	3.1	3.0
Years of education, mean	13.6	13.6
Less than high school (%)	9.7	9.8
High school or some college (%)	58.3	58.8
College degree or more (%)	32.0	31.4
Wife's years of education, mean	11.1	11.3
Urbanicity:		
Urban (%)	66.5	65.8
Suburban (%)	22.4	22.9
Rural (%)	11.1	11.3
Self-reported health, mean	3.9	3.9
Self-reported childhood health, mean	4.4	4.4
Parents smoked during childhood (%)	0.64	0.61
Sample size	2,368	2,493

[a] Intermediate sample; see Table 4.2

Table 4.2: Sample selection for regression analysis

	2003	2005
Male household heads	5,483	5,594
Health measurement sample	5,475	5,588
Not in next wave (for m_t)	600	540
Age $\geq$ 65	506	508
Uninsured	710	734
Non-working	451	427
Hourly wage $<$ \$4	16	17
Missing selection variables	561	594
Missing covariates	263	275
Intermediate sample size	2,368	2,493
Missing family total medical expenditures	975	1,065
Missing family out-of-pocket med.exp.	155	207
Missing hospital nights	8	12
Sample size total medical expenditures	1,393	1,428
Sample size out-of-pocket med.exp.	2,213	2,286
Sample size hospital nights	2,360	2,481

Note: Categories in lines 3–9 are mutually exclusive and assigned with higher rows taking precedence.

restrict our analysis to the insured only. Our sample is further limited to heads under 65 years of age since Medicare eligibility creates a very different set of circumstances for those 65 and over. Limiting the sample to those that meet the selection criteria excludes roughly 2,000 persons from each year's sample. Table 4.2 gives a breakdown of the sample selection. Summary statistics of the resulting sample are presented in Table 4.1.

4.6 Results

The outcome variables considered in this chapter are nonnegative, have a high frequency of zeros and have a right-skewed distribution. Furthermore, the number of nights spent in a hospital is an integer-valued variable. Common strategies to deal with such data structures are to estimate a two-part model or a bivariate sample selection model, and for nonnegative integer-valued variables to estimate count data models. For total medical expenditures and out-of-pocket expenditures, we estimated both Heckman-type bivariate sample selection models and two-part models. In both cases, the participation equation used a probit specification and the outcome equation was specified as a linear regression equation with the logarithm of the outcome variable as the dependent variable. The outcome equation is only estimated for observations with positive outcome values. The sample selection models explicitly estimate the correlation between the participation equation and the expenditure equation, but rely on strong assumptions about correctness of the functional form specification, homoskedasticity, and normality. The two-part models require weaker assumptions in this respect, but it is less clear to what extent they can accommodate correlation between the two equations. See Duan, Manning, Morris, and Newhouse (1983, 1984) and Hay and Olsen (1984) for an early discussion of some of these issues, and Jones (2000) for an overview of the subsequent debate. In our case, the results of these two approaches were very similar, and we only present the results from the bivariate selection models.

To account for the endogeneity of health, we have additionally estimated an instrumental variables version of the sample selection model. This estimates a system of three equations. The first two equations are the same as in the bivariate sample selection model:

$$\begin{aligned} y_1^* &= x\,\beta_1 + w'\gamma_1 + \epsilon_1, \qquad & y_1 &= 1(y_1^* > 0); \\ y_2^* &= x\,\beta_2 + w'\gamma_2 + \epsilon_2, \qquad & y_2 &= y_2^* \text{ if } y_1 = 1, \text{ and unobserved otherwise,} \end{aligned}$$

where y_1 indicates participation (positive OOP or total medical expenditures), y_2 is the outcome variable (logarithm of OOP or total medical expenditures), x is the endogenous explanatory variable (health), and w is the vector of exogenous covariates. The third equation of the IV model is the first-stage equation:

$$x = z'\pi + w'\gamma_3 + \epsilon_3,$$

where z is the vector of instrumental variables (childhood health and parental smoking during childhood). The model is completed by assuming that the error terms ϵ_1, ϵ_2, and ϵ_3 are jointly normally distributed. Allowing nonzero correlations between these error terms explicitly introduces correlations between the endogenous explanatory variable x and the error terms ϵ_1 and ϵ_2, which solves the endogeneity issue. We estimated this model with the `cmp` command in Stata (Roodman, 2009).

For the number of nights spent in a hospital we have estimated several count data models. The simplest count data model is poisson regression, which assumes that conditional on the explanatory variables x, the dependent variable follows a poisson distribution with both mean and variance equal to $\exp(x'\beta)$. However, equality of conditional mean and variance is often violated. Therefore, we have also estimated a negative binomial regression model, which is an extension of the poisson model that allows greater dispersion through an additional variance parameter. A typical characteristic of many count variables is a spike at zero that is not well reproduced by these count data regression models. Nights spent in hospital shows this quite extremely, with only 282 observations with a positive amount, out of a total sample of 4,841. To accommodate this, we also estimated a hurdle model. This is a two-part model, with the first equation being a probit for a positive amount versus a zero amount, and the second equation being a zero-truncated negative binomial model, that is, a negative binomial model conditional on the amount being positive. A more detailed discussion of the estimated models can be found in Cameron and Trivedi (2005).

Finally, we estimate a count data model that takes into account the endogeneity of health. This involves specifying models for both the dependent variable and the endogenous regressor. We assume

Table 4.3: First stage results (dependent variable: self-reported health)

	Model for Out-of-pocket costs		Model for Total medical expenditures		Model for Nights spent in a hospital	
Variable	Coeff	t-value	Coeff	t-value	Coeff	t-value
Self-reported childhood health	0.32	13.89	0.31	10.95	0.32	18.77
Parents smoked during childhood	-0.05	-1.59	-0.05	-1.32	-0.05	-1.84
Missing: self-reported childhood health[a]	0.74	3.90	0.71	1.98	0.63	1.67
Missing: parents smoked during childhood	0.13	1.93	0.06	0.68	0.12	2.08
Private insurance	-0.02	-0.32	-0.08	-0.90	-0.03	-0.53
Log wage	0.18	6.56	0.19	5.34	0.20	8.05
$a = (\text{Age} - 50)/10$	-0.15	-5.78	-0.14	-4.25	-0.16	-6.89
a^2	0.03	2.05	0.03	1.96	0.02	2.19
Secondary education	0.20	3.55	0.16	2.17	0.17	3.88
Tertiary education	0.34	5.13	0.31	3.69	0.31	5.93
Wife's education	0.02	2.80	0.03	3.07	0.02	3.27
Black	-0.14	-3.54	-0.13	-2.54	-0.14	-4.28
Other race	-0.07	-1.15	-0.04	-0.49	-0.06	-1.29
Married	-0.18	-1.41	-0.30	-1.89	-0.17	-1.66
Suburban	0.02	0.58	0.05	1.21	0.04	1.47
Rural	-0.05	-1.00	-0.01	-0.16	-0.02	-0.53
Household size	-0.01	-0.88	-0.01	-0.50	-0.02	-1.38
Constant	1.52	9.74	1.62	7.95	1.54	12.70
First-stage F statistic[a]	98.6		60.5		98.9	
N	4,499		2,821		4,841	

[a] When self-reported childhood health was missing we imputed a value of 3, corresponding to "good" childhood health. We included a binary variable indicating when the information was missing. We imputed that parents were not smoking during childhood when the information was missing. Again, we included a binary variable to indicate whether the information was missing. This is harmless as we are interested in the correlation between the instruments self-reported health.

a poisson model for the dependent variable and a linear reduced-form equation for the endogenous variable. The model is estimated using a two-step procedure as explained in Cameron and Trivedi (2009).

Table 4.3 presents the results of the first-stage regressions. As expected, better self-reported childhood health is associated with better self-reported later life health. Having parents who smoked during one's childhood is associated with a worse later life health. Note that in all cases the first stage F statistics are above the widely used threshold of 10 for separating weak and strong instruments as suggested by Staiger and Stock (1997).

Table 4.4 presents the results of the regression models for out-of-pocket medical costs. Columns 2–5 show the results for the standard bivariate sample selection model, whereas columns 6–9 show results for the IV bivariate sample selection model.

Table 4.4: Regression models for out-of-pocket costs

	Bivariate sample selection				IV bivariate sample selection[a]			
	Participation eq.		Outcome eq.		Participation eq.		Outcome eq.	
Variable	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Self-reported health	0.00	0.11	-0.12	-2.05	-0.06	-0.51	0.07	0.33
Private insurance	0.61	5.30	0.07	0.26	0.61	5.31	0.08	0.28
Log wage	0.02	0.40	0.13	1.42	0.04	0.68	0.08	0.99
$a = (\text{Age} - 50)/10$	0.28	6.32	-0.08	-1.12	0.27	5.41	-0.05	-0.66
a^2	0.01	0.56	-0.03	-0.73	0.01	0.66	-0.03	-0.87
Secondary education	0.11	1.19	-0.07	-0.35	0.12	1.24	-0.12	-0.54
Tertiary education	0.09	0.78	0.08	0.37	0.11	0.86	0.00	0.01
Wife's education	0.00	0.16	-0.01	-0.19	0.00	0.26	-0.01	-0.35
Black	-0.08	-1.25	-0.29	-2.27	-0.09	-1.34	-0.26	-2.09
Other race	-0.08	-0.81	-0.29	-1.45	-0.09	-0.87	-0.27	-1.39
Married	-0.03	-0.13	0.13	0.37	-0.04	-0.20	0.17	0.48
Suburban	-0.02	-0.44	0.16	1.61	-0.02	-0.43	0.15	1.58
Rural	0.00	-0.02	0.33	1.90	0.00	-0.03	0.33	1.91
Household size	-0.02	-0.84	-0.04	-0.91	-0.02	-0.85	-0.04	-0.87
Constant	0.15	0.69	6.29	16.74	0.32	0.78	5.77	8.70
N	4,499		3,224 – 3,247[b]		4,499		3,224 – 3,247[b]	

[a] The first stage F statistic for the IV regression is 98.6.

[b] The number of positive observations varies between imputed data sets.

The first observation from this table is that, despite the relatively large sample sizes, very few coefficients are statistically significantly different from zero. This is partly due to the imputation uncertainty, but we also interpret this as a sign that out-of-pocket medical expenditures contain a relatively large random component. For the time-varying covariates, another part of the explanation may be the large lag between observing the covariates and observing the dependent variable (2 years). We will return to the timing issue later in this section.

Examining the coefficients on self-reported health, we do not find an effect of health on having positive (vs. zero) out-of-pocket expenses in the non-IV model, but a statistically significant negative effect on the dollar amount for those who have positive out-of-pocket expenses, which replicates the general finding in the literature. Because the DRTS model predicts a negative sign and the CRTS model a positive sign, this is suggestive evidence in favor of the DRTS model. However, when we move to the IV estimates, we see that the coefficients on self-reported health are not statistically significantly different from zero and are of opposite sign for the participation and outcome equations.

Regarding the other coefficients in the model, in both the standard model and the IV model, having private insurance is positively related to positive medical out-of-pocket costs, but among those with positive out-of-pocket costs, the privately insured have no higher costs than the publicly insured. This positive relation between private insurance and out-of-pockets costs may be the results of the fact that public insurance programs, like Medicaid, pay for most of the medical expenses, whereas those with private insurance have to pay out of pocket for their own (higher quality) medical care (De Nardi, French, & Jones, 2010).

The results for age point to a positive relationship between age and out-of-pocket expenditures. As discussed above, within the pure investment model, this implies an important role of (and variation in) the residual term R_t. Within the pure consumption model, there is more ambiguity as to what the source of this positive coefficient may be. Besides the role of R_t, other explanations of the positive sign may be that we do not measure health precisely enough and thus the age coefficients pick up part of the health effect, or that there are other confounding factors that are correlated with age, such as cohort effects, although the latter would presumably point in the opposite direction.

The results for total medical expenditures are given in Table 4.5. They are largely similar to those for medical out-of-pocket costs. One exception is that the private insurance coefficients are now much smaller and not statistically significant. Another exception is that the age coefficients are now positive in the outcome equation as well, and statistically significant (at the 10 percent level in the standard model and at the 1 percent level in the IV model). The self-reported health coefficient is again negative and highly significant in the outcome equation of the standard model, but vanishes in the IV model.

In addition to the results presented here we have also performed a number of sensitivity analyses. We used household income instead of wages of the household head, and we included year dummies as covariates. Results are similar to those reported here and are omitted. We have also experimented with exploiting the panel nature of the data, estimating fixed effects logit models for the participation equation and fixed effects linear regression models for the outcome equations. For these models,

Table 4.5: Regression models for total medical expenditures

	Bivariate sample selection				IV bivariate sample selection[a]			
	Participation eq.		Outcome eq.		Participation eq.		Outcome eq.	
Variable	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Self-reported health	-0.04	-0.65	-0.18	-3.32	-0.09	-0.45	0.02	0.11
Private insurance	0.24	1.26	0.00	0.00	0.27	1.62	0.11	0.40
Log wage	0.05	0.63	0.12	1.29	0.07	0.67	0.07	0.60
$a = (\text{Age} - 50)/10$	0.24	3.68	0.24	2.27	0.25	3.66	0.33	3.60
a^2	0.00	-0.12	0.02	0.51	0.00	0.11	0.01	0.19
Secondary education	0.11	0.90	-0.08	-0.43	0.12	1.06	-0.10	-0.47
Tertiary education	0.17	1.15	-0.05	-0.23	0.20	1.44	-0.08	-0.30
Wife's education	0.00	0.02	0.00	0.15	0.00	-0.02	0.00	-0.12
Black	-0.04	-0.33	-0.37	-2.43	-0.05	-0.43	-0.38	-2.67
Other race	0.03	0.19	-0.50	-2.48	0.01	0.05	-0.51	-2.51
Married	-0.06	-0.20	-0.22	-0.47	-0.05	-0.18	-0.16	-0.35
Suburban	-0.01	-0.16	-0.01	-0.05	-0.01	-0.09	-0.02	-0.20
Rural	0.00	0.03	-0.01	-0.06	0.02	0.15	-0.01	-0.08
Household size	-0.01	-0.22	-0.11	-2.01	-0.01	-0.24	-0.11	-2.16
Constant	0.79	2.17	8.05	12.62	0.89	1.49	7.22	10.06
N	2,821		2,202 – 2,270[b]		2,821		2,202 – 2,270[b]	

[a] The first stage F statistic for the IV regression is 60.5.

[b] The number of positive observations varies between imputed data sets.

Table 4.6: Regression models for nights spent in a hospital

					Hurdle model					
	Poisson		Negative binomial		Probit		Zero-Truncated Negative Binomial		IV Poisson[a]	
Variable	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Self-reported health	-0.55	-4.46	-0.43	-3.79	-0.22	-5.85	-0.07	-0.62	-0.03	-0.06
Private insurance	-1.30	-3.64	-0.90	-2.59	-0.38	-3.12	-0.56	-1.66	-1.27	-3.32
Log wage	-0.14	-0.73	0.08	0.33	-0.07	-1.02	0.12	0.47	-0.27	-1.08
$a = (\text{Age} - 50)/10$	0.33	2.64	0.50	3.31	0.23	4.35	-0.04	-0.23	0.40	2.98
a^2	-0.04	-0.44	-0.02	-0.33	0.03	1.00	-0.13	-1.35	-0.05	-0.67
Secondary education	0.04	0.13	-0.51	-1.16	0.06	0.56	-0.22	-0.65	-0.07	-0.23
Tertiary education	0.28	0.58	-0.57	-1.22	0.03	0.24	-0.05	-0.11	0.12	0.28
Wife's education	-0.08	-1.88	-0.15	-2.46	-0.02	-1.16	-0.11	-1.30	-0.10	-1.89
Black	0.32	1.25	0.39	1.62	0.12	1.50	0.19	0.69	0.39	1.35
Other race	-0.49	-0.98	0.50	0.96	0.06	0.48	-0.05	-0.09	-0.47	-0.84
Married	0.08	0.12	0.89	0.94	0.02	0.08	0.59	0.51	0.22	0.32
Suburban	-0.18	-0.80	-0.04	-0.15	0.01	0.18	-0.17	-0.60	-0.21	-0.92
Rural	0.35	1.02	0.62	2.02	0.13	1.28	0.22	0.66	0.33	1.00
Household size	0.10	0.93	0.07	0.82	0.03	1.14	0.03	0.45	0.11	0.97
Constant	3.00	3.95	2.29	2.99	-0.09	-0.32	1.68	1.64	1.59	0.98
N	4,841		4,841		4,841		282		4,841	

[a] The first stage F statistic for the IV regression is 98.9.

almost all coefficients became insignificant. Our preliminary assessment of these results is that this is due to measurement error. It is well known (e.g., Wansbeek & Meijer, 2000, p. 140) that measurement error problems are greatly exacerbated in fixed effects estimators for panel data, and our possibilities for estimating panel data models are highly limited with only two effective waves of data.

Table 4.6 presents the results of the regressions for hospital nights. Not surprisingly, this shows a consistently negative effect of health. This replicates the findings in the literature. However with the IV estimator, it is not significant. Having private insurance is negatively related to spending nights in a hospital, in contrast with the positive sign we saw for out-of-pocket and total medical expenditures. Age is generally positively related to hospital nights, as it is with medical expenditures.

We compared our empirical results with those from the literature. This turned out to be difficult, because the empirical specifications generally do not correspond closely to the theoretical ones or because variables are used that are not closely related to those in the PSID. Examples of differences in specification are the inclusion of an additional health investment variable as an explanatory variable (e.g., doctor visits in the equation for hospital nights; Wagstaff, 1986a), or the use of multiple health measures as explanatory variables (e.g., both "permanent health" and number of sick days; Van de Ven & Van der Gaag, 1982). An example of a variable that is used and not closely related to those available in the PSID is the number of visits to the family doctor (e.g., Van de Ven & Van der Gaag, 1982; Wagstaff, 1986a; Erbsland et al., 1995).

With these caveats in mind, the closest comparisons appear to be models with hospital nights (or weeks) as the dependent variable, which is used by Van de Ven and Van der Gaag (1982), Wagstaff (1986a), and Erbsland et al. (1995). The first two use it in levels and the third uses the logarithm of 1 + the number of hospital nights, which should be very similar. Neither of these attempt to deal with the endogeneity of health, and thus their results should be most comparable to our non-IV results, in particular the (single equation) Poisson and negative binomial models. They obtain negative and highly statistically significant coefficients for health, with t-statistics between 3 and 19. Our t-statistics are between 4 and 6 and our results are thus, in terms of sign and statistical significance, in line with those in the literature.

Nocera and Zweifel (1998) use the logarithm of total medical expenditures, as reported by the insurance company, as their dependent variable, while accounting for the zeros through a tobit model but, again, not accounting for the endogeneity of health. Thus, their results should be most comparable to our non-IV sample selection model results for total medical expenditures, although arguably our respondent-reported measure is considerably more noisy than their measure. We find a statistically significant negative coefficient for self-reported health ($t = -3.3$). They find a negative but insignificant effect in their sample I, which (apart from the administrative data) is based on a single cross-sectional survey that reports on the past four years and is translated into four observations per respondent. Their sample II uses answers from two surveys (1981 and 1993) with partially the same respondents, who report about the past four years, and for the intervening years, the data are linearly interpolated, thus creating a data set with 12 "observations" per respondent. In this sample, the coefficient of health is positive and significant at the 5 percent level ($t = 2.26$). This t-statistic is likely inflated, because the covariates for a third of the observations are imputed in a way that does not take the variation into account, and standard errors do not appear to be adjusted for imputation uncertainty. Thus, unlike the results for hospital stays, which show a clear and strong relation with health, these, as well as our, results seem to indicate that medical expenditures are rather weakly related to health.

In our models, we used health investment from the next wave, because it is reported retrospectively and thus precedes the current health status, whereas in the model, health investment depends on health and thus should follow current health status. In the four papers mentioned, such corrections have apparently not been made. When we repeat our analyses for out-of-pocket and total medical expenditures with contemporaneously measured health investment as the dependent variable, the coefficients for health are more systematically negative, although still not significant for the IV models, despite the larger sample sizes (because we have three waves instead of two). One might argue that despite the reverse ordering (and thus the potential for reverse causality), contemporaneously reported medical expenses better approximate the desired measures, because health investment (esp. in the form of treatment) likely responds quickly to changes in health, and the contemporaneous report may be slanted towards the present and thus be closer to "current" investment than the report two years later.

4.7 Discussion

This chapter includes the first empirical tests of health capital theory with decreasing returns to scale. We start from the structural model developed by Galama (2011), which is an extension of the model of Grossman (1972a, 1972b) incorporating a decreasing returns to scale health production function as in Ehrlich and Chuma (1990). This model has a number of desirable characteristics that Grossman's constant returns to scale model does not have. In particular, there are substantive and empirical reasons to believe returns are decreasing. The model with decreasing returns to scale is able to reproduce stylized facts that run counter to the predictions of the constant returns model, and it has some theoretical properties that are more realistic.

The model leads to a derived structural relation between the demand for medical goods and services and health and other explanatory variables. After a few simplifying assumptions, a linear model equation is obtained that can be empirically estimated. We are the first to estimate an equation from a health capital model with decreasing returns to scale, and the first who take the endogeneity of health into account.

We obtain a statistically significant negative coefficient of health when the number of nights spent in a hospital is the dependent variable and when we do not take the endogeneity of health into account. Similarly, in the models for out-of-pocket medical expenditures or total medical expenditures, the dollar amounts are negatively related to health and statistically significant, although we do not find an effect for the participation equations (i.e., whether expenditures are positive or zero). This closely resembles the methodology and the findings in the literature and appears to support decreasing returns to scale, which predicts a negative coefficient, whereas constant returns to scale is associated with a positive coefficient.

However, when we attempt to control for the endogeneity of health by using instrumental variables methods, using childhood health and parental smoking during childhood as instruments, the coefficients become statistically insignificant and not consistently negative.

Our mixed findings suggest there is room for further research. Various technical issues need to be addressed in more detail, such as identification of sample selection equations by imposing sound exclusion restrictions, and measurement error in health investment. Also, the approximations through linearization we have employed in this chapter, although comparable to those in the literature, may not be accurate enough, and thus techniques such as GMM or nonlinear least squares may need to be employed to estimate more accurate nonlinear equations. Our current model equation does not allow us to derive the structural parameters as a function of the reduced form parameters, but additional equations can be derived from the theoretical model to estimate the structural parameters.

As in most of the literature on health capital models, our results only apply to working males. Understanding health investment in this population is of great interest, because health disparities are largely formed by age 50 (Smith, 2004; Case & Deaton, 2005). On the other hand, health costs are much larger later in life and thus it would be of interest to add the period after retirement, as well as the (endogenous) decision of when to retire. For persons 65 years and older, the availability of Medicare generally lowers out-of-pocket medical costs and premiums, encouraging health investment. In the other direction, after retirement, the production benefit of health—the effect of health on income—disappears. The steeply increasing medical costs may be the result of quickly decreasing returns to scale, of steeply accelerating biological aging rates, or of aggressive costly attempts to extend life. Models for females need to be much more involved than models for males, because the (dynamic) labor force participation decisions cannot be ignored at any age, and correspondingly, marriage and fertility need to be included in the model (Keane, 2011).

Chapter 5

Seek and ye shall find: how search requirements affect job finding rates of older workers

5.1 Introduction

The labor market position of older workers is usually solid if they are employed, as they are not very likely to be laid off. If they lose their job, however, it is very hard for them to find a new one. The latter is illustrated in the upper panel of Table 5.1, which shows that in the Netherlands the outflow rate of older unemployment insurance (UI) recipients is substantially smaller than that of prime age recipients. Gielen and Van Ours (2006) show that cyclical adjustments of the workforce in the Netherlands occur partly through fluctuations in separations for older workers. These separations are likely to be an one-way street out of the labor force into long-term unemployment. Moreover, it has been found that job loss among older workers results in large and lasting negative effects on future employment probabilities (see e.g., Chan & Stevens, 2001).

Receiving UI benefits is, in the Netherlands, as in most countries, conditional upon meeting criteria such as "availability for work" and "actively searching for a job" (Grubb, 2001). These criteria are, however, often less stringent for older workers in many European countries (OECD, 2006). In the Netherlands, UI recipients were for a long time exempted from the requirement to actively search for a job when they reached the age of 57.5. The reason for this was a relatively high unemployment rate among the young and the belief that older workers were holding their jobs. See also the lower panel of Table 5.1.

In this chapter we study how the exemption from the search requirement affected job search behavior of the UI recipients involved. We find that the removal of the search requirement had a large negative effect on the job finding rate. Furthermore, there is some evidence that already some time before the search requirement was removed the job finding rate goes down. Unemployed workers who

Table 5.1: Some labor market statistics

Outflow rate from unemployment into work					
Age	2001	2002	2003	2004	2005
15–25	0.24	0.26	0.27	0.26	0.23
25–35	0.97	0.87	0.76	0.78	0.74
35–45	0.89	0.76	0.64	0.59	0.62
45–55	0.73	0.66	0.52	0.48	0.53
55–65	0.60	0.40	0.31	0.27	0.27
Total	0.64	0.59	0.51	0.49	0.48
Unemployment rate					
Age	2001	2002	2003	2004	2005
15–25	5.0	5.4	7.3	9.0	9.4
25–35	2.2	2.9	4.1	4.7	4.7
35–45	1.9	2.5	3.5	4.4	4.6
45–55	2.1	2.4	3.2	4.1	4.1
55–65	1.7	2.4	3.0	3.9	4.5
Total	2.5	3.1	4.1	5.1	5.3

Source: Statistics Netherlands.

are getting close to the age of 57.5 seem to reduce their search intensity in anticipation of the removal of the search requirement. Apparently, even workers with a weak labor market position are able to react to the requirement that they should actively search for a job. Older workers thus seem to have at least some influence over their labor market position.

Our findings are relevant for a number of reasons. First, they are relevant for public policy in the context of aging. In recent years, OECD countries implemented several reforms which encouraged prolonged working lives for older workers in order to mitigate the adverse effects that population aging has on the financial sustainability of social security systems. An important aspect of these reforms are measures to get older unemployed workers back to work. Our results indicate that imposing job search requirements is an effective way to achieve this. Second, we contribute to the literature on the relationship between search requirements and job search. In previous studies, which are discussed in more detail in section 5.2, a change in search requirements often coincides with other measures, which makes it complicated to establish the separate effect of job search requirements.

This chapter is related to a recent paper by Bloemen, Hochguertel, and Lammers (2011), which utilizes the same variation in UI eligibility criteria as we do to examine how changes in search requirements for older UI recipients affect the transition rates to employment, early retirement and disability insurance. They find that stricter search requirements significantly increase the outflow to work, as well as outflow to disability insurance. This chapter differs from Bloemen et al. (2011) as follows: we do not only examine the effect of the removal of the search requirement, but we also study search behavior in the period just prior to the removal of the search requirement. Furthermore, we follow a different identification strategy for reasons to be discussed in more detail in section 5.4. Finally, we only focus on the transition from unemployment to work.

The remainder of this chapter is organized as follows. In section 5.2 we present and discuss previous studies on the relationship between search requirements and job search. In this section we also discuss the relationship between search requirements and job search from a theoretical point of view. In section 5.3 we provide details on the Dutch UI system. Section 5.4 describes the data and in section 5.5 we discuss our econometric approach. Empirical results and a range of sensitivity analyses are presented in section 5.6. Finally, section 5.7 concludes.

5.2 Search requirement and job search

5.2.1 Previous studies

Meyer (1995) gives an overview of a number of field experiments that have been conducted in the United States to evaluate alternative job search policies. These experiments typically combined more strict search requirements with job finding services, making it difficult to determine the relative importance of each of the measures.

The Washington Alternative Work-Search Experiment was designed to study the impact of different work-search policies for UI recipients. In particular, it was designed to assess the impact of three alternative work-search policies relative to the status quo policy requiring three job applications per week. The first alternative policy eliminated all work-search requirements, the second tailored the number of required job applications to personal characteristics, and the third involved job finding assistance at the beginning of the spell. The findings of this experiment, as reported by Johnson and Klepinger (1994), indicate that no work-search requirements significantly increase the length of the unemployment spell relative to the status-quo policy.

The Maryland UI Work Search Experiment was another experiment designed to study the effects of alternative search requirements. The experiment consisted of four different treatments: a first treatment required additional employer contacts, a second eliminated the requirement to report employer contacts, a third required UI recipients to attend a four-day job search workshop, and a fourth verified employer contacts. Klepinger et al. (2002) report findings from this experiment and conclude that stricter search requirements or employer contact verification reduce the length of the unemployment spell. Furthermore, they found no evidence that wages, earnings, or total income were affected, which suggests that higher non-monetary costs of continued benefit receipt were compensated by more intensive job search rather than by a reduction of the reservation wage.

Ashenfelter et al. (2005) report findings from another field experiment that was designed to measure whether stricter enforcement and verification of work search behavior alone decreases unemployment claims and benefits paid. In contrast to the results of Johnson and Klepinger (1994) and Klepinger et al. (2002), their results do not indicate that verification of search behavior led to shorter unemployment spells or lower total UI benefit payments.

This chapter is also related to a stream of literature that studies whether persons anticipate a future "treatment." For example, Katz and Meyer (1990) and Meyer (1990) find a dramatic increase in the probability of leaving unemployment just prior to the time of benefit exhaustion. Van Ours and Vodopivec (2006), studying UI benefits in Slovenia, find substantial spikes in the job finding rates around benefit exhaustion. The job finding rates in the month prior to benefit exhaustion are 2.2-2.5 times as high as in earlier months. Hairault, Sopraseuth, and Langot (2010) finds that an older worker's employment decision is affected by the distance to retirement. Black, Smith, Berger, and Noel (2003) study the effects of the Worker Profiling and Reemployment Services system which forces UI recipients with a high predicted probability of benefit exhaustion to enroll in a training program early in their unemployment spell. They not only find that the program reduces the number of weeks of UI benefit receipt, but also that once recipients learn that they may be enrolled in a job finding program they start to search for a job more intensively. Using Swiss data Lalive, van Ours, and Zweimüller (2005) examine the effect of benefit sanctions on the length of the unemployment spell. In the Swiss sanction system, a UI recipient must be notified that a sanction may be imposed and given the opportunity to clarify why he was not able to fulfill the eligibility requirements. Lalive et al. (2005) finds that this notification (or warning) has a positive impact on the job finding rate. Similar "threat effects" have been found in, among others, Geerdsen (2006), Rosholm and Svarer (2008), Van den Berg, Bergemann, and Caliendo (2009), Cockx and Dejemeppe (2010), and Crépon, Ferracci, Jolivet, and Van den Berg (2010).

5.2.2 Theoretical notions

The feature of the Dutch UI system that we study in this chapter, abolishing search requirements when UI recipients become 57.5 years old, introduces non-stationarity in a job search model. Furthermore, the Dutch UI system has sanctions that may be imposed when an UI recipient does not comply with the rules. Thus, a job search model that incorporates these features would need to combine aspects of a non-stationary search model, like Mortensen (1977) or Van den Berg (1990), with aspects of a job search model with sanctions, like Abbring, Van den Berg, and Van Ours (2005).

Job search requirements may affect the job finding rate for different groups of unemployed workers differently. First, UI recipients for whom the optimal search effort in the absence of sanctions is higher than the required minimum number of contacts (group 1) will exert the same effort regardless of whether or not a search requirements exists. Second, UI recipients for whom the optimal search effort in the absence of sanctions is lower than the required minimum number of contacts will either exert as much effort as required for as long as necessary (group 2) or will be willing to take the risk that a sanction is imposed (group 3).[1] Among UI recipients who will comply to the rules as long as necessary, search effort will show a discontinuity at the moment the search requirement is abolished. By contrast, among UI recipients who will take the risk that a sanction is imposed, search

[1] The fact that sanctions are actually imposed suggests that there are indeed UI recipients who search less than required. See for example Abbring et al. (2005).

effort will continuously decrease, for example because they think that the probability of a sanction decreases as they get closer to the moment at which the search requirement is abolished or because of the anticipated increase of the value of unemployment.

If additionally the reservation wage increases in anticipation of the higher value of unemployment, then, in the absence of duration dependence effects, the job finding rate will be constant for members of group 1. For members of group 2 the job finding rate will continuously fall until the moment of abolishment, discontinuously fall at the moment and be constant afterwards. Finally, for members of group 3 the job finding rate will continuously fall until the moment of abolishment and be constant afterwards.

The population of UI recipients will be a mixture of the three groups described above. Theoretical considerations suggest unambiguously that empirically we should find anticipation effects, but it is less clear that we should find a discontinuous jump in the job finding rate as soon as the search requirement is abolished.

5.3 The Dutch Unemployment Insurance System

In the Netherlands unemployed workers are entitled to unemployment benefits if they (1) are involuntarily and not culpably unemployed, i.e. they did not resign or were not fired for pressing reasons; (2) lose earnings for at least five working hours per week, or earnings for half of their working hours if employed for less than ten hours per week; (3) have been employed for at least 26 consecutive weeks out of the 39 weeks prior to unemployment; and (4) who are available for work.

If an unemployed worker meets the conditions mentioned above he or she is eligible for short-term benefits, which can be received for six months and which amount to 70 percent of the minimum wage. If in addition to the "26-out-of-39- weeks condition" he also received wages for at least 52 days in the four calendar years during the five years prior to unemployment, he qualifies for wage-related benefits. Depending on labor experience these benefits last for at least six months up to a maximum of five years.[2]

Labor experience is calculated as the number of years in the 5 calendar years prior to unemployment in which the person received wages for at least 52 days, plus the number of calendar years between the year that the person turned 18 and the 5 years prior to unemployment. As a result of the "4-out-of-5-years condition," the potential duration for wage-related benefits depends almost completely on the age at which the person becomes unemployed. A UI recipient with wage-related benefits receives 70 percent of the average wage received in the job from which he became unemployed in the 26 weeks prior to unemployment. The level is regularly adjusted to a general index of wages and limited to a maximum, which amounted to 152.62 Euro per day in 2001.

All persons who started to receive wage-related benefits before 11 August 2003, were eligible for extended benefits equal to 70 percent of the minimum wage. For the duration of extended benefits

[2]With a labor market experience of 30-35 years the potential duration of the wage-dependent benefits is 3 years, with 35-40 years of experience it is 4 years, and with more than 40 years of experience it is 5 years.

age was the only criterion. Persons who became unemployed before the age of 57.5, were entitled to two years of extended benefits, whereas older persons were entitled to 3.5 years. Thus, a person who became unemployed after the age of 57.5 and who met the conditions for wage-related benefits, had a potential benefit duration of at least four years, but it may have been as long as 7.5 years. He would therefore receive UI benefits until the mandatory retirement age of 65.[3] As of 11 August 2003, extended benefits are abolished, except for persons 50 years and older. However, for them extended benefits are means-tested such that with a household income of 70 percent of the minimum wage or more they do not receive extended benefits.

In most countries administrators of unemployment insurance typically impose job search requirements on the recipients to partially offset the negative impact unemployment benefits have on job search. In the Netherlands, up to 1 January 2004, UI recipients had to actively search for a job until the age of 57.5. Beyond that age they would still be obliged to accept job offers but a job search requirement was obsolete. From 1 January 2004 onwards, all new recipients were required to actively search for a job, and a search requirement was *activated* for recipients who were less than one year unemployed and who did not yet reach the age of the 62 years and 4 months at 31 December 2003. Regulations specified that a person needed to apply to at least four jobs in four weeks while these activities had to be reported at regular time-intervals to the employment office.[4]

In practice the number of job applications a recipient needed to make were tailored to: (1) the conditions of the regional labor market; (2) the number of "suitable" job opportunities; and (3) the medical situation and age of the person. Whether or not a recipient fulfilled the search requirement was decided upon by the case worker of the employment office. This worker also decides what is considered to be a "suitable" job, taking into account a person's previous job, his education, and earnings. Generally, the longer the unemployment duration lasts the broader the definition of "suitable" work becomes.

If a recipient does not comply with the rules, a benefit sanction may be applied. The noncomplying person can get a temporary or a permanent reduction in benefits (full or partial). Concerning the severity of the sanction Abbring et al. (2005) write: "In practice, the temporary partial reduction of the benefits ranges from 5 percent during 4 weeks to 25 or 30 percent during 13 weeks." These numbers are percentages of the previous wage or replacement-ratio percentage points. For example, a sanction of 20 percent on a worker with UI benefits equal to 70 percent of his previous wage is left with UI benefits equal to 50 percent of his previous wage.[5]

[3]It is known that in the past collecting UI benefits has been used as a retirement pathway (Heyma, 2001; Gruber & Wise, 1998).

[4]The stated employer contacts may have been verified by the employment office but this did not occur across the board.

[5]See Abbring et al. (2005) for further details.

5.4 Data

We use administrative data from Statistics Netherlands which are informative about unemployed persons who collect UI benefits. The data contain the start and end date of UI-benefit spells (with censoring at 31 December 2005); the reason for termination of UI-benefits; the type of benefits; the number of hours for which UI-benefits are collected; and whether unemployment is entered from employment or otherwise, like a period of sickness. Personal characteristics in the data include gender; age; marital status; educational levels; nationality; and a profiling indicator consisting of four levels of increasing "distance to the labor market." Based on the reported educational level, persons are stratified into three different classes: primary, secondary and tertiary education. The indicator for marital status treats cohabitors as married. "Distance to the labor market" is included by the dummy variable (phase 1) that equals one if the person has a "strong" labor market position according to the classification of the employment office.

To study the effect of the removal of the search requirement we use a sample which consists of persons who are required to search at the moment of inflow and who are exempted from it as soon as they reach the age of 57.5. As a consequence of the policy reform of January 2004, UI recipients who became unemployed before 31 December 2003 and who did not reach the age of 57.5 by then, needed to continue searching when turning 57.5. Persons who became unemployed in 2001 and who were 55.5 years or older at inflow, were at least 57.5 years old at 31 December 2003 and hence only subject to initial regulations. Recipients in the same age range (at inflow) who became unemployed in 2002 or 2003 may have been affected by the policy reform of 1 January 2004. For these reasons and because we do not have data on the years prior to 2001, we select workers who became unemployed in 2001 at an age between 55.5 and 57.5 for our main analysis. We select workers who became unemployed in 2004 at an age between 55.5 and 57.5 for a sensitivity analysis.

As indicated in the introduction, this chapter is closely related to Bloemen et al. (2011). That paper takes a difference-in-difference approach and compares two age groups, 55.5-57.5 and 57.5-59.5, between two years, 2001 and 2004. The following proportional hazard specification is a simplification of the model actually used by them, but it contains the elements essential to understand why and how our approach differs from theirs. Ignoring the effects of observed and unobserved personal characteristics, the job finding rate, $\lambda(t)$, is specified as

$$\begin{aligned}\lambda(t) = \lambda_0(t)\exp(&\delta_1 1(y = 2004) + \delta_2 1(\tau \geq 57.5) + \delta_3 1(y = 2004)1(\tau \geq 57.5) \\ &+ \delta_4 1(\tau(t) \geq 57.5) + \delta_5 1(y = 2004)1(\tau(t) \geq 57.5)),\end{aligned}$$

where y stands for year of inflow, τ for age at inflow, and $\tau(t)$ for age at time (duration) t. Furthermore, $1(.)$ is the indicator function that equals 1 if the logical is true and zero otherwise, $\lambda_0(t)$ is the baseline hazard, and the δ's are parameters.[6] Bloemen et al. (2011) argue that the parameter δ_3 is informative

[6]Bloemen et al. (2011) specify a mixed proportional hazard model for each of the transitions they consider (from unemployment to work, to early retirement, and to disability insurance.) The unobserved heterogeneity component is assumed to

about the impact of job search requirements as of the beginning of the the unemployment spell for persons 57.5 years and older at inflow. They further argue that the parameter δ_5 is informative about the impact of being required to continue searching for a job when turning 57.5 (as opposed to being allowed to stop searching). However, the parameter estimates of δ_3 and δ_5 may be biased to the extent that different rules for extended benefits affected the composition of both age groups and to the extent that these rules affected the job finding rate. Furthermore, the parameter estimates of δ_3 and δ_5 may be biased to the extent that the reform of 11 August 2003 affected the composition of both age groups and thereby invalidating the common trend assumption of difference-in-difference identification strategy.[7]

Since the focus in our main analysis is on workers who became unemployed in 2001 at an age between 55.5 and 57.5, we avoid that our estimate of the impact of job search requirements on the job finding rate is biased due to the reform of 11 August 2003. As a consequence, the "treatment" we analyze is in fact different from those considered in Bloemen et al. (2011).

In our analyses, we further focus on persons with a Dutch nationality, thus abstaining from issues on the labor market position of immigrants. Finally, we select unemployed workers who are non-seasonally full-time unemployed.

Table 5.2: Descriptive statistics

Variable	2001	2004
Age	56.4	56.4
Sex male (%)	84.6	82.8
Primary education (%)	15.9	8.1
Secondary education (%)	65.1	70.2
Tertiary education (%)	19.0	21.7
Not alone (%)	74.8	73.8
Phase 1 (%)	85.4	82.8
Completed spell (%)	40.0	31.1
of which:		
Type 1 (%)	7.6	11.7
Type 2 (%)	8.6	9.3
Type 3 (%)	76.9	69.2
Type 4 (%)	5.4	7.6
Type 5 (%)	1.4	2.1
Sample size	1,606	2,778

Note that the types of completed spells are explained in the main text.

follow a discrete distribution with two points of support, again for each of the transitions they consider.

[7]Recall from section 5.3 that before 11 August 2003 the potential length of the UI spell was dependent on whether a person became unemployed before or after 57.5. This age-dependency of the potential UI spell was abolished with the reform of 11 August 2003. Not only the age-dependency was abolished, all workers who became unemployed in 2004 faced different rules for the extended benefits (both for the amount and the time they can be received) than those who became unemployed in 2001.

Figure 5.1: Monthly job finding rates by duration of unemployment

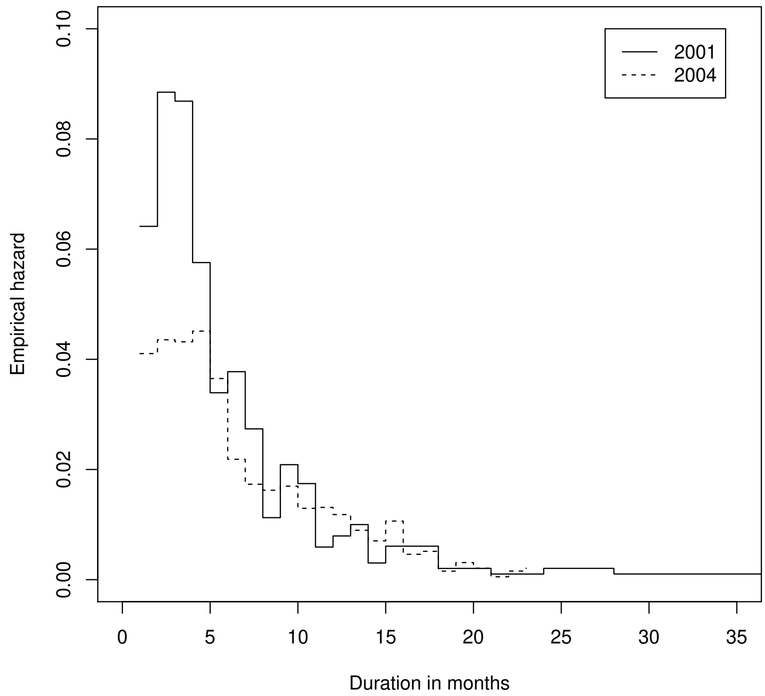


Table 5.2 contains descriptive statistics for both 2001 and 2004 on the covariates we use in the analysis. Perhaps most interesting is the high percentage of the UI recipients in our sample who have a "strong" connection to the labor market according to the employment office. Figure 5.1 shows the empirical hazard for our sample. The outflow to a job is relatively high in the first six months and then quickly drops to zero in the period afterwards.

5.5 Econometric approach

We assume that differences in transition rates from unemployment to work can be characterized by observed time-varying covariates $(z(t))$, age at inflow (τ), other observed time-invariant covariates (x), an unobserved random variable (v), and the elapsed duration of unemployment itself (t). The job finding rate at time t conditional on $z(t)$, τ, x and v, denoted by $\lambda(t|z(t), \tau, x, v)$, is assumed to have a mixed proportional hazard specification (see, e.g., Lancaster, 1990; Van den Berg, 2001)

$$\lambda(t|z(t), \tau, x, v) = \lambda_0(t) \exp(\alpha' z(t) + \beta\tau + \gamma' x + v),$$

where β is a parameter and α' and γ' are vectors of parameters. Furthermore, $\lambda_0(t)$ represents duration dependence – the baseline hazard – in the form of a flexible piecewise-constant function

$$\lambda_0(t) = \exp\left(\sum_{j=1}^{J} \mu_j 1_j(t)\right),$$

where $1_j(t)$ is the indicator function that equals 1 if t is in the j-th interval and 0 otherwise, and where the μ's are parameters to be estimated. We distinguish four time intervals: less than 3 months, 3–6 months, 6–12 months, and 12 and more months. Because we also estimate a constant term, we normalize the duration dependence parameter of the first interval to zero ($\mu_1 = 0$).

Let t^a be the time until a person starts to anticipate the abolishment of the search requirement and let t^s be the time until abolishment, then in our baseline specification we specify

$$\alpha' z(t) = \alpha_1 1(t^a \leq t < t^s) + \alpha_2 1(t^s \leq t).$$

The effect of the anticipation of the abolishment of the search requirement is captured by α_1 and the effect of the abolishment itself by α_2. In our baseline specification we assume that $t^a = t^s - 3$ (months). Since there is (a priori) no reason why UI recipients start anticipating the removal 3 months in advance, we also consider specifications with different anticipation intervals as part of the sensitivity analyses.

All persons in our sample have a search requirement at the start of their unemployment spell, i.e., $t^0 < t^s$, where t^0 is the moment at which a person becomes unemployed. Furthermore, when our sample members are unemployed long enough the search requirement is removed. We do not require that a person is unemployed before the moment at which we conjecture he starts to anticipate the removal of the search requirement, thus it can be that $t^a < t^0$ for some persons in our sample. Let t^u denote the completed duration of the unemployment spell. Then, the following situations can occur within our sample:

1. $t^0 < t^a < t^s < t^u$: the worker enters unemployment before the start of the anticipation period and finds a job after age 57.5;
2. $t^0 < t^a < t^u < t^s$: the worker enters unemployment before the start of the anticipation period and finds a job before age 57.5;
3. $t^0 < t^u < t^a < t^s$: the worker enters and leaves unemployment before the start of the anticipation period;
4. $t^a < t^0 < t^s < t^u$: the worker enters unemployment during the anticipation period and finds a job after age 57.5;
5. $t^a < t^0 < t^u < t^s$: the worker enters unemployment during the anticipation period and finds a job before age 57.5;

The identification of the anticipation and the 57.5+ (or treatment) effect is determined by which of the five above cases apply. The identification of the anticipation effect is through cases 1, 2, 4 and 5. For the identification of the 57.5+ effect cases 1 and 4 are important. In cases 4 and 5 there is a correlation between anticipation effect and the "age at inflow effect." Table 5.2 provides information about the relative size of each of these five groups, with $t^a = t^s - 3$ (months).

Note that it is possible to allow the hazard rate to depend on time (t), and the time until the removal of the search requirement (t^s), because t^s is a time-varying covariate that varies within the sample. This is analog to the reasoning in papers studying the impact of the potential benefit duration on the duration of unemployment; they can allow the hazard to depend on time and the time until exhaustion (see e.g. Katz & Meyer, 1990; Meyer, 1990). Note also that it is not possible to separately identify the anticipation effect and the age effect without imposing some structure; we assume that age at inflow enters "linearly" into the hazard rate.

The unobserved heterogeneity component v is assumed to follow a discrete distribution with two points of support, (v_1, v_2), where $\Pr(v = v_1) = p$, $\Pr(v = v_2) = 1 - p$, and $p = \exp(\eta)/(1 + \exp(\eta))$. We normalize v_1 to zero and estimate v_2.[8]

A key variable in our analyses is age at inflow. If age at inflow would be precisely measured, the loglikelihood contribution of the i-th person would be

$$L_i = d_i \log f(t_i^u | z_i(t), \tau_i, x_i, t_i^a, t_i^s) + (1 - d_i) \log S(t_i^u | z_i(t), \tau_i, x_i, t_i^a, t_i^s),$$

where $d_i = 1$ if the i-th person's spell is uncensored and $d_i = 0$ if censored, where $f(\cdot)$ denotes the probability density function and $S(\cdot)$ the survivor function, and where t_i^a and t_i^s depend on the age at inflow. In the data, however, age at inflow is reported in years and months only, and hence it is imprecisely measured. Thus, the *observed* age at inflow is lower than the *actual* age at inflow. How much it is lower depends on the month of birth, but unfortunately we do not have this information nor can it be inferred from the information we do have. Therefore, we proceed as follows: let τ denote the actual age at inflow, $\hat{\tau}$ the observed age at inflow and assume that $\tau = \hat{\tau} + \varepsilon$, where ε follows a discrete uniform distribution with $\{0, 1, \ldots, 30\}$ as possible outcomes, each with probability $1/31$. The loglikelihood contribution for person i can then be written as

$$L_i = \frac{1}{31} \sum_{k=0}^{30} [d_i \log f(t_i^u | z_i(t), \tau_i + k, x_i, t_i^a, t_i^s) + (1 - d_i) \log S(t_i^u | z_i(t), \tau_i + k, x_i, t_i^a, t_i^s)].$$

5.6 Parameter estimates

In the absence of detailed information on how the removal of the search requirement was operationalized, we assume in our baseline specification that a person was exempted from it as of the first day of the month in which he turned 57.5. Furthermore, we assume that UI recipients start to anticipate the removal of the search requirement three months before it actually removed. The baseline parameter estimates based on the 2001 inflow sample are reported in Table 5.3. The first set of results,

[8] We have also estimated our baseline models with three points of support for the unobserved heterogeneity, but these points collapsed to two.

specification (1), indicate that, in anticipation of the removal of the search requirement, UI recipients lower their search efforts. The effect is, however, statistically insignificant at conventional levels. At the moment the search requirement is removed, i.e. the difference between the 57.5+ effect and the anticipation effect, the job finding rate decreases by another 27 percentage points ($\approx 1 - \exp(-0.32)$, t = -1.67). After the removal of the search requirement, the job finding rate is significantly lower (the 57.5+ effect), showing that even for persons with a rather weak labor market position a job search requirement is an effective instrument in getting them back to work. Male workers have a higher job finding rate than otherwise equivalent female workers. Unemployed workers with secondary or tertiary education have a lower job finding rate. Also, the job finding rate for UI recipients with a relatively strong labor market position is more than twice as high as that of a recipient with a weaker labor market position.

The second set of results in Table 5.3, specification (2), shows the relevant parameter estimates if ignore the anticipation effect. The negative effect of the abolishment of the search requirement is still significantly different from zero but substantially smaller than in specification (1).

Table 5.3: Baseline parameter estimates for 2001

	(1)		(2)	
	Coeff	t-value	Coeff	t-value
Anticipation <3 months	-0.21	-1.31	-	-
57.5+ effect	-0.53	-2.76	-0.42	-2.43
Male	0.24	1.88	0.24	1.87
Age at inflow / 10	-1.17	-1.15	-1.89	-2.19
Secondary education	-0.49	-4.39	-0.49	-4.39
Tertiary education	-1.05	-6.31	-1.05	-6.32
Not alone	-0.07	-0.72	-0.07	-0.71
Phase 1	0.99	5.83	0.99	5.82
Constant	0.31	0.05	4.38	0.90
Duration dependence				
3-6 months	-0.31	-2.37	-0.32	-2.49
6-12 months	-1.12	-5.36	-1.15	-5.62
$>$ 12 months	-3.39	-10.30	-3.46	-10.73
v_2	-20[a]		-20[a]	
η	1.94	1.35	1.91	1.37
$\Pr(v = v_2)$	0.13		0.13	
Log-likelihood	-4,726.46		-4,727.35	
N	1,606		1,606	

[a] Estimating v_2 sometimes gave numerical problems for certain models. We therefore plug in $v_2 = -20$, approximately the estimated value in cases without numerical problems.

In our baseline specifications we postulated that UI recipients start to anticipate the removal of the search requirement 3 months before its actual removal and that the anticipation effect is constant during this period. Table 5.4 presents the results of several models that make different assumptions

Table 5.4: Specifications with different anticipation intervals; 2001 inflow

	(1)		(2)		(3)	
	Coeff	t-value	Coeff	t-value	Coeff	t-value
Anticipation 2-3 months	-	-	-	-	0.11	0.56
Anticipation 1-2 months	-	-	-0.48	-1.85	-0.45	-1.71
Anticipation <1 months	-0.37	-1.45	-0.44	-1.71	-0.41	-1.58
57.5+ effect	-0.48	-2.72	-0.57	-3.10	-0.54	-2.84
Log-likelihood	-4,726.20		-4,724.26		-4,723.91	
N	1,606		1,606		1,606	

Note that the same specifications are used as in Table 5.3 but only the relevant parameters are reported.

regarding the anticipatory behavior of UI recipients. In specification (1), we assume that recipients start to anticipate the change in policy only 1 month before the actual removal. The anticipation effect is larger than in the baseline specification, but again it is not statistically significant.

In specification (2), we assume that recipients start to anticipate the change in policy in the two months prior to it. Furthermore, we allow the anticipation effect to be different in the first month of anticipation relative to the second month. The estimates of both anticipation effects are similar however, suggesting that the anticipation effect is constant within the two month period. By contrast to the previous results, the anticipation effects are now significant at the 10 percent level. The effect of the removal of the search requirement is again the same as in the baseline specification. In specification (3), we assume that recipients start to anticipate the change in policy three months prior to the removal of the search requirement. As in specification (2) we allow the anticipation effect to differ across "anticipation months." The results indicate that unemployed workers start to anticipate the removal of the search requirement only in the two months prior to the actual removal.

In none of the specifications the effect of moving from just before 57.5 to just after (i.e., the difference between the 57.5+ effect and the anticipation effect in the last month) is statistically different from zero. At the same time the results do indicate that the job finding rate is significantly lower when UI recipients are not required to actively search for a job (57.5+ effect).

The results presented so far may be biased when persons are able to choose the age at which they get unemployed. The discussion in section 5.3 pointed out that, at least before 11 August 2003, there were strong incentives for persons to get unemployed after the age of 57.5, since in that case they can receive unemployment benefits until the mandatory retirement age of 65. A first indication of this may be provided by the inflow rates by age. These do not show a drop in inflow just before 57.5, but do show a spike just after 57.5. Tuit and Van Ours (2010) conclude, based on the same data as we use, that "workers had some influence on the timing of their unemployment spell and when possible they use this influence to their advantage." So, our results may be biased. Assuming that persons cannot postpone their inflow into unemployment by more than six (three) months, analyzing a sample which consists of person younger than 57 (57 and 3 months) at inflow (and older than 55.5) provides evidence of the sensitivity of our conclusions to the sample. The relevant parameter estimates are presented in Table 5.5, specifications (1) and (2). The anticipation effect is much smaller than in the baseline specification and remains statistically insignificant. In both specifications there is a significant effect at 57.5 (the difference between the 57.5+ effect and the anticipation effect, t = -2.72 and -3.12) which is substantially larger than in the baseline specification. Also, the effect of not having a search requirement (57.5+ effect) is significant and the size is the same between the two specifications, but in both cases the effect is larger than in the baseline specification. These results suggest that our conclusions are not sensitive to the "sample selection problem."

So far we have assumed that UI recipients were exempted from the search requirement as of the first day of the month in which they turned 57.5. However, as we do not know when exactly the abolishment was implemented, we estimate two models in which we assume a different moment of abolishment. In specification (3) of Table 5.5 we assume that the search requirement is removed at the end of the month in which a person turn 57.5, whereas in specification (4) we assume that it is removed at a person's 57.5 "birthday." Across both specifications, the size of the anticipation effect and 57.5+ effect are the same, and in both cases the effects are larger than in the baseline specification and now also the anticipation effects are statistically significant. Also, there is a significant negative effect on the job finding rate at 57.5.

As a final sensitivity analysis we estimated our baseline model on inflow data from 2004, to examine whether our findings are indeed due to the removal of the search requirement. The parameter estimates are given in the second column of Table 5.6. There is no significant anticipation effect and also no significant effect at 57.5 (t = -1.08) at the job finding rate. Surprisingly however, the results presented in specification (1) indicate that there is still a negative and significant 57.5+ effect although it is substantially smaller than for the 2001 inflow sample and only statistically significant at the 10 percent level. However, the results of specification (2) indicate that the 57.5+ effect disappears once we do not allow for an anticipation effect. Furthermore, once we restrict the sample to UI recipients flowing in after 1 July 2004, we find a 57.5+ effect that does not differ significantly from zero (results not shown). These results support the view that the finding of a negative and significant 57.5+ effect when using all inflow from 2004, is due to a delayed implementation of the reform of 1 January 2004. Another explanation may be that HRM departments of firms are, in the beginning, not used to receiving job application letters from persons older than 57.5 and therefore put them aside.

Table 5.5: Other sensitivity analyses

	(1) Age at inflow < 57		(2) Age at inflow < 57 + 3 months		(3) Removal at end of the month		(4) Removal at 57.5 birthday	
	Coeff	t-value	Coeff	t-value	Coeff	t-value	Coeff	t-value
Anticipation <3 months	-0.08	-0.37	-0.17	-1.06	-0.40	-2.39	-0.39	-2.39
57.5+ effect	-0.96	-3.03	-0.91	-3.85	-0.79	-3.72	-0.84	-4.11
Log-likelihood	-3,938.65		-4,438.16		-4,722.42		-4,721.01	
N	1,274		1,487		1,606		1,606	

Note that the same specifications are used as in Table 5.3 panel a. but only the relevant parameters are reported.

Table 5.6: Parameter estimates of models for 2004

	(1)		(2)	
	Coeff	t-value	Coeff	t-value
Anticipation <3 months	-0.12	-0.89	-	-
57.5+ effect	-0.28	-1.71	-0.20	-1.45
Male	0.68	4.23	0.69	4.24
Age at inflow / 10	-0.75	-0.62	-1.25	-1.16
Secondary education	-0.51	-2.68	-0.52	-2.68
Tertiary education	-0.88	-3.96	-0.88	-3.97
Not alone	0.07	0.54	0.07	0.55
Phase 1	1.42	7.93	1.43	7.99
Constant	-2.36	-0.35	0.44	0.07
Duration dependence				
3-6 months	0.35	2.72	0.35	2.74
6-12 months	0.07	0.35	0.07	0.31
$>$ 12 months	-0.05	-0.18	-0.07	-0.25
v_2	-3.08	-11.96	-3.12	-12.38
η	-0.94	-3.65	-0.93	-3.81
$\Pr(v = v_2)$	0.72		-0.72	
Log-likelihood	-6,790.64		-6,791.04	
N	2,778		2,778	

5.7 Conclusions

Older workers face a difficult labor market position when they lose their job; the prospects of finding a new job are not very good. Therefore, unemployment durations among older unemployed workers are relatively long. For a long time the labor market position of older workers in the Netherlands was considered to be so poor, that from age 57.5 onwards, UI recipients were no longer required to actively search for a job although they still had to to accept job offers. In 2004 this rule was abolished and the active search requirement was also imposed upon unemployed workers beyond age 57.5.

In this chapter we analyze job finding rates of workers who became unemployed in 2001 over a relatively short age span, from 55.5 to 57.5 years. We are particularly interested in the evolution of the job finding rates as these workers got close to the point were the active search requirement was abolished to find out whether any anticipation effects occurred. Also, we are interested in the magnitude of the effect of the abolishment on the job finding rate. It is not easy to identify an anticipation effect as job finding rates are influenced by duration dependence as well as age dependence. Overall, our results are somewhat sensitive to model specification. Nevertheless, investigating the job finding rates of these older workers we find that the abolition of the requirement to actively search for a job had a relatively large negative effect. Furthermore, although the results are not unambiguous it seems that unemployed workers who are getting close to the age of 57.5 reduce their search intensity. Thus, there is some evidence of a negative anticipation effect.

Our evidence for a downward shift in search intensity at the moment the search requirement was abolished and thereafter suggests that there is a large group of older UI recipients who felt obliged to stick to the rules of the game and actually kept on searching until as long as was necessary. Furthermore, we find strong evidence that not imposing a search requirement has a large negative effect on the job finding rate. This is remarkable as job search is among the activities which are not very popular among unemployed workers. Knabe, Rätzel, Schöb, and Weimann (2010) for example study life satisfaction measured as a general feeling and momentary satisfaction related to specific activities and find that UI recipients consider being employed as a desirable state but they do not value the activities which would speed up the transition to this state sufficiently. Thus, it makes sense to impose an obligation to unemployed workers to actively search for a job.

Although the absolute increase in the job finding rates among older workers for whom the search requirement is reinstalled is modest, the fact that there is an increase at all is remarkable given the relatively weak labor market position of older workers. Apparently, even older workers have some influence over their job finding. Extrapolating our findings to younger age categories it is clear that it is important to have well specified search requirements which should be enforced to make sure that UI recipients keep searching for a job irrespective of how long they have been unemployed. Even workers with seemingly poor job prospects seem to benefit from the requirement to actively search for a job.

References

Abbring, J. H., Van den Berg, G. J., & Van Ours, J. C. (2005). The effect of unemployment insurance sanctions on the transition rate from unemployment to employment. *Economic Journal*, *115*, 602–630.

Akerlof, G. A. (1970). The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, *84*, 488–500.

Ashenfelter, O., Ashmore, D., & Deschênes, O. (2005). Do unemployment insurance recipients actively seek work? Evidence from randomized trials in four U.S. States. *Journal of Econometrics*, *125*, 53–75.

Attanazio, O. P., & Weber, G. (1989). Intertemporal substitution, risk aversion and the Euler equation for consumption. *Economic Journal*, *99*(Suppl.), 59–73.

Bago d'Uva, T., Van Doorslaer, E., Lindeboom, M., & O'Donnell, O. (2008). Does reporting heterogeneity bias the measurement of health disparities? *Health Economics*, *17*, 351–375.

Bago d'Uva, T., Van Doorslaer, E., Lindeboom, M., & O'Donnell, O. (2011). Slipping anchor? Testing the vignette approach to identification and correction of reporting heterogeneity. *Journal of Human Resources*, *46*, 872–903.

Battistin, E., Brugiavini, A., Rettore, E., & Weber, G. (2009). The retirement consumption puzzle: Evidence from a regression discontinuity approach. *American Economic Review*, *99*, 2209–2226.

Black, D., Smith, J., Berger, M., & Noel, B. (2003). Is the threat of training more effective than training itself? Evidence from random assignments in the UI system. *American Economic Review*, *93*, 1313–1327.

Bloemen, H., Hochguertel, S., & Lammers, M. (2011). *Job search requirements for older unemployed: Transitions to employment, early retirement and disability benefits* (Working Paper No. 5442). Bonn, Germany: IZA.

Bound, J. (1991). Self-reported versus objective measures of health in retirement models. *Journal of Human Resources*, *26*, 106–138.

Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In J. Heckman & E. Leamer (Eds.), *Handbook of Econometrics, volume 5* (pp. 3705–3843). Elsevier Science.

Buchmüller, T., Grumbach, K., Kronick, R., & Kahn, J. (2005). The effect of health insurance on medical care utilization and implications for insurance expansion: A review of the literature.

Medical Care Research and Review, *62*, 3–30.

Buckley, J. (2008). *Survey context effects in anchoring vignettes* (Working Paper). Retrieved from `http://polmeth.wustl.edu/media/Paper/surveyartifacts.pdf`

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. New York, NY: Cambridge University Press.

Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics using Stata*. College Station, TX: Stata Press.

Card, D., Dobkin, C., & Maestas, N. (2008). The impact of nearly universal insurance coverage on health care utilization: Evidence from Medicare. *American Economic Review*, *98*, 2242–2258.

Card, D., Dobkin, C., & Maestas, N. (2009). Does Medicare save lives? *Quarterly Journal of Economics*, *124*, 597–636.

Case, A., & Deaton, A. (2005). Broken down by work and sex: How our health declines. In D. A. Wise (Ed.), *Analyses in the economics of aging* (pp. 185–212). Chicago, IL: University of Chicago Press.

Chan, S., & Stevens, A. H. (2001). Job loss and employment patterns of older workers. *Journal of Labor Economics*, *19*, 484–521.

Chiappori, P.-A., Durand, F., & Geoffard, P.-Y. (1998). Moral hazard and the demand for physician services: First lessons from a French natural experiment. *European Economic Review*, *42*, 499–551.

Chiappori, P.-A., & Paiella, M. (2011). Relative risk aversion is constant: Evidence from panel data. *Journal of the European Economic Association*, *9*, 1021–1052.

Chiappori, P.-A., & Salanie, B. (2000). Testing for asymmetric information in insurance markets. *Journal of Political Economy*, *108*, 56-78.

Cockx, B., & Dejemeppe, M. (2010). *The threat of monitoring job search: a discontinuity design* (Working Paper No. 5337). Bonn, Germany: IZA.

Colombo, F., & Tapay, N. (2004). *Private health insurance in OECD countries: The benefits and costs for individuals and health systems* (Working Paper No. 15). Paris: OECD.

Crépon, B., Ferracci, M., Jolivet, G., & Van den Berg, G. J. (2010). *Analyzing the anticipation of treatments using data on notification dates* (Working Paper No. 5265). Bonn, Germany: IZA.

Cropper, M. L. (1977). Health, investment in health, and occupational choice. *Journal of Political Economy*, *85*, 1273–1294.

Cutler, D., & Zeckhauser, R. (2000). The anatomy of health insurance. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of Health Economics* (Vol. 1A).

Datta Gupta, N., Kristensen, N., & Pozzoli, D. (2010). External validation of the use of vignettes in cross-country health studies. *Economic Modelling*, *27*, 654–865.

De Nardi, M., French, E., & Jones, J. B. (2010). Why do the elderly save? The role of medical expenses. *Journal of Political Economy*, *118*, 39–75.

Deveugele, M., Derese, A., Van den Brink-Muinen, A., Bensing, J., & De Maeseneer, J. (2002).

Consultation length in general practice: Cross sectional study in six European countries. *British Medical Journal*, *325*, 472–477.

Dowd, J., & Zajacova, A. (2007). Does the predictive power of self-rated health for subsequent mortality vary by socioeconomic status in the U.S.? *International Journal of Epidemiology*, *36*, 1214–1221.

Duan, N., Manning, W. G., Jr., Morris, C. N., & Newhouse, J. P. (1983). A comparison of alternative models for the demand for medical care. *Journal of Business & Economic Statistics*, *1*, 115–126.

Duan, N., Manning, W. G., Jr., Morris, C. N., & Newhouse, J. P. (1984). Choosing between the sample-selection model and the multi-part model. *Journal of Business & Economic Statistics*, *2*, 283–289.

Ehrlich, I., & Chuma, H. (1990). A model of the demand for longevity and the value of life extension. *Journal of Political Economy*, *98*, 761–782.

Erbsland, M., Ried, W., & Ulrich, V. (1995). Health, health care, and the environment. Econometric evidence from German micro data. *Health Economics*, *4*, 169–182. (Reprinted in A. M. Jones & O. O'Donnell, 2002, *Econometric analysis of health data*, Wiley, New York, pp. 25–36.)

Fang, H., Keane, M., & Silverman, D. (2008). Sources of advantageous selection: Evidence from the Medigap insurance market. *Journal of Political Economy*, *116*, 303–350.

Feldstein, P. J. (1993). *Health care economics* (4th ed.). Albany, NY: Delmar.

Finkelstein, A., Luttmer, E. F. P., & Notowidigdo, M. J. (2008). *What good is wealth without health? The effect of health on the marginal utility of consumption* (Working Paper No. 14089). Cambridge, MA: National Bureau of Economic Research.

Finkelstein, A., & McGarry, K. (2006). Multiple demensions of private information: Evidence from the long-term care insurance market. *American Economic Review*, *96*, 938–958.

French, E. (2005). The effects of health, wealth, and wages on labour supply and retirement behaviour. *Review of Economic Studies*, *72*, 395–427.

Galama, T. J. (2011). *A contribution to health capital theory* (Working Paper No. WR-831). Santa Monica, CA: RAND Corporation.

Galama, T. J., & Kapteyn, A. (2011). Grossman's missing health threshold. *Journal of Health Economics*, *30*, 1044–1056.

Geerdsen, L. P. (2006). Is there a threat effect of labour market programmes? A study of ALMP in the Danish UI system. *Economic Journal*, *116*, 738 – 750.

Geil, P., Million, A., Rotte, R., & Zimmerman, K. (1997). Economic incentives and hospitalization in Germany. *Journal of Applied Econometrics*, *12*, 295–311.

Gerdtham, U.-G., & Johannesson, M. (1999). New estimates of the demand for health: Results based on a categorical health measure and Swedish micro data. *Social Science & Medicine*, *49*, 1325–1332.

Gerdtham, U.-G., Johannesson, M., Lundberg, L., & Isacson, D. (1999). The demand for health:

Results from new measures of health capital. *European Journal of Political Economy*, *15*, 501–521.

Gielen, A., & Van Ours, J. C. (2006). Age-specific cyclical effects in job reallocation and labor mobility. *Labour Economics*, *13*, 493–504.

Goldman, D. P., Jena, A. B., Lakdawalla, D. N., Malin, J. L., Malkin, J. D., & Sun, E. (2010). The value of specialty oncology drugs. *Health Services Research*, *45*, 115–132.

Grossman, M. (1972a). *The demand for health: A theoretical and empirical investigation* (Occasional Paper No. 119). New York, NY: National Bureau of Economic Research.

Grossman, M. (1972b). On the concept of health capital and the demand for health. *Journal of Political Economy*, *80*, 223–255.

Grossman, M. (1998). On optimal length of life. *Journal of Health Economics*, *17*, 499–509.

Grossman, M. (2000). The human capital model. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of Health Economics* (Vol. 1, pp. 347–408). Amsterdam, The Netherlands: North-Holland.

Grubb, D. (2001). Eligibility criteria for unemployment benefits. In *Labour Market Policies and the Public Employment Service* (pp. 187–216). Paris: OECD.

Gruber, J., & Wise, D. (1998). Social security and retirement: An international comparison. *American Economic Review*, *88*, 158–163.

Hahn, J., Todd, P., & Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression discontinuity design. *Econometrica*, *69*, 201–209.

Hairault, J.-O., Sopraseuth, T., & Langot, F. (2010). Distance to retirement and older workers' employment: The case for delaying retirement age. *Journal of the European Economic Association*, *8*, 1034–1076.

Hansen, L. P., & Singleton, K. J. (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica*, *50*, 1269–1286.

Hay, J. W., & Olsen, R. J. (1984). Let them eat cake: A note on comparing alternative models of the demand for medical care. *Journal of Business & Economic Statistics*, *2*, 279–282.

Herzog, A. R., & Wallace, R. B. (1997). Measures of cognitive functioning in the AHEAD study. *Journal of Gerontology Series B*, *52B*, 37–48.

Heyma, A. (2001). *Dynamic models of labour force retirement: An empirical analysis of early exit in the Netherlands*. Unpublished doctoral dissertation, University of Amsterdam.

Hopkins, D., & King, G. (2010). Improving anchoring vignettes: Designing surveys to correct for interpersonal incomparability. *Public Opinion Quarterly*, *74*, 201–222.

Huisman, M., Van Lenthe, F., & Mackenbach, J. (2007). The predictive ability of self-assessed health for mortality in different educational groups. *International Journal of Epidemiology*, *36*, 1207–1213.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: A review of twenty-seven community studies. *Journal of Health and Social Behavior*, *38*, 21–37.

Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*, 467–475.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to pratice. *Journal of Econometrics*, *142*, 615–635.

Johnson, T., & Klepinger, D. (1994). Experimental evidence on unemployment insurance work-search policies. *Journal of Human Resources*, *29*, 694–717.

Jones, A. M. (2000). Health econometrics. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of Health Economics* (Vol. 1, pp. 265–344). Amsterdam, The Netherlands: North-Holland.

Jürges, H. (2009). Health insurance status and physician behavior in Germany. *Schmoller's Jahrbuch*, *129*, 297–307.

Kapteyn, A., Smith, J. P., & Van Soest, A. (2007). Vignettes and self-reports of work disability in the United States and the Netherlands. *American Economic Review*, *97*, 461–473.

Kapteyn, A., Smith, J. P., & Van Soest, A. (2009). *Work disability, work, and justification bias in Europe and the U.S.* (Working Paper No. WR-696). Santa Monica, CA: RAND Corporation.

Kapteyn, A., Smith, J. P., Van Soest, A., & Voňková, H. (2011). *Anchoring vignettes and response consistency* (Working Paper No. WR-840). Santa Monica, CA: RAND Corporation.

Katz, L. F., & Meyer, B. D. (1990). The impact of the potential duration of unemployment benefits on the duration of unemployment. *Journal of Public Economics*, *41*, 45–72.

Keane, M. P. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature*, *49*, 961–1075.

King, G., Murray, C. J. L., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, *98*, 191–207.

King, G., & Wand, J. (2007). Comparing incomparable survey responses: Evaluating and selection anchoring vignettes. *Political Analysis*, *15*, 45–66.

Klepinger, D., Johnson, T., & Joesch, J. (2002). Effects of unemployment insurance work-search requirements: the Maryland experiment. *Industrial and Labour Relations Review*, *56*, 3–22.

Knabe, A., Rätzel, S., Schöb, R., & Weimann, J. (2010). Dissatisfied with life but having a good day: time-use and well-being of the unemployed. *Economic Journal*, *120*, 867–889.

Kotlarski, I. (1967). On characterizing the gamma and normal distribution. *Pacific Journal of Mathematics*, *20*, 69–76.

Lalive, R., van Ours, J. C., & Zweimüller, J. (2005). The effect of benefit sanctions on the duration of unemployment. *Journal of European Economic Association*, *3*, 1386–1417.

Lancaster, T. (1990). *The econometric analysis of transition data*. Cambridge, UK: Cambridge University Press.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*, 281–355.

Leu, R. E., & Doppmann, R. J. (1986). Gesundheitszustandsmessung und Nachfrage nach Gesundheitsleistungen. In E. Wille (Ed.), *Informations- und Planungsprobleme in öffentlichen Auf-*

gabenbereichen (pp. 1–90). Frankfurt am Main, Germany: Lang.

Leu, R. E., & Gerfin, M. (1992). Die Nachfrage nach Gesundheit: Ein empirischer Test des Grossman-Modells. In P. Oberender (Ed.), *Steuerungsprobleme im Gesundheitswesen* (pp. 61–78). Baden-Baden, Germany: Nomos.

Levy, H., & Meltzer, D. (2004). What do we really know about whether health insurance affects health. In C. McLaughlin (Ed.), *Health Policy and the Uninsured.*

Llewellyn, D. J., Lang, I. A., Langa, K. M., & Huppert, F. (2008). Cognitive function and psychological well-being: Findings from a population-based cohort. *Age and Ageing*, *37*, 685–689.

Lungen, M., Stollenwerk, B., Messner, P., Lauterbach, K. W., & Gerber, A. (2008). Waiting times for elective treatments according to insurance status: A randomized empirical study in Germany. *International Journal for Equity in Health*, *7*, 1–7.

MaCurdy, T. E. (1981). An empirical model of labor supply in a life-cycle setting. *Journal of Political Economy*, *89*, 1059–1085.

Manning, W., Morris, C., & Newhouse, J. (1981). A two-part model of the demand for medical care: Preliminary results from the health insurance study. In J. van der Gaag & M. Perlman (Eds.), *Economics and health economics.* Amsterdam: North Holland.

Manning, W., Newhouse, J., Duan, N., Keeler, E., Leibowitz, A., & Marquis, M. (1987). Health insurance and the demand for medical care: Evidence from a randomized experiment. *American Economic Review*, *77*, 251–277.

Manski, C. F. (2004). Measuring expectations. *Econometrica*, *72*, 1329–1376.

Mehta, K. M., Yaffe, K., Langa, K. M., Sands, L., Whooley, M. A., & Covinsky, K. E. (2003). Additive effects of cognitive function and depressive symptoms on mortality in elderly community-living adults. *Journal of Gerontology Series A*, *58A*, 461–467.

Meyer, B. D. (1990). Unemployment insurance and unemployment spells. *Econometrica*, *58*, 757–782.

Meyer, B. D. (1995). Lessons from U.S. unemployment insurance experiments. *Journal of Economic Literature*, *83*, 91–131.

Mortensen, D. T. (1977). Unemployment insurance and job search decisions. *Industrial and Labor Relations Review*, *30*, 505–517.

Murray, C. J., Özaltin, E., Tandon, A., Salomon, J. A., Sadana, R., & Chatterji, S. (2003). Empirical evaluation of the anchoring vignette approach in health surveys. In C. J. Murray & D. Evans (Eds.), *Health systems performance assessments: Debates, methods and empiricism* (pp. 369–399). World Health Organization.

Newhouse, J. (1974). A design for a health insurance experiment. *Inquiry*, *11*, 5–27.

Newhouse, J. (1993). *Free for all? Lessons from the RAND health insurance experiment*. Harvard University Press.

Nocera, S., & Zweifel, P. (1998). The demand for health: An empirical test of the Grossman model using panel data. In P. Zweifel (Ed.), *Health, the medical profession, and regulation* (pp. 35–

49). Boston, MA: Kluwer.

OECD. (2006). *Live longer, work longer, ageing and employment policies*. Paris, France: OECD.

Perlroth, D. J., Goldman, D. P., & Garber, A. M. (2010). The potential impact of comparative effectiveness research on U.S. health care expenditures. *Demography*, *47*, S173–S190.

Pohlmeier, W., & Ulrich, V. (1995). An econometric model of the two-part decisionmaking process in the demand for health care. *Journal of Human Resources*, *30*, 339-361.

Quanjer, P. H., Lebowitz, M., Gregg, I., Miller, M., & Pedersen, O. (1997). Peak expiratory flow: Conclusions and recommendations of a working party of the European Respiratory Society. *European Respiratory Journal*, *24*, 2S–8S.

Ried, W. (1998). Comparative dynamic analysis of the full Grossman model. *Journal of Health Economics*, *17*, 383–425.

Riphahn, R., Wambach, A., & Million, A. (2003). Incentive effects in the demand for health care: A bivariate panel count data estimation. *Journal of Applied Econometrics*, *18*, 387–405.

Roodman, D. (2009). *Estimating fully observed recursive mixed-process models with* `cmp` (Working Paper No. 168). Washington, DC: Center for Global Development.

Rosholm, M., & Svarer, M. (2008). The threat effect of active labour market programmes. *Scandinavian Journal of Economics*, *110*, 385–401.

Rothschild, M., & Stiglitz, J. (1976). Equilibrium in competitive insurance markets: An essay in the economics of incomplete information. *Quarterly Journal of Economics*, *90*, 624–649.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Rust, J., & Phelan, C. (1997). How Social Security and Medicare affect retirement behavior in a world of incomplete markets. *Econometrica*, *65*, 781–831.

Salomon, J. A., Tandom, A., & Murray, C. J. (2004). Comparability of self rated health: Cross-sectional multi-country survey using anchoring vignettes. *British Medical Journal*, *328*, 258–260.

Sen, A. (2002). Health: perception versus observation. *British Medical Journal*, *324*, 860–861.

Sirven, N., Santos-Eggimann, B., & Spagnoli, J. (2008). *Comparability of health care responsiveness in Europe using anchoring vignettes from SHARE* (Working Paper No. 15). Paris.

Smith, J. P. (2004). Unraveling the SES–health connection. *Population and Development Review*, *30*(Suppl.), 108–132.

Smith, J. P. (2009). Reconstructing childhood health histories. *Demography*, *46*, 387–403.

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*, 557–586.

Thistlethwaite, D., & Campbell, D. (1960). Regression discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, *51*, 309–317.

Tuit, S., & Van Ours, J. C. (2010). How changes in unemployment benefit duration affect the inflow into unemployment. *Economics Letters*, *109*, 105–107.

Usher, D. (1976). Comments on "The correlation between health and schooling". In N. E. Terleckyj (Ed.), *Household production and consumption* (pp. 212–220). New York, NY: National Bureau of Economic Research. Retrieved July 4, 2011, from http://www.nber.org/chapters/c3962.pdf

Van de Ven, W. P. M. M., & Van der Gaag, J. (1982). Health as an unobservable: A MIMIC-model of demand for health care. *Journal of Health Economics*, *1*, 157–183.

Van den Berg, G. J. (1990). Nonstationarity in job search theory. *Review of Economic Studies*, *57*, 255–277.

Van den Berg, G. J. (2001). Duration models: Specification, identification, and multiple durations. In J. J. Heckman & E. Leamer (Eds.), *Handbook of econometrics* (Vol. V, pp. 3381–3460). Amsterdam, the Netherlands: North-Holland.

Van den Berg, G. J., Bergemann, A., & Caliendo, M. (2009). The effect of active labor market programs on not-yet treated unemployed individuals. *Journal of the European Economic Association*, *7*, 606–616.

Van den Berg, G. J., Deeg, D. J. H., Lindeboom, M., & Portrait, F. (2010). The role of early-life conditions in the cognitive decline due to adverse events later in life. *Economic Journal*, *120*, F411–F428.

Van den Berg, G. J., Lindeboom, M., & Portrait, F. (2006). Economic conditions early in life and individual mortality. *American Economic Review*, *96*, 290–302.

Van den Berg, G. J., & Van der Klaauw, B. (2006). Counseling and monitoring of unemployed workers: Theory and evidence from a controlled social experiment. *International Economic Review*, *47*, 895–936.

Van der Klaauw, W. (2009). Regression discontinuity anlaysis: A survey of recent developments in economics. *Labour*, *22*, 219–245.

Van Doorslaer, E. (1987). *Health, knowledge and the demand for medical care: An econometric analysis*. Assen, Netherlands: Van Gorcum.

Van Doorslaer, E., Koolman, X., & Jones, A. M. (2004). Explaining income-related inequalities in doctor utilization in Europe. *Health Economics*, *13*, 629–647.

Van Ours, J. C., & Vodopivec, M. (2006). How shortening the potential duration of unemployment benefits affects the duration of unemployment: Evidence from a natural experiment. *Journal of Labor Economics*, *24*, 351–378.

Van Soest, A., Delaney, L., Harmon, C., Kapteyn, A., & Smith, J. P. (2011). Validating the use of vignettes for subjective threshold scales. *Journal of the Royal Statistical Society, Series A*, *174*, 575–595.

Varaiya, P. P. (1998). *Lecture notes on optimization*. University of California, Berkeley. Retrieved May 31, 2011, from http://paleale.eecs.berkeley.edu/~varaiya/papers_ps.dir/NOO.pdf (Previously published in 1972 as *Notes on optimization*, Van Nostrand Reinhold, New York.)

Wagstaff, A. (1986a). The demand for health: Some new empirical evidence. *Journal of Health*

Economics, *5*, 195–233.

Wagstaff, A. (1986b). The demand for health: Theory and applications. *Journal of Epidemiology and Community Health*, *40*, 1–11.

Wagstaff, A. (1993). The demand for health: An empirical reformulation of the Grossman model. *Health Economics*, *2*, 189–198.

Wakker, P. P. (2008). Explaining the characteristics of the power (CRRA) utility family. *Health Economics*, *17*, 1329–1344.

Walendzik, A., Gress, S., Manouguian, M., & Wasem, J. (2008). Vergütungsunterschiede im ärztlichen Bereich zwischen PKV und GKV auf Basis des standardisierten Leistungsniveaus der GKV und Modelle der Vergütungsangleichung. *Diskussionsbeitrag No. 165, University of Duisburg-Essen*.

Wansbeek, T., & Meijer, E. (2000). *Measurement error and latent variables in econometrics*. Amsterdam, The Netherlands: North-Holland.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Zweifel, P., & Breyer, F. (1997). *Health economics*. New York, NY: Oxford University Press.

Zweifel, P., Breyer, F., & Kifmann, M. (2009). *Health economics* (2nd ed.). Berlin, Germany: Springer.

Zweifel, P., & Manning, W. G. (2000). Moral hazard and consumer incentives in health care. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of Health Economics* (Vol. 1A, pp. 409–459). Amsterdam, The Netherlands: North-Holland.