ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis [International Bachelor in Economics and Business Economics]

Income and (un)healthy consumption:

Estimating the effect of income on smoking, drinking and sports by exploiting policy changes in the income-dependent combination-reduction (IACK)

Name Student: Mareen Bastiaans

Student ID number: 414723

Supervisor: Dr. J.L.W. van Kippersluis

Second assessor: Prof. Dr. E.K.A van Doorslaer

Date final version: 18 Augustus 2017

Summary

This thesis evaluates the effect of income on smoking, drinking and sports by exploiting policy changes in the income dependent combination reduction ("inkomensafhankelijke combinatiekorting", IACK), a Dutch type of EITC. It thereby adds to the existing, small body of literature on the effect of income on health behaviour with a novel approach. As far as I know, this is the first time the IACK has been used in this type of analysis. Moreover, it is a first empirical test of a model which could potentially explain the socioeconomic status health gradient (SES health gradient).

The existence of the SES health gradient is undisputed, but the causal mechanisms are difficult to uncover (Smith, 1999). Galama and van Kippersluis (2017) propose a model that aims to explain this health gradient. This paper tests one of their predictions, namely that wealthier individuals engage more in healthy and moderately unhealthy consumption than less wealthy individuals, but engage less in severely unhealthy consumption.

The research question of this paper is as follows: *'To what extent does income influence the level of (un)healthy consumption by individuals?'* The research question is tested with panel data from the LISS Panel, covering 7 waves in the period 2008-2015 and consisting of 28,838 observations for 8,657 individuals. A 3-step analysis is used. First, OLS models with and without controls for age, gender, marital status and changes in health are tested. Secondly, FE models with individual and year fixed effects are estimated. Thirdly, policy changes in the IACK are exploited as a source of external variation in income.

This paper finds some evidence in support for the Galama van Kippersluis' (2017) prediction. The coefficient of income on the external margin of alcohol consumption is significant in the FE model. The average marginal effect of a 10% income increase is a 5.88 percentage point increase on the likelihood of alcohol consumption. However, this effect seems to be mainly driven by 16 to 24 year olds, which cannot be explained by income. For this group the average marginal effect of a 10% income increase is a 20 percentage point increase in the probability of drinking. It is difficult to say whether this effect is causal, as the main group that drives the effect is not affected by the IACK policy. Moreover, for other types of unhealthy and healthy consumption no evidence for an effect of income is found.

The main limitations of this thesis are that it cannot successfully use the IACK to exploit external variation in income, and therefore cannot make claims on causality. Moreover, it focuses on three types of unhealthy and healthy behaviour. Whereas its findings suggest that these different types cannot be generalized. Further research should focus on expanding the range of behaviours, isolating external variation and empirically testing other predictions of Galama and van Kippersluis (2017).

Acknowledgments

I would like to take this opportunity to thank my thesis supervisor, Dr. Hans van Kippersluis, for his help and support throughout this process. The many meetings were immensely helpful, they helped me structure the work and keep my spirits up. Moreover, the comments on previous versions and the many questions I got to ask made it able for me to improve my work.

Furthermore, I make use of data of the LISS (Longitudinal Internet Studies for Social sciences) panel administered by CentERdata (Tilburg University, the Netherlands). I am very thankful that I was able to use this data.

Finally, I would like to thank the many people (family, friends, and acquaintances) that have had the pleasure of listening to me talking about my thesis throughout this process. Some more than others. I hope I didn't bore you.

Table of Contents

Chapter 1: Introduction and research question

This thesis describes the relation between income and smoking, drinking and sports by exploiting policy changes in the IACK, a Dutch type of EITC. It thereby adds to the existing literature on this topic with a novel approach and tests a hypothesis from a new model which aims to explain the socioeconomic status health gradient (SES health gradient).

The existence of the SES health gradient is well described in the economic literature (Smith, 1999). The SES health gradient is the strong correlation between socioeconomic status (a combination of among other education, income and wealth) and health status or health outcomes. This gradient is substantial, as Case and Deaton (2005) describe that 20 year old men in the lowest quartile of income report worse health than fifty year old men in the highest quartile. The SES health gradient and particularly the causal mechanisms are of interest to policy makers. The Grossman health capital model (Grossman, 1972a, b) is often used to describe and try to explain the gradient, moreover many theoretical extensions have been made to this model in an effort to better explain the gradient (see for an overview Grossman, 2000). Moreover, a large body of empirical literature exists that looks at the correlations between various dimensions of SES and health in an effort to untangle these causal relations (for an overview, see Cutler, Lleras-Muney and Vogl, 2008). However, as causations between SES and health go both ways and there are other underlying mechanisms that influence both it is hard to establish the causal relations (Feinstein, 1993; Smith, 1999). As Case and Deaton (2005) and Cutler, Lleras-Muney and Vogl (2008) point out, a theoretical framework is needed to explain the SES health gradient and to interpret empirical findings.

Galama and van Kippersluis (2017) aim to provide such a framework. The authors develop a Grossman health capital model in which many additional forms of health related behaviour are incorporated. The goal of the paper is that the SES health gradient will follow as the outcome of rational choices of individuals in this model.

This paper tests one of their predictions, namely that wealthier individuals engage more in healthy and moderately unhealthy consumption than less wealthy individuals, but engage less in severely unhealthy consumption. Galama and van Kippersluis (2017) argue that this is the case as increases in income have an overall positive effect on healthy consumption, but a positive direct, monetary effect and a negative indirect, health effect on unhealthy consumption. The authors argue that the negative health effect increases in the severity of the type of unhealthy consumption, and thus the negative health effect will dominate for severely unhealthy consumption but the positive monetary effect will dominate for moderately unhealthy consumption. The research question of this paper is as follows:

'To what extent does income influence the level of (un)healthy consumption by individuals?'

I consider the effect of income rather than wealth, which is prominent in the prediction. This choice is made because data on income is better available than wealth, moreover it suffers from less reporting issues. Furthermore, income affects wealth as unused income will flow to wealth, thus wealth is largely made up of previous income streams. Therefore, I argue that the choice to focus on income or wealth will not affect the results substantially. This thesis considers three types of behaviour: smoking behaviour and alcohol consumption as forms of unhealthy consumption, sports as a form of healthy consumption. These behaviours are chosen as they can clearly be categorized as healthy or unhealthy, moreover data is available. For all three variables, both the effect on the extensive and intensive margin will be considered. These effects are not necessarily the same, they might capture the difference between the effects on moderately and severely unhealthy consumption as hypothesized above.

The research question is tested with panel data from the LISS Panel, covering 7 waves in the period 2008-2015 and consisting of 28,838 observations for 8,657 individuals. The analysis is a 3-step process in which potential bias is reduced in every step. First, OLS models with and without controls for age, gender, marital status and changes in health are tested. Secondly, FE models with individual and year fixed effects are estimated. Thirdly, an effort is made to isolate exogenous variation in income by exploiting policy changes. The inkomens afhankelijke combinatie korting (IACK)[1], a Dutch type of EITC for certain working parents, is used as a source of exogenous policy changes and the type of analysis used is inspired by Gruber and Saez (2002). To my knowledge, this paper is the first that uses IACK as a source of external variation in income.

The main contributions of this thesis are twofold. Firstly, as emphasized before, it tests a prediction of Galama and van Kippersluis' (2017) model. This model could potentially explain the SES health gradient. It is important that such theoretical models are tested on their empirical validity, this paper aims to do so by serving as a first empirical test of the model. Secondly, it adds to the existing knowledge of the effect of income on smoking, drinking and sports. A large part of empirical literature looks at the effect of income and other dimensions of the SES on health outcomes. Far fewer focus on the effect on health behaviour rather than health outcomes. Health behaviour does deserve specific attention. The effect on health behaviour does not have to be the same as the effect on health, as health behaviours translate imperfectly into health outcomes or delayed. Furthermore, the approach

[1] Loosely translated: income dependent combination reduction, a more detailed discussion of the policy follows in the background chapter.

of this paper is different as it uses the IACK to exploit external variation in income, rather than macroeconomic variables, inheritances and lottery winnings or instrumental variables such as work experience as was used in the existing literature. This paper thus adds to a small body of literature with a novel approach in this field.

This paper finds some evidence in support for the Galama van Kippersluis hypothesis. The coefficient of income on the external margin of alcohol consumption is significant in a FE model with individual and year fixed effects and controls for age, gender, marital status and change in health. The average marginal effect of a 10% income increase is a 5.88 percentage point increase in the probability of drinking. The effect seems to be mainly driven by 16-24 year olds. For this group the average marginal effect of a 10% income increase is a 20 percentage point increase in the probability of drinking. However, these individuals are not affected by the IACK policy and therefore policy changes in IACK cannot be used to test the causality of this effect. Moreover, for other types of unhealthy and healthy consumption no evidence for an effect of income is found. Overall, there is thus limited support for the hypothesis. There is thus evidence for an effect of income on one dimension of unhealthy consumption, and in line with the direction Galama and van Kippersluis' model suggests. However, the other types of unhealthy and healthy consumption seem to be unaffected by income, this is not in line with Galama and van Kippersluis' hypotheses. This highlights that different types of unhealthy and healthy consumption cannot be generalized.

The rest of this thesis is as follows. In chapter 2 the relevant background to this paper is described: the theoretical and empirical literature and the IACK policy. In chapter 3 the data is discussed, followed by an explanation of the methodology in chapter 4. The fifth chapter covers the results. In chapter 6 these results will be discussed. Finally, chapter 7 forms the conclusion.

Chapter 2: Theoretical and empirical background and policy setting

This chapter describes the relevant background to this thesis. First, theoretical literature is discussed to describe a theoretical framework in which the results can be embedded. Secondly, empirical literature is described followed by the contributions of this paper. Lastly, the IACK policy which is used as a source of external variation in income is described.

2.1 Theoretical models and concepts

This discussion of theoretical models starts with a discussion of the Grossman model and some of its extensions. This model is the standard model for the discussion of health and health behaviour. Moreover, it serves as a basis for the Galama and van Kippersluis paper from which I will test a prediction and which will be discussed here subsequently. Next, I will move on to discuss some other views on the relation between the SES or income and health. Based on this, I form my hypotheses.

2.1.1 The Grossman Health Capital Model of the Demand for Health

Grossman (1972a,b;2000) proposed a health model inspired by human capital models with a new view of health: consumers demand 'good health' rather than medical care directly. Health can be seen as capital stock that decreases over time. To consumers it is both a consumer good, as it provides direct utility (or, the disutility of being sick), and an investment good, as it determines the time available for other market and non-market activities.

In the model, individuals are born with a stock of health. This stock depreciates over time by an exogenous depreciation rate. However the stock can be increased through investments. If the stock of health falls below a certain threshold death occurs. Individuals thus choose the length of their life through investments. Investments in health stock are produced by household production functions, the inputs are the time of the individual and market goods (e.g. medical care, diet). Furthermore the production function depends on 'environmental variables' such as education; education increases the efficiency of production. The optimal stock of health capital at any age is determined by equating marginal efficiency of this capital, the marginal benefit, to its user cost in terms of the prices of gross investment, marginal costs.

Grossman (1972a) predicted that the more educated would have more health stock, as they are more efficient health producers and thus have lower health costs. The effect of age on the demand would depend on assumptions on the role of age on the depreciation rate. Furthermore an increase in income would loosen the budget constraint in this model and thus allow for more investments.

The Grossman model has become the standard model in health economics and has been extended in many ways. It is not the purpose of this study to discuss these extensions (for an overview of

theoretical extensions to the model see e.g. Grossman, 2000). Rather, I will focus on the model proposed by Galama and van Kippersluis (2017), as the purpose of this paper is to test one of their predictions.

This paper tests a prediction of Galama and van Kippersluis (2017) construct a life-cycle model that incorporates medical care and other types of (health) behaviour which could help explain the SES-health gradient. They are not the first to develop a life-cycle Grossman model (see e.g. Wagstaff 1986; Ehrlich and Chuma 1990). Nor are they the first to develop predictions, for example Ried (1998) have done so with comparative dynamic analysis. However, what is unique is that they include many different types of behaviours. The goal of the paper is to construct a model that explains the SES-health gradient as a result of rational choices.

The demand for health in their model is based on Grossman's model with some extensions. First, they allow for decreasing returns to scale in the health production process (as done by Galama, 2015). Moreover, based on a review of the literature of multiple disciplines they add additional behaviours and corresponding choices that are relevant to the SES-health gradient. The most relevant addition to the model for this paper, is that they allow consumption to affect the deterioration rate. Consumption which decreases the deterioration rate of health capital is labelled healthy consumption, and consumption which increases the deterioration rate of health capital is labelled unhealthy consumption. Furthermore, they perform comparative dynamics analysis to generate predictions regarding the relation between SES dimensions and health.

The prediction that this paper aims to test is that wealthy individuals engage in more healthy and moderately unhealthy consumptions but engage less in severely unhealthy consumption. Central to this prediction is Galama and van Kipperluis' (2017) concept of 'relative marginal value of health'. The relative marginal value of health is the marginal benefit of health investment. It is defined as the marginal value of health divided by the marginal value of wealth. It captures the idea that there is a trade-off between health and wealth. Moreover, the first conditions of healthy and unhealthy consumption in the Hamiltonian that is used to solve their model are relevant.

The first condition of healthy consumption is formed by the marginal monetary cost of healthy consumption and the marginal benefit of healthy consumption. The former is the direct monetary cost of healthy consumption, it is a function of the price of healthy consumption and services and the opportunity cost of time. The marginal benefit of healthy consumption is the marginal value of the health saved by the healthy consumption. It is a product of the relative marginal value of health and the 'amount' of health that is saved by the consumption.

The first condition of unhealthy consumption depends on two types of costs. First the direct monetary costs of unhealthy consumption, these are similar to the direct monetary costs of healthy consumption. The second is the marginal value of health lost by unhealthy consumption. This is similar to the marginal benefit of healthy consumption.

Galama and van Kippersluis (2017) form their hypothesis as follows. First, Galama and van Kippersluis proof that in their model wealthy individuals live longer and that they value health more and are healthier at all ages. Secondly, the authors discuss the effect of wealth on healthy and unhealthy consumption.

For healthy consumption there is a direct and indirect effect. The direct effect is that an increase in wealth leads people to be able to afford more healthy consumption. The indirect effect is that the increase in wealth increases the relative marginal value of health and thus the health benefit of healthy consumption. These effects thus point in the same direction; the overall effect of wealth on healthy consumption is positive.

The effect of wealth on unhealthy consumption it is more complex. The direct effect is again positive: people can afford more unhealthy consumption after an increase in wealth. But the indirect effect is negative. An increase in wealth increases the relative marginal value of health which in turn increases the health cost of unhealthy consumption. These two effects work in the opposite direction, the effects are competing. Therefore it is unknown what the overall effect will be.

Galama and van Kippersluis point out that the health cost of unhealthy consumption increase in the severity of the unhealthy consumption's impact on health. They therefore hypothesise that for moderately unhealthy goods the direct wealth effect would dominate, while for severely unhealthy goods the indirect wealth effect would dominate. This leads them to the following prediction: *"Wealthy individuals shift consumption toward healthy consumption: they consume more healthy and moderately unhealthy consumption goods and services, but fewer severely unhealthy consumption goods and services"* (Galama and van Kippersluis, 2017, p. 24).

2.1.2 Alternative theoretical views and models

It is worth mentioning that this is not the only possible way to view the relation between income and unhealthy/healthy consumption. On a broader level, there are three categories of theories that explain the SES-health gradient, and the related correlation between income and health. First, those that theorize a causal effect from SES on health. Second, those that theorize a causal effect from health on SES. Finally, those that do not point to a causal effect, but rather to a confounding variable that truly drives the correlation between SES and health.

The prediction discussed here, falls in the first category. An example of the second category is that poor health affects labour market opportunities and thus income. This could because poor health leads to labour force withdrawal (e.g. Case and Deaton, 2005), because people in good health are more productive (Currie and Madrian, 1999) or effects of health as a foetus or child on later income (Case, Fertig and Paxson, 2005). These theories expect that rather than an effect of income on health behaviour, health behaviour will affect health and therefore labour market opportunities and income. An example of the third category is Fuchs (1982). Fuchs (1982) argues that the relationship between schooling and health is driven by time preference. A lower time preference leads to more schooling, more healthy behaviour and better health, as both these behaviours require that one is willing to put off current benefits for future benefits. Either time preference is established beforehand, or schooling affects time preference and thereby also leads to more investments in time (see also Becker and Mulligan, 1997 for a model of endogenous time preference). This would thus not expect an effect of income on health.

2.1.3 Hypotheses

There are thus three broad categories of theories that explain the SES health gradient differently. Moreover, the theories predict a different effect for income on health. Theories in the first category will predict such an effect, but theories in the second and third categories would not. However, these three theories are not mutually exclusive. As Smith (1999) points out, it is not a question of which of the three broad categories of theories is valid, but rather to what extent each category attributes to the overall SES health gradient.

This paper focuses on the first category, the effect of the SES on health. In particular, the effect of income on healthy and unhealthy consumption. Based on the prediction described above by Galama and van Kippersluis (2017), I hypothesise the following:

Hypothesis 1: Income has a positive effect on healthy and moderately unhealthy consumption.

Hypothesis 2: Income has a negative effect on severely unhealthy consumption.

2.2 Empirical literature

There exists a large body of literature related to the SES health gradient. It is not possible to describe all of it here, nor would this be relevant to this paper. In this literature review I first give a short overview of the general literature on the SES and health. Afterwards, I move on to describe the most relevant work for this research. These are the papers that try to establish the causal effect of income on health behaviours rather than income, this part of the literature review relies heavily on the overview given by van Kippersluis and Galama (2014).

2.2.1 Empirical literature on the SES health gradient and health outcomes

The existence of the SES health gradient is widely accepted in the literature and described (Feinstein, 1993; Smith, 1999). However, much of this literature is focused on correlations and cannot successfully describe the causal relations (Feinstein, 1993). This is caused by econometric difficulties in separating the three possible channels through which the SES health gradient can arise, which were described before, and which can operate simultaneously (Smith, 1999).

A large part of the literature on the SES health gradient tries to uncover these causal relations between health outcomes and the SES, or dimensions of the SES. Cutler, Lleras-Muney and Vogl (2008) provide an overview of the empirical literature related to four dimensions of the SES, education, financial resources, race and ethnicity, and rank, and health. The authors desribe that strong associations between education and health are found in the literature, but it is unknown which part is causality and which part reverse-causality or due to omitted factors. Financial resource seem to have an effect on health in early stages of life, foetal and child, but afterwards it is mainly health which affects financial resources rather than the other way around. Others have emphasized the importance of the reverse causality from health to income, mainly through labour market opportunities and working hours (Currie & Madrian, 1999; Smith, 2004, 2005; Case and Deaton, 2005). There has also been work that tries to estimate the effect of income on various health outcomes (see e.g. Deaton and Paxson, 2001). Moreover, there have been attempts to use external variation in income, for example by using lottery winnings (Lindahl, 2005). This literature is to some extent relevant to this paper. As it sets the general effect of income on health (outcomes).

2.2.2 Empirical literature on income and health behaviour

The literature described in the previous section is relevant to the paper as it sets the background of the general effect of income on health (otucomes). However, this paper focuses its attention on the effect of income on health behaviour rather than health outcomes, this is not necessarily the same. Health behaviour might not directly or might imperfectly translate into health outcomes. Moreover, it is interesting to note that this branch of literature is relatively small compared to the literature review concerned with health outcomes.

Ettner (1996) estimates the effect of income on health and some types of health behaviour by using instrumental variables. The instrumental variables used are state unemployment rates, work experience, parental income and spousal characteristics. The findings are a positive effect on mental and physical health, but also a positive effect on alcohol consumption. However, the use of this instrumental variables has been heavily criticized (e.g. Kawachi, Adler & Dow 2010). Adda, Banks and

van Gaudecker (2009) estimate the effect of permanent income innovations on health. The authors do so by exploiting structural and exogenous changes in the incomes of cohorts. The authors find little effect for an effect on health, but a positive effect on risky behaviour (smoking and alcohol) and mortality. Kim and Ruhm (2012) use inheritances to estimate the effect of wealth on health related behaviour. The authors compare big and small inheritances and conclude that bigger inheritances have a positive effect on alcohol consumption, but find no effect on smoking or exercising. Apouey and Clark (2015) use lottery winnings to estimate the effect of wealth shocks on physical and mental health. The authors find a positive effect of larger winnings on the likelihood of smoking and the probability of (social) drinking.

Van Kippersluis and Galama (2013) propose a theory of unhealthy consumption which is later incorporated in Galama and van Kippersluis (2017), the paper that serves as a basis for this thesis. The authors describe a model with a negative health effect of unhealthy consumption and thus a direct and indirect effect of income on unhealthy consumption. The authors empirically test the prediction by using inheritances and lotteries as sources of external variation. A positive effect of both lotteries and inheritances is found on the prevalence of drinking, but no effect on the number of drinks or heavy drinking. For smoking, a positive effect on the prevalence and the number of cigarettes is found for inheritances, but not for lottery winnings.

Another approach to the effect of income on health behaviour, is to focus on the effect of macroeconomic changes in income on health. Ruhm (2000) finds evidence that health improves in economic downturns. This could be explained by a decrease in working hours (Ruhm, 2005). Moreover, smoking increases during a boost (Ruhm, 2000) and decreases in a recession (Ruhm, 2005). Furthermore, excess weight decreases and leisure activities increase during a downturm (Ruhm, 2005).

As in the literature regarding income and health, there are reverse causality concerns for income and health behaviour. Van Ours (2004) synthesises the existing research for the effect of smoking and alcohol on income and estimates the effect of both on wages. The findings are that there is a positive wage effect of alcohol and a negative wage effect of smoking, of equivalent size.

Finally, there also exists literature which focuses on smoking behaviour and alcohol consumption. Cook and More (1999) give an overview of the empirical literature on alcohol consumption, however this literature focuses on the specific characteristics of alcohol consumption such as its social context, delayed consequences and habit formation rather than the effect of income, which is the interest of

this paper. Chaloupka and Warner (200) provide a similar overview for smoking behaviour. However, this literature is mainly concerned with price elasticity and smoking in relation to taxes. To my knowledge, such an overview does not exist for physical activity or sports. There are some papers that relate physical activity and income, however these are mainly concerned with the effect of physical activity on income (see e.g. Lechner, 2009).

2.2.3 Contributions of this paper to the empirical literature

There are three main contributions of this paper to the existing empirical literature. Firstly, the body of literature on the effect of income on health behaviour is relatively small. Therefore it is already valuable that this paper estimates this effect as a check of previous findings. Secondly, this thesis takes a different approach to establish the effect compared to existing papers. Previous work has used macroeconomic variables, lotteries and inheritances and instrumental variables such as work experience to isolate the effect of income. This paper uses policy changes in the IACK, a Dutch type of EITC, to isolate exogenous variation in income. As far as I know, this paper is the first to use the IACK in such a way. The EITC has been used to estimate the effect of income on smoking, but these papers mainly focus on maternal or female smoking (see e.g. Averett & Wang, 2013; Kenkel, Schmeiser & Urban, 2014 is an exception to this). Finally, most papers focus on the effect of income on unhealthy behaviour. This paper examines the effect of unhealthy and healthy behaviour.

2.3 Relevant policy setting: the income-dependent combination-reduction (IACK)

The IACK will be used as a source of external variation in income in this paper. Here, I will discuss this policy in some detail to serve as background knowledge for the rest of the paper. IACK stands for income dependent combination reduction ('inkomensafhankelijke combinatiekorting' in Dutch). The IACK is a 'heffingskorting' in the Dutch tax system. This is an amount of money that can be deducted from the payable tax. The IACK is aimed to stimulate parents of young children to work, especially single parents and parents that have a working partner. For these parents, starting to work or increasing the number of hours of work, usually implies that they need day-care or after school care for their children. Thus there are high opportunity costs of working that can act as a barrier to work. By giving these parents a tax benefit if they work, the government aims to stimulate working among them.

To receive IACK an individual must meet the following conditions: have at least one child under 12; and work; and be a single parent or earn less than his/her partner. The amount of IACK received dependents on a number of variables. First, the income of the individual over period t, P_{it}. Second a number of variables that are set each year by the government: the minimum and maximum amount

of IACK, s_t^{min} and s_t^{max}; the upper and lower bound to income, B_t^H and B_t^L; the percentage of income received as IACK, τ_t. The scheme is as follows (see equation 1). The amount of IACK received is zero if the income is below or equal to a lower bound[2] and the maximum amount if the income is above an upper bound. If the income is above the lower bound, but below or equal to the upper bound, the amount received is equal to the minimum amount plus a percentage of the income in excess of the lower. This thus creates a scheme with two flat parts and a part that connects the two, linearly increasing in income.

$$S_{it}\begin{cases} 0 \; if \; P_{it} \leq B_t^L \\ s_t^{min} + \tau_t * (P_{it} - B_t^L) \; if \; B_t^L < P_{it} \leq B_t^H \\ s_t^{max} \; if \; P_{it} > B_t^H \end{cases} \tag{6}$$

[2] Some individuals do receive an amount even though their income is below the lower bound. This is the case if they are eligible for another tax deduction, the 'zelfstandigenaftrek' (a tax deduction for entrepeneurs/self-employed). I ignore this condition, because it is difficult to construct and would only affect a small part of the sample.

Chapter 3: Data

3.1 Data source and dataset

The data that is used in this study comes from the LISS (Longitudinal Internet Studies for the Social Sciences) panel administered by CentERdata (Tilburg University, The Netherlands). The LISS panel consists of 7000 individuals, forming 4500 households. The population is based on a true probability sample of households drawn from the population register by Statistics Netherlands. All data is self-reported; panel members complete online questionnaires monthly for financial compensation.

This study makes use of data from three core studies: Health, Family and Household, and Social Integration and Leisure, which are conducted annually as well as background variables that are updated on a monthly basis. These surveys are not conducted at the same time. [3] A particular issue with this is that the Social Integration and Leisure survey, containing the dependent variables for sports, is conducted at the beginning of the year and would therefore precede independent variables; this could lead to reverse causality. To prevent this, lags of the variables in the Social Integration and Leisure survey are included; effectively, the Health and Family and Household surveys are combined with the Social Integration and Leisure survey of the following year. Furthermore, background variables are used from June. June was chosen in line with the Income survey which would originally contain the income variables (see discussion of this later in this section), and it also ensures that the independent variables are recorded before or at the same time as the dependent variables; again to prevent reverse causality.

The panel data set covers the period 2008-2015, and consists of 7 waves. A wave for 2014 is absent, as the Health survey was not conducted in this year. All panel members 16 years and older are asked to complete the core studies, but not all do. To ensure a consistent sample across specifications, only observations that contain all the dependent and independent variables in a given year are kept. Moreover, observations with negative values for income are removed.[4] After this data selection and cleaning process, the dataset consists of 29,838 observations for 8,657 individuals. The dataset is unbalanced as not all individuals complete all surveys in each year. On average, the number of waves per individual is 3.4.

A sub-sample of this dataset is used in the analysis exploiting policy changes in the IACK, a policy that has started in 2009. The IACK is a Dutch type of EITC that is meant to stimulate working among young parents. The details of this policy will be discussed in the methodology section. The dataset is

[3] A table with the specific months for the first and second measure per survey that was included can be found in the Appendix, Table 7.

[4] In the case of income, specific negative values are used to indicate that a participant does not know their income or is unwilling to share it.

restricted to IACK recipients: working parents, with at least one child under 12 and who are either single or the least earning parent. Data necessary for IACK recipient identification comes from the Family and Household survey and the background variables. The type of analysis used requires that an individual is identified as an IACK recipient in two consecutive years as well as that dependent and independent variables are available in two consecutive years. This also excludes the 2015 wave as 2014 or 2016 is not available. As these restrictions already lead to a small sample, I do not impose the restriction of a consistent sample as this would further reduce the number of observations and thus would lead to a loss of power. The number of observations varies between 604 and 760 per specification, from 287 to 343 individuals. Again the sample is unbalanced.

3.2 Variables

The dependent variables are three types of (un)healthy consumption: smoking behaviour and alcohol consumption for unhealthy consumption and playing sports for healthy consumption. These are chosen because they are clearly unhealthy or healthy respectively, and data is available. For all three a measure of the extensive and intensive margin is used, as the effect of income might differ in these dimensions.

The smoking variables come from the Health survey. A binary variables that indicates whether someone is currently smoking is used for the extensive margin. For the intensive margin, a variable is constructed that sums the number of cigarettes, pipes and cigars smoked per day into one variable containing the number of "smokes" per day. Data on alcohol consumption also come from the Health survey. The extensive margin is less obvious than for smoking, a timeframe needs to be chosen here. A binary variable that indicates alcohol consumption in the past seven days is used for the extensive margin, as this is seen as more relevant than alcohol consumption in the past year. A variable that captures the average number of drinks per week or month would be ideal for the intensive margin, but is not available within the LISS panel. Rather the number of drinks consumed on the day with the most alcohol consumption in the past seven days, the "peak-day", is used. A limitation of this variable is that it ignores the frequency of this drinking behaviour, no difference is made between drinking 5 units once a week or every day. The number of days with alcohol consumption in the last seven days is another option, but this would ignore the number of drinks consumed. Consumption on peak-day was chosen for the intensive margin, as this is seen as less flawed and capturing a more important dimension of alcohol consumption. The sports variables are obtained from the Social Integration and Leisure survey. The extensive margin is a binary variable that indicates whether someone engages in sports activities. The intensive margin is a variable that consists of the number of hours someone engages in sports activities per week.

The independent variable is income. Data on income can be collected from the Income survey and the background survey. The Income survey contains very rich data, but this makes the construction of an income variable complex. The background variables survey contains ready-made income variables. These are easier to use, contain less risk of making a mistake and complex enough for the goal of this research. Net income is chosen instead of gross income, as this is closer to the disposable income that is hypothesized to influence (un)healthy consumption and thus will be more able to capture this effect, if it exists. The distribution of net income has a long right-side tail. To prevent these observations from biasing the results, a log transformation is used. This concentrates the range of observations. Therefore the regression will be a log-linear model and thus the regression coefficients will be semi-elasticities. One is added to net income before taking the log, this prevents losing observations with zero income. The independent variable is thus constructed as follows:

$$\mathrm{x} = \ln(net\ income + 1)$$

Furthermore, control variables are included in the models. Individual and year fixed effects are eventually added to the specification, therefore the controls can be limited to time-varying, individual factors. This is why for example education is not included, although this is obviously a confounding variable. Age, gender, marital status and change in health are included. Health status is a problematic variable, it is both a confounding factor and a mechanism. I should only control for the confounding part; if I would control for the mechanism part this would lead to an underestimation of the effect of income. To do so, a variable from the Health survey is used that asks participants to compare current health to their health a year ago. As this captures the change in health from period t-1 to period t, it cannot be influenced by income at t. Therefore it is a way to control only for the potentially confounding part and not for the mechanism part. As a robustness test I have also tested the models without change in health and did not find substantial differences between the estimated coefficients (see Appendix, Table 8). The variable is a categorical variable with the following categories: considerably poorer, somewhat poorer, the same, somewhat better and considerably better. The categories are included as dummies with “the same” as base category. Age is included in years in the standard specifications, but in six categories for the specification with interaction effects for age and income[5]. Gender is a dummy variable that indicates female. Marital status is a categorical variable with the following values: married, separated, divorced, widower or widow and never been married. These categories are included as dummies, with never been married as reference category.

For the second part of the analysis, IACK-recipients must be identified and the IACK amount constructed. The policy was explained in the background chapter, and the exact identification of

[5] The age categories Statistics Netherlands uses are chosen: 1) 14 years and older, 2) 15-24 years old, 3) 25-34 years old, 4) 35-44 years old, 5) 45-54 years old, 6) 55-64 years old, 7) 65 years and older. The sample contains 16 year-olds and older, thus the first category is not used and the second contains 16-24 year-olds.

recipients will be discussed in the methodology chapter. The necessary information is obtained from the Family and Household and background survey. Birth years of children are used to construct a dummy indicated that an individual has a child under 12, deceased children are not considered. Occupation and marital status are used to obtain work and partner status. Partner's income is obtained by subtracting the individual's income from the household income.

3.3 Summary and descriptive statistics

Table 1 Summary statistics of complete sample

	Mean	Standard deviation	Minimum	Maximum	N
Smoking (extensive)	0.188	0.391	0	1	29412
Smoking (intensive)	2.320	5.860	0	60	29412
Alcohol (extensive)	0.725	0.447	0	1	29412
Alcohol (intensive)	2.676	3.366	0	66	29412
Sports (extensive)	0.544	0.498	0	1	29412
Sports (intensive)	2.186	3.310	0	157	29412
Income	1566.766	3997.102	0	285759	29412
Log of income	6.443	2.324	0	12.563	29412
Age	49.858	17.257	15	95	29412
Female	0.516	0.500	0	1	29412
Marital status					29412
Married	0.582	0.493	0	1	29412
Separated	0.004	0.059	0	1	29412
Divorced	0.087	0.282	0	1	29412
Widow or Widower	0.051	0.219	0	1	29412
Never been married	0.277	0.447	0	1	29412
Health change					29412
Considerably worse	0.014	0.118	0	1	29412
Somewhat worse	0.158	0.365	0	1	29412
The same	0.694	0.461	0	1	29412
Somewhat better	0.111	0.315	0	1	29412
Considerably better	0.022	0.147	0	1	29412 [i]

[i] Data is a panel data set, data comes from the following LISS Panel surveys: Health, Family and Household, Social Integration and Leisure and background, and covers the time period 2008-2015 (with the exception of 2014). Sample is restricted to observations that contain all independent and dependent variables in a given year. Smoking (extensive) is a binary variable that indicates some currently smokes, smoking (intensive) indicates the number of cigarettes, cigars and pipes smoked per day. Alcohol (extensive) is a binary variable indicating alcohol consumption in the past seven days. Alcohol (intensive) the units of alcohol consumed on the day with most alcohol consumption in the past seven days. Sports (extensive) is a binary variable indicating engaging in sports activities in a week, sports (intensive) indicates the number of hours per week spend on playing sports. Summary statistics are made on the level of observations rather than individuals.

Summary statistics give an overview of the demographic characteristics of the sample as well as the extent to which smoking behaviour, alcohol consumption and sports behaviour are present within the sample (Table 1). The sample consists of slightly more observations for women than men (51.6%), aged between 15 and 95, on average almost 50 years old. The categorical values are considered for all years together, and therefore do not add up to 1. Income is between zero and 285,759 euros per month, on average 1.567 euros per month with a standard deviation more than 2.5 as large, which further motivates the choice to use a log transformation. The descriptive statistics give the following indication of the extent of smoking, alcohol and sports. Almost 20% of the observations smoke, individuals smoke on average 2.32 units per day (this includes non-smokers) with a maximum of 60 per day. 72.5% of the observations had consumed alcohol in the past week, individuals drink on average 2.7 units of alcohol on peak-day (again, this average includes non-drinkers). Slightly more than half (54.4%) engage in sports activities, the average number of hours of sports individuals play per week is 2.2.

Table 2 Correlation matrix of Log of income and Dependent variables

	Log of Income	Smoking (extensive)	Smoking (intensive)	Alcohol (extensive)	Alcohol (intensive)	Sports (extensive)	Sports (intensive)
Log of Income	1						
Smoking (extensive)	0.0082	1					
Smoking (intensive)	0.0245	0.8231	1				
Alcohol (extensive)	0.0954	0.0539	0.0345	1			
Alcohol (intensive)	-0.0042	0.1315	0.1086	0.4903	1		
Sports (extensive)	-0.0192	-0.1451	-0.1618	0.0709	0.077	1	
Sports (intensive)	-0.0628	-0.0941	-0.1031	0.0519	0.0863	0.6049	1

[ii]

[ii] Data is for 2008-2015 (without 2014) and obtained from the LISS Panel. The data set contains 29,412 observations.

Two types of descriptive statistics are used to get a first indication of the relation between income and smoking, drinking and sports: correlations, which only allow for a monotonic relation, and bar charts of the means by income quintile, which also allows for non-monotonic relationships.

Overall, the correlations are low, thus indicating a weak relationship (Table 2). The correlations for smoking behaviour are particularly low: 0.0082 for the extensive margin of smoking and 0.0245 for the intensive margin. Both are positive, thus suggesting that high income is associated with a higher probability of being a smoker and more smoking units. The correlation between log of income and the extensive margin of alcohol consumption is the highest (0.0954), the correlation with the intensive margin is considerably smaller (-0.0042). Furthermore, the correlation between log of income is

positive with the extensive margin of alcohol consumption, but negative with the intensive margin. This suggests that high incomes are more frequently drinkers, but drink less on peak-day. The correlations between the dimension of playing sports and log of income are both negative and small. Furthermore, the results indicate low correlations between the different types of (un)healthy consumption, which is in line with the existing literature (see e.g. Cutler and Glaeser, 2005).

These correlation can only capture monotonic associations between income and smoking, drinking and sports. It is possible however, that some of these relationships are non-monotonic. In that case, the effect of income is not constant across the income distribution, rather it changes sign or there is only an effect for parts of the income distribution. By constructing bar charts of the mean of the variables of (un)healthy consumption by income quintile, we can see if there are indications of such a pattern. The quintiles are constructed by dividing the observations in 5 quintiles based on net income.

There is a non-monotonic pattern for smoking (Figure 1, Appendix). The frequency of smokers jumps from the 1st (lowest) income quintile to the second quintile. From the second to the third, fourth and fifth quintile the proportion of smokers decreases, but only in the fifth quintile is the proportion lower than in the first. A similar pattern is found in the number of cigarettes, cigars and pipes used per day. An initial jump from the first to second quintile, followed by a drop. This is an interesting pattern, as it suggests that there may be a non-monotonic relation between smoking and income: the lowest income quintile smoke the least often and the fewest cigarettes, the second income quintile the most, but following quintiles again fewer.

The proportion of alcohol-consumers increases for each income quintile and the increase is similar for each move (Figure 2, Appendix). The number of alcoholic drinks shows a different pattern. The number of drinks decreases from the first quintile (2.7) to the second quintile (2). But then increases for the subsequent income quintiles. The average number of drinks is higher in the fourth (2.9) and fifth quintile (3.1) than in the first. Thus, these charts suggest a monotonic positive association between income and the extensive margin of alcohol consumption, but a non-monotonic association between income and the intensive margin.

The proportion of people playing sports and the number of hours played exhibit a similar pattern (Figure 3, Appendix). There is a drop from the first income quintile to second, but an increase for the subsequent income quintiles. This thus suggests a non-monotonic relation.

What is interesting is that in all cases of a somewhat non-monotonic association, the difference between the first and second quintile is different from the differences between the subsequent income quintiles. This suggest a ‘specific difference’ between the first and second income quintile.

But this may be related to other characteristics rather than income, such as education and age. However, these bar charts do not indicate whether the differences between these income quintiles are significant.

Chapter 4: Methodology

The goal of this paper is to analyse the effect of income on (un)healthy consumption, represented by smoking behaviour, alcohol consumption and playing sports. This is done in three steps that increase in the ability to isolate the effect of income on the variables: an OLS regression of income on the dependent variables without and with controls, a FE model with individual and year fixed effects and finally exploiting external changes to income due to policy changes in the IACK. The first two parts use the complete sample, the third part only uses a subsample of the population. I will also repeat the first two analyses on this subsample.

4.1 Regression analysis

The first part of the analysis consists of OLS regressions of income on the six dependent variables. The dependent variables are numerical variables for the intensive margins, and binary variables for the extensive margins. For binary variables, typically probit or logit models are used as these are more apt to estimate probabilities. However, eventually fixed effects are added to the model. This is possible in a logit model, but complicated. Moreover, average marginal effects are fairly similar to the coefficients in an OLS model, so these can be used.

Firstly, basic regressions of the dependent variable on the independent variables are run without any control variables, estimating equation (2).

$$y_{i,t} = a + \beta I_{i,t} + \epsilon_{i,t} \quad (2)$$

y_{it} is the dependent variable for individual i at point t; the intensive or extensive margin of smoking, alcohol consumption or sports. a is the intercept on the y-axis, and thus the estimated level of (un)healthy consumption for an individual without income. In this model, this is assumed to be time invariant and consistent across individuals. The independent variable, I_{it}, is the log of income of individual i at point t in time. The effect of income on the dependent variable is estimated by β. As the independent variable is a log variable and the dependent variable linear, this model is thus a log-linear model. Therefore β captures the effect of a percentage change in income on the absolute level of the dependent variable; a semi-elasticity. This effect is assumed to be time-invariant. Furthermore, ϵ_{it} is the error term. As there are no controls in this model, the error term is likely to be correlated with the independent variable, because of omitted confounding variables such as age and gender. Therefore the estimate for β is biased and unreliable. It does not give a causal effect, but rather captures the correlation between the dependent variable and the log of income. This model is included as starting point for the analysis, since it gives an insight into the basic associations between the variables.

The second model that is estimated is an expansion of the first model, but with the inclusion of a set of control variables.

$$y_{it} = a + \beta I_{it} + \gamma \boldsymbol{\chi}_{it} + \epsilon_{it} \quad (3)$$

$$y_{i,t} = a + \beta I_{i,t} + \gamma_1 Age + \gamma_2 Female + \sum \gamma_k \sum Marital\ status_{i,t} + \sum \gamma_k \sum Change\ in\ health\ status + \epsilon_{it}$$

The control variables that are added are the numerical variable age, the dummy variable gender (1 for female) and a set of dummy variables for marital status and change in health. The reference categories for marital status and change in health are “never been married” and “the same” respectively. The control variables are indicated in (3) by the vector χ. These categorical variables are included as dummies, because the effect from one category to the next is not assumed to be constant. Furthermore, the effects of the control variables are assumed to be time invariant and independent of the level of income. The control variables control for the factors that in equation (2) could be correlated with both the error term and the independent variable and thus leading to biased results. This bias is reduced by the inclusion of the control variables. Thus the estimates of the second model are more reliable. However, factors that are not included can still bias the estimated coefficients. Therefore these results can indicate an association, but cannot be interpreted as a causal effect.

In the third model, time-invariant, individual specific factors and year shocks are controlled for by including individual and time fixed effects.

$$y_{it} = a_i + \beta I_{it} + \gamma \boldsymbol{\chi}_{it} + \delta_t + \epsilon_{it} \quad (4)$$

The individual fixed effects are a_i and replace the common intercept α in the previous models. The individual fixed effects capture time-invariant, individual-specific factors that influence (un)healthy consumption, examples of this are education and family background. By including these individual fixed effects, only within-individual variation in income, dependent and independent variables is used to estimate the coefficients. This reduces the potential bias of omitted variables. However, it ignores across-individual variation and is therefore not suitable if there is limited within-individual variation. This is not the case in this study, within-individual variation in income is an interesting source of variation. Furthermore, time fixed effects are included by including time dummies, δ_t. These capture year-specific factors that affect (un)healthy consumption across individuals. An example is the smoking ban in bars and restaurants that was put in place in October 2014. By including the two types of fixed effects, the time-invariant, individual-specific factors and year shocks are captured and thus their bias is removed from the estimates. The only remaining bias can come from time-varying, individual-specific factors. The control variables that were present in model (3) are still present in

model (4) to capture these effects. However, it should be noted that there could still be other time-varying, individual-specific factors that are not included in this model that can bias the estimates.

The above described models only allow for a monotonic effect of income: a percentage change in income has the same effect irrespective of the absolute level of income. As the bar charts of the distribution of smoking, drinking and sports per income quintile suggest, there may be a non-monotonic effect: the effect of income is not constant across the income distribution, it may change sign or only have an effect on parts of the distribution. To allow for this the three regression models, OLS with and without controls and the FE model, are estimated with income in in quintiles. The first quintile is excluded and serves as the reference category.

The above described regression models estimate the effect of income on (un)healthy consumption for the complete population. It is possible that this effect however differs across individuals, called heterogeneous effects. Particularly, the effect may be dependent upon gender or age. To consider this possibility, two models are estimated. The fourth model estimates heterogeneous effects for gender:

$$y_{it} = a_i + \beta I_{it} + \gamma_1 * Female_i + \gamma_2 * Female * I_{i,t} + \gamma \boldsymbol{\chi}_{it} + \delta_t + \epsilon_{it} \quad (5)$$

In this model, β captures the full effect of log of income for males on the dependent variable. The full effect of income for women consists of the coefficient for log of income, β, and the interaction effect, γ_2. The effect of being female, irrespective of income, on the dependent variable is captured by γ_1. If heterogeneous effects for gender exist, this will be indicated by a significant interaction term, this indicates that effect of income on the dependent variable is significantly different for women than men and thus the interaction coefficient is significantly different from zero.

To consider heterogeneous effects based on age, age categories are used. There are multiple ways to estimate interaction terms for age and income. I choose to estimate an interaction term per age category rather than a linear interaction term. Interaction terms per age category allow for more flexibility and non-monotonicity. The age category 35-44 years old is used as reference category, as this is prime age and thus the most relevant group to compare the others to.

$$y_{it} = a_i + \beta I_{it} + \sum \gamma_k Age\ Category_k + \sum \zeta_k Age\ Category_k * I_{i,t} + \gamma \boldsymbol{\chi}_{it} + \delta_t + \epsilon_{it} \quad (6)$$

In this model, β captures the full effect of log of income for the reference category (35-44 year-olds) on the dependent variable. The gamma coefficients capture the effect of belonging to a specific age category on the dependent variables, compared to belonging to the reference group. The full effect of income for the other age groups, is captured by the sum of β and the interaction term of that age category, ζ. It is only tested whether the effect of income for an age category is significantly different from the effect of income on the dependent variables for the reference category.

4.2 External variation in income: the IACK

In the previous analysis, differences in income are used to estimate the effect of income on the dependent variables. First between-individual and within-individual variation in the OLS models, later only within-individual variation in the FE model. In these models, controls for omitted variables are included to isolate the effect of income, but we cannot be sure if these controls are sufficient. Another approach is to only use exogenous variations in income; this eliminates the possibility of confounding variables. Here, I use policy changes in the IACK to establish exogenous variation in income. The IACK policy setting has been discussed in the background chapter, in this section the method of analysis is discussed, which is inspired by the approach of Gruber and Saez (2002). Furthermore, the first three regression models are repeated with the IACK subsample

4.2.1 Exploiting policy changes in the IACK

Every year the upper and lower bound, minimum and maximum amount and the percentage in the IACK scheme can change. This creates exogenous variation in income that can be used to analyse the effect of income on (un)healthy consumption. The method I use is based on the approach introduced by Gruber and Saez (2002). The policy changes are used to isolate the variation in IACK that is truly exogenous, and this is regressed on the change in (un)healthy consumption.

Consider, that (7) is the 'true' formula, which captures (un)healthy consumption, $y_{i,t}$, for individual i at point t. It consists of an individual fixed effect captured by $\alpha_{i,t}$, formed by among others education and gender. The effect of income at time t for individual i, $I_{i,t}$, is captured by β. This effect is assumed to be time invariant. $\boldsymbol{\chi_{it}}$ is a vector of time-varying individual characteristics that affect the outcome variable, γ captures the effects of these characteristics on (un)healthy consumption. δ_t is a time dummy, that capture year shocks on (un)healthy consumption. Lastly $\epsilon_{i,t}$ represents the error term.

$$y_{it} = \alpha_i + \beta I_{i,t} + \boldsymbol{\chi_{it}}\gamma + \delta_t + \epsilon_{i,t} \quad (8)$$

Furthermore, consider that the income is determined as represented by (9).

$$I_{it} = P_{it} + IACK_{it} \quad (9)$$

Income is the sum of the pre-IACK income, P_{it}, and the $IACK_{it}$. In reality the $IACK_{it}$ is a tax deduction, and formula (5) is thus slightly more complex. However, in essence the IACK is still "received", and as more IACK is better this simplification can be used. The IACK, in turn, is determined by equation (1) as described in the background chapter. Changes in the IACK amount received can come from various sources. Exogenous variation due to changes in the policy: the minimum and maximum amount received, the lower and upper bound and the percentage. However, part of the variation is also endogenous. Individual can respond to or anticipate the policy by changing their income for example

by working more. It is not unlikely that these behavioural responses exist; it is in fact the goal of the policy. To be able to use the IACK as a source of exogenous variation in income it is necessary to isolate the exogenous part of the variation in the IACK.

First, the sample is restricted to IACK recipients. I only include people for the years in which I can identify them as eligible. The reason for this is twofold: an individual can only be affected by the policy if he is eligible. Secondly, changes in IACK eligibility can be endogenous, for example dropping out of the labour force. Eligibility is defined as someone that has a child under twelve, and is a working-single parent or a working-least earner. Income is not used as a criteria for this, even though there is a minimum amount of income that is necessary to receive IACK. However, this varies over the years, and thus I consider this as part of the policy variation.

Next, a 'synthetic' change in IACK is constructed. The actual change in the IACK is as (10), this consists of the exogenous and endogenous change.

$$\Delta IACK_{it} = S_{i,t}\left(P_{it}, B_t^L, B_t^H, s_t^{min}, s_t^{max}, \tau_t\right) - S_{i,t-1}(P_{it}, B_{t-1}^L, B_{t-1}^H, s_{t-1}^{min}, s_{t-1}^{max}, \tau_{t-1}) \quad (10)$$

A synthetic change is constructed by only letting the policy rules change, but keeping the income constant. This simulates the variation in IACK if people were unable to respond to the change in policy, thereby isolating the exogenous variation due to policy change.

$$\Delta\widehat{IACK}_{it} = S_{i,t(-1)}\left(P_{i,t-1}, B_t^L, B_t^H, s_t^{min}, s_t^{max}, \tau_t\right) - S_{i,t-1}(P_{i,t-1}, B_{t-1}^L, B_{t-1}^H, s_{t-1}^{min}, s_{t-1}^{max}, \tau_{t-1}) \quad (11)$$

The simulated change in IACK is then regressed on the change in the dependent variables for (un)healthy consumption. In this model a control for the lagged income level is included, as the absolute income level is expected to influence the consumption. Moreover, year dummies are included. Thus equation (12) is estimated.

$$\Delta y_{i,t} = \beta\Delta\widehat{IACK}_{i,t} + \gamma P_{i,t-1} + \delta_t + \Delta\epsilon_t \quad (12)$$

Chapter 5: Results

5.1 Results from OLS and FE models: smoking, drinking and sports

5.1.1 Smoking behaviour and income

Income is shown to have no significant effect on smoking behaviour, both on the extensive and intensive margin (Table 3). The coefficient of log of income on smoking behaviour on the extensive margin, whether someone smokes, changes sign throughout the specifications. In the first, naïve regression without any controls, the coefficient is negative and insignificant (Column 1, Table 3). There is thus no correlation found between logs of income and engaging in smoking behaviour. After adding age, gender, marital dummies and health dummies, the coefficient becomes insignificantly positive and larger (Column 2, Table 3). Age, gender, being married and the health dummies are significantly negative and thus take the negative effect that was absorbed by log of income in the first model. In the last model, individual and year fixed effects are added and therefore only within subject variation is exploited (Column 3, Table 3). The coefficient of log of income becomes insignificantly negative again. Furthermore, control variables age and gender change sign as well as some marital status dummies. The dummies for health change are not substantially affected. The years 2010 through 2015 have a significantly negative effect on smoking, compared to base year 2008. The three specifications give no indication that a change in income affects the probability of smoking.

The estimated coefficient of log of income on the number of "smokes" is positive and insignificant throughout the three specifications, however the magnitude varies (Column 4-6, Table 3). The coefficient becomes larger after the first set of controls are included (Column 5, Table 3). However, after the individual and year fixed effects are included the coefficient drops below the coefficient of the first two estimations (Column 6, Table 3). But the main point is, that in none of these specifications the estimated effect is significant. Thus, as for the extensive margin, there is no evidence that a change in income affects the intensive margin of smoking.

Overall, no evidence is found that income or changes in income affect smoking behaviour. A possible explanation could be that smoking is addictive, and therefore does not change much over time. This would explain the non-significant effects in the models with individual fixed effect, as these only consider within-subject variation. However, we would still expect to find significant effects in the models without any controls. As from the literature we known that smoking is correlated with income. It is thus surprising that such a correlation is not found in this data.

Table 3 Results for regressions of log of income on smoking, extensive and intensive margin

VARIABLES	(1) Smoking (extensive)	(2) Smoking (extensive)	(3) Smoking (extensive)	(4) Smoking (intensive)	(5) Smoking (intensive)	(6) Smoking (intensive)
Log of Income	-0.000666	0.00107	-0.000281	0.0165	0.0238	0.00403
	(0.00101)	(0.00105)	(0.00125)	(0.0146)	(0.0151)	(0.0178)
Age		-0.00230***	0.00451		-0.0176***	0.0985**
		(0.000257)	(0.00282)		(0.00376)	(0.0401)
Gender		-0.0377***	0.317***		-0.496***	0.880
		(0.00832)	(0.109)		(0.123)	(1.555)
Constant	0.204***	0.331***	-0.132	2.283***	3.458***	-2.015
	(0.00760)	(0.0129)	(0.144)	(0.110)	(0.189)	(2.050)
Observations	29,412	29,412	29,412	29,412	29,412	29,412
R-squared			0.018			0.017
Number of individuals	8,369	8,369	8,369	8,369	8,369	8,369
Marital Dummies	No	Yes	Yes	No	Yes	Yes
Health Dummies	No	Yes	Yes	No	Yes	Yes
Indiviudal FE	No	No	Yes	No	No	Yes
Year FE	No	No	Yes	No	No	Yes

[iii]

[iii] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Smoking (extensive) is a binary variable indicating currently smoking, smoking (intensive) is the number of cigarettes, cigars and pipes smoked per day. Marital dummies indicate the marital status and health dummies indicate the change in health compared to the previous year.

5.1.1 Alcohol consumption and income

Change in income is found to positively affect alcohol consumption on the extensive margin, but such an effect is not found on the intensive margin (Table 4). The estimated coefficient for log of income on alcohol consumption on the extensive margin is significantly positive in the first estimation without any controls (Column 1, Table 4). The coefficient remains significantly positive but drops from 0.0141 to 0.00591 when controls for age, gender, marital dummies and health dummies are added (Column 2, Table 4). This drop in the effect size is likely due to age, it is the only control variable with a positively, significant effect. Moreover, age is positively correlated with log of income. Therefore, the exclusion of this variable let to an upward bias in the coefficient in the first column. The addition of individual and year fixed effects, does not substantially affect the coefficient. It does increase the standard error, but this is not surprising as many dummies are added to the model. Moreover, the estimated coefficient remains significant. This points towards a positive effect of a change in income on the likelihood that an individual consumed alcohol in the past seven days. The estimated coefficient of income for the extensive margin of drinking is 0.00588 in the FE model. Thus the estimated effect of a 10% increase in income is 0.0588. As the dependent variable is a binary variable, this implies that

the estimated average marginal effect of a 10% income increase is 5.88%. Thus a 10% income increase increases the probability of drinking with 5.88 percentage points.

The estimated coefficient of log of income on the number of alcoholic drinks on peak-day is negative and significant in the regression without any controls (Column 4, Table 4). Thus log of income is negatively correlated with the amount of drinks consumed on peak-day. Higher income is associated with fewer drinks on peak-day, but more often associated with alcohol consumption in the last week as was found in the first three columns. However, the coefficient loses its significance when control variables are included (Column 5, Table 4). Age, gender, marital dummies and a deterioration in health all have a significantly negative effect on the number of drinks. This must have led to a downward bias in the coefficient in the fourth column. Particularly, age could have driven this as age is positively correlated with log of income, whereas for gender the correlation with income is negative. After including individual and year fixed effects, the coefficient of log of income becomes positive (Column 6, Table 4). Thus, no evidence is found for an effect of income on the number of drinks consumed on peak-day.

Table 4 Results for regressions of log of income on alcohol consumption, extensive and intensive margin

VARIABLES	(1) Alcohol consumption (extensive)	(2) Alcohol consumption (extensive)	(3) Alcohol consumption (extensive)	(4) Alcohol consumption (intensive)	(5) Alcohol consumption (intensive)	(6) Alcohol consumption (intensive)
Log of Income	0.0141***	0.00591***	0.00588**	-0.0224**	-0.0152	0.00821
	(0.00143)	(0.00151)	(0.00245)	(0.0108)	(0.0111)	(0.0165)
Age		0.00290***	0.000628		-0.0218***	-0.0745**
		(0.000296)	(0.00554)		(0.00225)	(0.0374)
Gender		-0.136***	0.251		-1.973***	-0.0536
		(0.00828)	(0.215)		(0.0643)	(1.449)
Constant	0.620***	0.643***	0.605**	2.880***	5.467***	6.875***
	(0.00994)	(0.0145)	(0.283)	(0.0764)	(0.110)	(1.910)
Observations	29,412	29,412	29,412	29,412	29,412	29,412
R-squared			0.004			0.005
Number of individuals	8,369	8,369	8,369	8,369	8,369	8,369
Marital Dummies	No	Yes	Yes	No	Yes	Yes
Health Dummies	No	Yes	Yes	No	Yes	Yes
Indiviudal FE	No	No	Yes	No	No	Yes
Year FE	No	No	Yes	No	No	Yes

[iv]

[iv] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Alcohol (extensive) is a binary variable indicating alcohol consumption in past seven days, alcohol (intensive) is the number of alcoholic drinks on the day with the most alcohol consumption in the last seven days. Marital dummies indicate the marital status and health dummies indicate the change in health compared to the previous year.

The results in the first three columns of Table 4 show positive estimated coefficients for log of income on the binary variable that indicates alcohol consumption in the last seven days. These results thus suggest that income has a positive effect on the extensive margin of alcohol consumption. Higher incomes are more often associated with alcohol consumption. Moreover, changes in one's income are also positively associated with alcohol consumption. The estimated average marginal effect of a 10% income increase is 5.88%. However, it is difficult to say whether this is a causal effect. I cannot conclude on such an effect on the intensive margin. Initially a significant, negative correlation is found, but this effect disappears as controls are added. Therefore it is suggested that higher income and increases in income make engaging in alcohol consumption more likely, but no effect is found for the amount of drinks consumed on the peak-day. The results thus suggest an interesting difference in the effect on the extensive and intensive margin of alcohol consumption.

5.1.3 Sports and income

The results seem to suggest a negative effect of income on the intensive margin of playing sports, but no consistent effect of income on the extensive margin. The estimated coefficient of income on engaging in sports is significantly negative in the model without controls (Column 1, Table 5). Thus there is a negative correlation between income and engaging in sports. However, this result seems to suffer from a downward bias. In the next column, the coefficient becomes significantly positive after the inclusion of control variables. The coefficient of age is significantly negative, and age is positively correlated with income, therefore it is likely that age causes the upward bias in the estimate of log of income in the first column. After including year and individual fixed effects, the estimated coefficient of log of income becomes insignificant and negative (Column 3, Table 5). By including year and individual fixed effects separately, it is discovered that this is due to the individual fixed effects (Table 10, Appendix). As the coefficient is not affected by only including year fixed effects, but the coefficient does change sign when only individual fixed effects are added. This indicates that there are some personal characteristics that lead to a strong upward bias. The coefficient of age also loses its significance.

The last three columns of Table 5 show a negative association between log of income and the hours of sports played. The estimated coefficient of log of income is negatively, significant in the model without controls (Column 4, Table 5). This coefficient drops from -0.0926 to -0.0759 in the model with controls and remains significant (Column 5, Table 5). The estimated coefficient further drops as year and individual fixed effects are included but remains significant and negative (Colum 6, Table 5). The estimated effect of a 10% increase in income is a -0.446 hours or a 27 minutes decrease in time spent playing sports per week in the last model.

Table 5 Results for regressions of log of income on sports variables, extensive and intensive margin

VARIABLES	(1) Sports (extensive)	(2) Sports (extensive)	(3) Sports (extensive)	(4) Sports (intensive)	(5) Sports (intensive)	(6) Sports (intensive)
Log of Income	0.00568***	0.00360**	-0.00356	-0.0926***	-0.0759***	-0.0446***
	(0.00151)	(0.00159)	(0.00227)	(0.0103)	(0.0109)	(0.0163)
Age		0.00418***	-0.00196		0.00810***	0.000140
		(0.000332)	(0.00512)		(0.00222)	(0.0368)
Female		0.00835	-0.0749		-0.617***	-0.355
		(0.00963)	(0.199)		(0.0633)	(1.428)
Constant	0.591***	0.747***	0.782***	2.829***	3.769***	3.248*
	(0.0107)	(0.0163)	(0.262)	(0.0723)	(0.109)	(1.883)
Observations	29,412	29,412	29,412	29,412	29,412	29,412
R-squared			0.006			0.006
Number of Individuals	8,369	8,369	8,369	8,369	8,369	8,369
Marital Dummies	No	Yes	Yes	No	Yes	Yes
Health Dummies	No	Yes	Yes	No	Yes	Yes
Individual FE	No	No	Yes	No	No	Yes
Year FE	No	No	Yes	No	No	Yes

[v]

[v] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Sports (extensive) is a binary variable indicating engaging in sports activities, sports (intensive) is the number of hours of sports played per week. Marital dummies indicate the marital status and health dummies indicate the change in health compared to the previous year.

The results in the fifth table show a negative effect of income on the amount of time spend playing sports. Higher incomes are associated with lower numbers of hours playing sports. In the most restrictive specification, this effect is found to be almost a half-hour decrease per week for a 10% increase in income. However, interestingly, this effect of income on the hours spend on playing sports is not translated into an effect on the extensive margin. This suggests that an increase in income does lead people to decide to spend fewer hours on playing sports, but it does not lead them to quit playing sports altogether. Thus there is no conclusive evidence that income affects the likelihood of playing sports. Although higher incomes are associated with lower occurrence of playing sports, but this effect seems to be driven by other factors such as age, marital status and health deterioration.

5.2 Non-monotonic effect of income

The means of the dependent variables by income quintiles (Figures 1-3, Appendix) suggest that income may have non-monotonic rather than monotonic effect. Particularly for the extensive and intensive margin of smoking, the intensive margin of drinking and the extensive and intensive margin of playing sports, such an effect is suggested. To allow for this, the above described regression are repeated, with the income in quintiles as independent variables.

Table 11 (Appendix) for smoking illustrates, overall, no evidence for an effect of income on the extensive or intensive margin. On the extensive margin, without any controls the coefficient of the fifth quintile is significantly negative (Column 1). After including basic controls, the coefficient remains significantly negative and the coefficient of the second quintile is significantly positive. However, the coefficients are no longer significant in the FE model, all are insignificantly positive. On the intensive margin, without any controls the coefficient of the second coefficient is significantly positive. This holds in the model with basic controls (Column 5), but not in the FE model (Column 6). Thus, there is no evidence that smoking is significantly different in one of the quintiles from the first quintile. In other words, there is no evidence of an effect, non-monotonic or monotonic, of the income quintiles on smoking behaviour.

For drinking, also no evidence is found that points towards a non-monotonic effect of income on drinking (Table 12, Appendix). On the extensive margin, without any controls the third, fourth and fifth income quintile have a significantly positive estimated coefficient (Column 1). The coefficients for the fourth and fifth quintile remain significant and positive in the model with basic controls, the coefficient of the third coefficient is no longer significant (Column 2). In the FE model, none of the coefficients is significant (Column 3). This thus gives no evidence for the existence of a non-monotonic effect of income on the extensive margin of drinking. In the regression with the log of income, a monotonic effect was found. Therefore, I would have expected that (at least some of) the income quintiles would have been significantly positive, to reflect this monotonic effect. On the intensive margin, without any controls the coefficients of the second and third quintile are significantly negative (Column 4). In the model with basic controls, the coefficients of the second, third and fourth quintile are significantly negative (Column 5). However, in the FE model none of the estimated coefficients are significant (Column 6).

For sports, on the extensive margin, in the first model the coefficients of the second, third and fourth quintile are significantly negative (Table 13, Appendix). In the model with basic controls, the fourth and fifth model are significantly positive (Column 2). However, in the FE model none of the coefficients are significant (Column 3). On the intensive margin, in the first two models all coefficients for income quintiles are significantly negative (Column 4 and 5). In the FE model only the coefficients of the third and fourth income quintile are significantly negative (Column 6). Thus these income quintiles sport significantly less than the people in the first quintile, even after controlling for age, gender, marital status, change in health and individual and year fixed effects. The regressions of the log of income on the intensive margin of sports, suggested a monotonic effect of income on sports. However, these results suggest that this effect is mainly driven by the third and fourth income quintile, rather than the entire distribution.

Even though the distribution of smoking behaviour, alcohol consumption and sports behaviour by income quintile gave rise to some suspicion that there may exist a non-monotonic effect of income on the behaviours, these regression results do not indicate so. For smoking no evidence for any effect is found. Neither is there for alcohol consumption, which is surprising as it somewhat seems to contradict the effect of income on the extensive margin of alcohol found in the log-linear models. For sports, there is no evidence for an effect of income on the extensive margin. There is evidence for a negative effect on the intensive margin, as in the previous estimations with the log of income. However, in this model it seems to be mainly driven by the third and fourth quintile.

5.3 Heterogeneous effects: exploring interaction effects for gender and age

Furthermore, heterogeneous effects based on gender and age are considered in the FE model. The results show heterogeneous effects for men and women for the log of income on the hours of sports played by women (Table 14, Appendix). The interaction effect for female and log of income is significantly positive (Column 7, Table 14). The coefficient for log of income is significantly negative, it gives the full effect of income for men. The sum of the normal and interaction coefficient gives the full effect of income for women, this is still negative but smaller than for men. The estimated coefficients suggest that a 10% increase in income reduces the time spent playing sports by an hour for men, but only by 0.253 hours or 15 minutes for women. Thus this shows that income has a larger effect on the decision of hours to spend playing sports for men than for women, but perhaps men play more sports on average. For the other dependent variables, no significant interaction effects are found. Which is not surprising, as the coefficients of log of income were not significant without the interaction effects. The coefficient of the extensive margin of alcohol consumption was significantly positive in the model without interaction effects. By including the interaction term, the normal coefficient for log of income remains positive but loses its significance and the interaction term is insignificantly negative.

Secondly, the heterogeneous effects based on age were considered (Table 15, Appendix). First, consider the two dependent variables that had significant coefficients in the model without interaction terms: binary variable for alcohol consumption and hours of sports per week. The coefficient of log of income for alcohol consumption is not significant when interaction effects for age and income are included (Column 3, Table 15, Appendix). This captures the full effect of income on alcohol consumption for the age category 35-44 years old. For this group, income thus does not affect their alcohol consumption. The age coefficients and interaction terms for age categories above 45 years are not significant either. Thus their alcohol consumption is not significantly different from 35 to 44 year-olds, nor do they respond differently to an income increase than the aforementioned

group. The estimated coefficients for 16 to 34 years old suggest different behaviour for these groups. The interaction terms for the 16 to 24 year olds and the 25 to 34 year olds are both significantly positive. This suggests that individuals in these age groups are more likely to consume alcohol if their income increases. The estimated age coefficients for these age categories are however significantly negative, thus individuals in this age group are less likely to consume alcohol than the 35 to 44 year olds. Both the age coefficient and the interaction term are larger for the 16-24 years old category, thus the difference between this group and the 35 year olds and older is strongest. The results show a difference in response to income regarding alcohol consumption based on age. Individuals in the age category 16 to 34 years old are less likely to consume alcohol. However, if faced with an income change they are estimated to respond positively in this with regards to alcohol consumption. The average marginal effect of income on the probability of alcohol consumption for the 16 to 24 year olds is the sum of the coefficient for income and the interaction term, in all three models, thus with or without education, this average marginal effect is around 20 percentage points for a 10% increase in income. The older individuals, here 35 years and above, are more likely to consume alcohol, but they do not change their alcohol consumption decision based on income/when their income changes. This is an interesting difference. A possible explanation could be that young people do not really respond to income, but rather to education. As, especially in the younger age group, a higher income compared to others with the same age might be because they are already working and no longer studying. The positive effect of income on alcohol consumption might hide the actual negative effect of education. Moreover, education should be captured in the individual fixed effects for the older individuals as their education does not vary over time anymore but for this young category their education may still change over time.

To test this explanation the same regression was run but with dummies for education categories. Higher secondary education, intermediate and higher vocational education and university all have a significantly positive effect on alcohol consumption. However, the age categories remain significantly negative. Thus 16 to 34 year olds are significantly less likely to consume alcohol than 35 to 44 year olds, even after controlling for education. The interaction term for the youngest group, 16 to 24 year olds, remains significantly positive. Thus this effect cannot be explained by education. The interaction term for 25 to 34 year olds does lose its significance, and was thus caused by education. These effects are found irrespective of if education is defined as the highest obtained degree or if the current degree is also considered. Thus education cannot explain the story for the youngest group. A possible explanation may be that older individuals' alcohol consumption has become a habit that they do not change based on income. However, for young people this behaviour is still responsive to change in

income. This might be because if they do not have income or low income, they simply cannot buy alcohol, whereas for older people it is always within their budget.

The interaction terms for age and hours of sports, the other variable for which a significant effect was found in the previous specifications, are not significant. Moreover, the coefficient for log of income loses its significance. However, the coefficient of age was not significant in the specification without interaction terms. Thus the coefficient of log of income might simply lose its significance because ten irrelevant dummies are added.

Apart from drinks and hours of sports. Smoking produces some significant results. 25 to 34 year olds are more likely to engage in smoking, but income does not affect their smoking behaviour nor is others smoking behaviour affected by income. 16 to 24 year olds smoke fewer cigarettes, pipes and cigars, however this number increases if income increases. No such effects are found for other age groups.

5.4 External variation in a subsample: exploiting policy changes in the IACK

In the following part I exploit the policy changes in IACK as external variation in income. This is done on a subsample of the data: IACK recipients. To be able compare the results of this analysis to the OLS and FE models, I repeat the models on the IACK subsample.

For smoking none of the estimated coefficients of income on the extensive or intensive dimension of smoking is significant (Table 16, Appendix). This is in line with the results for the full sample as discussed earlier. Similarly for drinking, none of the estimated coefficients of income on the extensive or intensive dimension of alcohol consumption are significant (Table 17, Appendix). In the FE specification, the signs are positive as in the models on the full sample. However, for the full sample the estimated coefficient on the extensive margin is significant, unlike for the IACK sample. Lastly, for sports the coefficients for the extensive margin are insignificantly positive in the OLS model with and without controls, but significantly negative in the FE model (Table 18, Appendix). On the intensive margin, the coefficients are negative in all three specifications, but only significant in the FE model. In the full sample, the coefficient in the FE model for the extensive margin is insignificant, but does have the same sign as the IACK sample estimate. The estimate for the intensive margin is significantly negative in the full sample and the IACK sample, however there is a substantial difference in the estimated coefficient's size. For the full sample the coefficient is -0.0446, for the IACK it is -0.505. For a 10% income increase, the estimate of the full sample estimates a 0.45 hour, or 27 minutes, decrease in time spend per week playing sports; the estimate of the IACK sample estimates a 5 hour decrease in time spend per week playing sports, an unlikely high amount.

By using policy changes in IACK as a source of exogenous variation in income, none of the estimated coefficients for synthetic change in IACK are significant (Table 6). The estimated coefficients on smoking behaviour, both the extensive and intensive margin, are insignificant (Column 1 and 2, Table 6). This is in line with the results found on the full sample (Table 3) and on the IACK-subsample (Table 16, Appendix), and thus not surprising. The estimated coefficient for the number of drinks is insignificant (Column 4, Table 6), as in the earlier regressions on the full sample and IACK-subsample. The estimated coefficient for alcohol consumption on the extensive margin is not significant (Column 3, Table 6). The effect for this coefficient was found to be significant in the complete sample (Table 4). This could indicate that the effect found earlier was not driven by income but by confounding factors. Another possible explanation is that the effect of income is not present in the IACK population. This is likely as the regression with interaction effects for age and income, showed that the effect of income on the extensive margin of alcohol consumption was driven by the age group 16 to 24 year olds. The IACK is aimed at parents of young children, the minimum age in the subsample is 24 and thus the age group that drives the effect is not present in the subsample. This explanation is further supported by that the effect is also not found in the FE model on the IACK-subsample. Therefore, I argue that the coefficient of income is insignificant because young people are not affected by the IACK. However, I also cannot conclude on the causality of the effect found in the FE model, as the IACK is not a suitable source of policy change to simulate exogenous changes in income in this group.

The estimated coefficients for playing sports and for the hours of sports played are not significant in the IACK specification either (Column 4 and 5, Table 6). In the complete sample I find a significant effect of income on the hours of sports played. It is later determined that this effect is significant for both men and women, but much stronger for men (Table 5). As only single, working parents or working parents who earn less than their partners receive IACK, IACK recipients are largely women (almost 85%). The effect may thus be lost because the sample consists largely of women, for whom the effect is small. However, the coefficients of income on the extensive and intensive margin of sports were significant in the FE models on the IACK-subsample. This makes the explanation that the effect disappears because of the absence of men in the subsample less likely, and makes it more likely that the effect is no longer found because it is not caused by exogenous variation in income but rather by confounding factors or endogenous variation in income.

Table 6 Results of regressions of synthetic change in IACK on the change in the dependent variables

VARIABLES	(1) Smoking (extensive)	(2) Smoking (intensive)	(3) Drinking (extensive)	(4) Drinking (intensive)	(5) Sports (extensive)	(6) Sports (intensive)
Synthetic change in IACK	-6.11e-05	-0.00228	0.000339	0.000949	0.000141	0.00194
	(0.000294)	(0.00254)	(0.000636)	(0.00335)	(0.000453)	(0.00188)
Log of net income	0.00736	0.208***	-0.00704	0.00337	-0.0110	-0.0378
	(0.00838)	(0.0805)	(0.0177)	(0.0934)	(0.0138)	(0.0572)
Observations	760	727	604	604	772	772
Number of individuals	343	338	287	287	341	341
Marital Dummies	NO	NO	NO	NO	NO	NO
Health Dummies	NO	NO	NO	NO	NO	NO
Individual FE	NO	NO	NO	NO	NO	NO
Year FE	YES	YES	YES	YES	YES	YES

[vi]

[vi] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Dependent variables are the difference in variable y in t and t-1: $y_t - y_{t-1}$. Smoking (extensive) is a binary variable indicating currently smoking, smoking (intensive) is the number of cigarettes, cigars and pipes smoked per day. Alcohol (extensive) is a binary variable indicating alcohol consumption in past seven days, alcohol (intensive) is the number of alcoholic drinks on the day with the most alcohol consumption in the last seven days. Sports (extensive) is a binary variable indicating engaging in sports activities, sports (intensive) is the number of hours of sports played per week. The synthetic change in IACK is as follows: IACK amount with policy of t and income of $t-1$ minus the IACK amount with policy of $t-1$ and income of $t-1$.

Exploiting external variation due to changes in IACK policy thus suggest that there are no significant effects of income changes on healthy and unhealthy consumption. However, the IACK recipients are a specific sub-sample of the population, largely women and within 24-61 years old, on average almost 40. Thus it is not a relevant policy to consider effects in specific other groups of the population.

Chapter 6: Discussion

No evidence for an effect of income on smoking is found in any of the models employed. This could be because smoking is a habit or addiction and therefore not much affected by income. Moreover, if people start smoking at an early age, it may not be influenced by income yet. Surprising is that there is no significant correlation between smoking and income.

I do find evidence for a positive effect of income on the extensive margin of alcohol consumption. The coefficient of log income is significant in the FE model, but the coefficients of the income quintiles are not. However, this may be because relevant variation within income quintiles is lost in the latter model. The average marginal effect of a 10% income increase is a 5.88 percentage point increase in the likelihood of drinking. Furthermore, no evidence is found for an effect of income on the intensive margin. This suggests an interesting difference in the effect of income on alcohol consumption on the extensive and intensive margin. Positive income changes are associated with an increase in the likelihood of consuming alcohol, but such changes are not a factor that influences the number of alcohol units consumed. Moreover, it turns out that the suggested effect of income on the extensive margin is driven by 16 to 24 year olds. They are less likely to engage in alcohol consumption than 35+, but income increases this probability. The total average marginal effect of a 10% increase on income is a 20 percentage point increase in the likelihood of engaging in alcohol consumption for 16 to 24 year olds. It is shown that education cannot explain this effect. Another explanation for the difference in the effect of income on drinking behaviour for different age groups might be that for the older group alcohol is always within budget, whereas the younger groups income is needed to be able to purchase it and therefore they do respond to an income increase. No evidence for an effect is found in the IACK sample, by applying the OLS and FE models or by exploiting external variation in IACK. However, this could be because of the sample. Thus, evidence is found for an effect of income on the likelihood of consuming alcohol. However, it is difficult to conclude on causality as I am unable to successfully exploit policy changes as a source of exogenous variation in income.

For sports, the coefficient of the log of income on the intensive margin is significant in the FE model, but only the income quintiles of the third and fourth quintile are significant in the FE model with income quintiles. This may be because the effect is driven by these two groups, but the model discards within income quintile variation, thus I cannot be sure. Moreover, the coefficient of the extensive margin is insignificant. This suggest that in response to an income increase, people reduce the hours of sports played per week, but do not decide to quit altogether. Furthermore, the effect is found to be significantly stronger for men than for women. The effect is not found when the external variation in IACK is exploited. This may be because the sample mainly consists of women. However, this is unlikely as the coefficient of log of income is significant in this subsample. Thus it is more probable

that the effect is not causal. Concluding, some evidence is found that income may have an effect on the intensive margin of sports but the analysis on the IACK subsample suggest that this is not a causal effect.

The hypotheses of this thesis were that income has a positive effect on healthy and moderately unhealthy consumption, but a negative effect on strongly or severely unhealthy consumption. This thesis finds some evidence for a positive effect of income on the extensive dimension of alcohol consumption and negative on the intensive dimension of sports, but finds no such evidence for the other types of consumption: the extensive and intensive margin of smoking, the intensive margin of drinking and the extensive margin of sports. Thus some evidence is found for the first hypothesis, no evidence is found in support of the second hypothesis.

There is some, limited support for the first hypothesis that wealthier individuals engage more in healthy and moderately unhealthy consumption. In the FE model, I find a positive effect of income on the extensive margin of alcohol consumption which can be seen as a form of moderately unhealthy consumption. However, such an effect is not found for the types of healthy consumption: the intensive and extensive margin of sports. For the intensive margin of sports, some evidence for a negative effect of income is found. However, the IACK analysis suggest that this effect is not causal. Still, the association between positive income changes and negative changes in hours of sports played are not in line with the hypothesis. As playing sports is healthy, I would have expected to have found a positive association or effect. However, the reason might be that the cost of playing sports is mainly opportunity costs of time, especially when considering the intensive margin. Therefore, if income increases people may be able to afford more healthy consumption, but spending hours on sports becomes more expensive. The intensive margin of sports thus may not be representative for other types of healthy consumption that do not have these high opportunity costs in terms of time. I thus cannot clearly reject or accept the first hypothesis. Some evidence in support for it is found, but for some other indicators the effect is not found. This points towards that different forms of healthy and moderately unhealthy consumption cannot be generalized.

No evidence is found in support for the second hypothesis that income has a negative effect on severely unhealthy consumption. For both dimensions of smoking and the intensive margin of alcohol consumption no effect of income is found. Thus, I can reject the second hypothesis, at least for these specific types of unhealthy behaviour.

The research question of this paper was to see to what extent income affects healthy and unhealthy consumption. This paper points towards that it is hard to generalize the effect of income on different

types of healthy and unhealthy consumption. Strong support for an effect of income on the extensive margin of smoking is found, but do not find evidence for an effect on other types of healthy and unhealthy consumption. Therefore it is not possible to make a general statement on the effect of income on healthy and unhealthy behaviour. This might not be surprising as the correlation between different types of healthy and unhealthy behaviour is low (see Table 2, Data chapter, or e.g. Cutler and Glaeser, 2005). Thus it is unlikely that these behaviours are caused by the same factor, moreover it is unlikely that income affects them in the same way. Another point this thesis makes is that the effect is different between groups. The effect of income on the extensive dimension of alcohol consumption seems to be largely driven by young people, 16-24 year olds.

This thesis also aims to serve as a first empirical test of the model for the SES health gradient proposed by Galama and van Kippersluis (2017) by testing one of their predictions. As mentioned before, this thesis finds some support for their prediction, in the form of a positive effect of income on the prevalence of drinking, but does not find such an effect for other types of healthy and unhealthy consumption. It is thus difficult to interpret whether this is confirms their model or not. Moreover, this thesis is only able to consider a few forms of unhealthy and healthy consumption, these effects may not be generalized to other types of behaviour. Furthermore, Galama and van Kippersluis (2017) hypothesis is based on wealth rather than income. Although wealth and income are closely linked, this could be a potential explanation for why the prediction does not hold. Concluding, this thesis finds limited support for the prediction of the Galama and van Kippersluis (2017) framework. However, more empirical results are needed to make statements on the empirical validity of this framework.

The positive effect of income on the extensive margin of alcohol consumption is in line with previous findings such as Ettner (1996), Adda, Banks and van Gaudecker (2009), van Kippersluis and Galama (2014) and Apouey and Clark (2015). Van Kippersluis and Galama (2014) also find a similar difference in the effect on the prevalence of drinking and the number of drinks, a positive effect of inheritances and lottery winnings on the prevalence, the extensive margin, but no such effect on the number of drinks, the intensive margin. This thesis does not find effects of income on smoking, nor do Kim and Ruhm (2012) by using inheritances or van Kippersluis and Galama (2014) for lottery winnings. The latter authors however, do report a positive effect of inheritances on the prevalence of smoking, extensive margin, and the number of cigarettes, intensive margin. Apouey and Clark (2015) also find support for such an effect by using lottery winnings. This makes it even more surprising that the results in this thesis do not even indicate a correlation between income and smoking. However, this might be due to the different type of income or wealth that is considered in the different papers. The effect of

an income increase on smoking may not be comparable to a lottery winning or inheritance, if people treat these sources of income differently ("mental accounting", see Thaler, 1985).

This paper has some limitations. First, there is a lack of external variation which makes it difficult to conclude on causality. The FE models are a good attempt to control for many factors that could bias the results, however there may still be omitted time-varying personal effects that are not captured by the model. In an effort to isolate the effect of external variation in income, the effect of external changes in the IACK was exploited. However, due to the specific target group of this policy, it was later established that this policy is not suitable for this specific analysis. It is however is a promising source of external variation in income that could be used particularly if the group of interest is women with young children. Secondly, due to data and time constraints I was only able to analyse the effects of income on smoking, alcohol and sports. These are important dimensions of healthy and unhealthy behaviour, but these effects cannot necessarily be generalized to other types of unhealthy and healthy consumption such as food and nutrition.

Based on this thesis, there are some suggestions for further work. First, this thesis was unable to use external variation in income by exploiting the IACK. However, the type of analysis that was aimed for would contribute to the existing empirical literature. Therefore a suggestion is to consider other policies similar to the IACK that could be used to exploit external variation in income in a similar setup. Secondly, it would be worthwhile to extend the scope of healthy and unhealthy behaviour. This paper focuses on smoking, drinking and sports, and many papers have looked at smoking and drinking. However, it would also be interesting to see what the effect is on for example food consumption, as this is a different dimension of behaviour. Lack of detailed data might prove difficulties here. Thirdly, this paper uses Dutch data, it is always interesting to see if these results hold in different country-settings. Finally, this paper looks at one prediction of Galama and van Kippersluis' model in detail, but to fully test and ssess their model other predictions should be tested empirically as well.

Chapter 7: Conclusion

This paper examined the relation between income and smoking, drinking and sports. Thereby the paper aims to test a prediction from Galama and van Kippersluis' model that tries to explain the SES-health gradient.

Evidence for an effect of income on the extensive margin of alcohol consumption was found in the form of a significant coefficient of income for the extensive margin of alcohol consumption in a FE model with individual and year fixed effects and controlling for age, gender, marital status and health change. It seems that this effect is mainly driven by young people, which cannot be explained by the different link between education and income at that age. It is difficult to assess whether this effect is causal, as these young people are not IACK recipients and are therefore not affected by the policy. The estimated coefficient for the intensive margin of sports is significant in the FE model, however the analysis with policy changes in IACK casts serious doubts over the causality of this effect. The coefficients for the forms of unhealthy and healthy behaviour are not significant, or are not significant in the FE model.

Overall, the paper thus finds limited support for the hypotheses of Galama and van Kippersluis. The suggested positive effect of income on the extensive margin of alcohol does point toward a positive effect of income on moderately unhealthy consumption. However, for other types of unhealthy and healthy consumption I do not find an effect of income. This thus does not support the hypothesis. This makes it difficult to accept or reject the hypotheses and to answer the research question. The findings suggest that it should be emphasized that different types of unhealthy and healthy consumption cannot be generalized.

The contributions of this paper are as follows. Firstly, it adds to the small body of literature on the effect of income on unhealthy and healthy behaviour. This thesis uses policy changes in the IACK to exploit external variation in income, this is a new approach in this field. It confirms the findings of other papers on a positive effect of income on the extensive margin of alcohol consumption. It does not find an effect for smoking, this is in line with some of the literature but contradicts some papers that did find an effect. Secondly, this thesis is a first empirical test of the Galama and van Kippersluis (2017) framework for the SES health gradient. It provides limited support for the framework, but more research is needed to conclude on the empirical validity of the framework.

Based on this paper I can only make limited policy suggestions. It is of interest to policymakers to know what causes different type of healthy and unhealthy consumption in order to be able to promote or discourage these types of behaviour. This paper only finds such an effect of income for the extensive margin of alcohol consumption, and not for the other types of unhealthy and healthy consumption that were studied. Therefore, policymakers could take this effect on alcohol consumption, and related health effects, into account when designing policy that affects income. However, to make concrete policy suggestions, more knowledge on what causes these different types of healthy and unhealthy behaviour is needed.

Reference list

Adda, J., Banks, J., & von Gaudecker, HM (2008). *The impact of income shock on health: evidence from cohort data* (No. w3329). IZA Discussion Papers.

Apouey, B., & Clark, A. E. (2015). Winning big but feeling no better? The effect of lottery prizes on physical and mental health. *Health economics*, *24*(5), 516-538.

Averett, S., & Wang, Y. (2013). The effects of Earned Income Tax Credit payment expansion on maternal smoking. *Health economics*, *22*(11), 1344-1359.

Becker, G. S., & Mulligan, C. B. (1997). The endogenous determination of time preference. *The Quarterly Journal of Economics*, *112*(3), 729-758.

Case, A., & Deaton, A. S. (2005). Broken down by work and sex: How our health declines. In *Analyses in the Economics of Aging* (pp. 185-212). Chicago: University of Chicago Press.

Case, A., Fertig, A., & Paxson, C. (2005). The lasting impact of childhood health and circumstance. *Journal of health economics*, *24*(2), 365-389.

Chaloupka, F. J., & Warner, K. E. (2000). The economics of smoking. *Handbook of health economics*, *1*, 1539-1627.

Cook, P. J., & Moore, M. J. (1999). *Alcohol* (No w6905). National Bureau of Economic Research.

Cutler, D., & Glaeser, E. (2005). *What explains differences in smoking, drinking and other health-related behaviors* (No. w11100). National Bureau of Economic Research.

Cutler, D. M., Lleras-Muney, A., & Vogl, T. (2008). *Socioeconomic status and health: dimensions and mechanisms* (No. w14333). New York: National Bureau of Economic Research.

Currie, J., & Madrian, B. C. (1999). Health, health insurance and the labor market. *Handbook of labor economics*, *3*, 3309-3416.

Ehrlich, I., & Chuma, H. (1990). A Model of the Demand for Longevity and the Value of Life Extension. *Journal of Political economy*, *98*(4), 761-782.

Ettner, S. L. (1996). New evidence on the relationship between income and health. *Journal of health economics*, *15*(1), 67-85.

Galama, Titus J., A Contribution to Health-Capital Theory (April 23, 2015). CESR-Schaeffer Working Paper No. 2015-004. Available at SSRN: https://ssrn.com/abstract=2601121 or http://dx.doi.org/10.2139/ssrn.2601121.

Galama, T. J., & van Kippersluis, H. (2017). *A theory of socioeconomic disparities in health over the life cycle.* Manuscript in preparation.

Grossman, M. (1972a). On the concept of health capital and the demand for health. *Journal of Political economy*, *80*(2), 223-255.

Grossman, M. (1972b). *The demand for health – A theoretical and empirical investigation.* New York: National Bureau of Economic Reasearch.

Grossman, M. (2000). The human capital model. In *Handbook of health economics*, *1*, 347-408.

Gruber, J., & Saez, E. (2002). The elasticity of taxable income: evidence and implications. *Journal of public Economics*, *84*(1), 1-32.

Fuchs, V. R. (1982). Time preference and health: an explanatory study. In V. R. Fuchs (ed.), *Economic aspects of health* (pp. 93-120). Chicago: University of Chicago Press.

Kawachi, I., Adler, N. E., & Dow, W. H. (2010). Money, schooling, and health: Mechanisms and causal evidence. *Annals of the New York Academy of Sciences*, *1186*(1), 56-68.

Kenkel, D. S., Schmeiser, M. D., & Urban, C. (2014). Is smoking inferior? Evidence from variation in the earned income tax credit. *Journal of Human Resources*, *49*(4), 1094-1120.

Kim, B., & Ruhm, C. J. (2012). Inheritances, health and death. *Health Economics*, *21*(2), 127-144.

Lechner, M. (2009). Long-run labour market and health effects of individual sports activities. *Journal of health economics*, *28*(4), 839-854.

Ried, W. (1998). Comparative dynamic analysis of the full Grossman model. *Journal of Health Economics*, *17*(4), 383-425.

Ruhm, C. J. (2000). Are recessions good for your health?. *The Quarterly journal of economics*, *115*(2), 617-650.

Ruhm, C. J. (2005). Healthy living in hard times. *Journal of health economics*, *24*(2), 341-363.

Smith, J. P. (1999). Healthy bodies and thick wallets: the dual relation between health and economic status. *The journal of economic perspectives: a journal of the American Economic Association*, *13*(2), 144.

Thaler, R. (1985). Mental accounting and consumer choice. *Marketing science*, *4*(3), 199-214.

Van Kippersluis, H., & Galama, T. J. (2014). Wealth and health behavior: Testing the concept of a health cost. *European economic review*, *72*, 197-220.

Van Ours, J. C. (2004). A pint a day raises a man's pay; but smoking blows that gain away. *Journal of health economics*, *23*(5), 863-886.

Wagstaff, A. (1986). The demand for health: some new empirical evidence. *Journal of health economics*, *5*(3), 195-233.

Appendix

Table 7 Overview of the recording time of the various surveys, categorized by wave in the panel data set

Dataset	Variables	First measurement	Second measurement
	Wave 1: 2008		
Health	Smoking, alcohol (dependent); change in health (control)	2008-11	2008-12
Social Integration and Leisure	Sports (dependent)	2009-02	2009-03
Family and Household	IACK identification	2008-03	2008-06
Background	Age, gender, marital status (control)	2008-06	-
	Wave 2: 2009		
Health	Smoking, alcohol (dependent); change in health (control)	2009-11	2009-12
Social Integration and Leisure	Sports (dependent)	2010-02	2010-03
Family and Household	IACK identification	2009-03	2009-04
Background	Age, gender, marital status (control)	2009-06	-
	Wave 3: 2010		
Health	Smoking, alcohol (dependent); change in health (control)	2010-11	2010-12
Social Integration and Leisure	Sports (dependent)	2011-02	2011-03
Family and Household	IACK identification	2010-03	-
Background	Age, gender, marital status (control)	2010-06	-
	Wave 4: 2011		
Health	Smoking, alcohol (dependent); change in health (control)	2011-11	2011-12
Social Integration and Leisure	Sports (dependent)	2012-02	2012-03
Family and Household	IACK identification	2011-03	2011-04
Background	Age, gender, marital status (control)	2011-06	-
	Wave 5: 2012		
Health	Smoking, alcohol (dependent); change in health (control)	2012-11	2012-12
Social Integration and Leisure	Sports (dependent)	2013-03	2013-03
Family and Household	IACK identification	2012-03	2012-04
Background	Age, gender, marital status (control)	2012-06	-
	Wave 6: 2013		
Health	Smoking, alcohol (dependent); change in health (control)	2013-11	2013-12
Social Integration and Leisure	Sports (dependent)	2014-02	2014-03
Family and Household	IACK identification	2013-03	2013-04
Background	Age, gender, marital status (control)	2013-06	-
	Wave 7: 2015		
Health	Smoking, alcohol (dependent); change in health (control)	2015-07	2015-08
Social Integration and Leisure	Sports (dependent)	2015-10	2015-11
Family and Household	IACK identification	2015-09	2015-10
Background	Age, gender, marital status (control)	2015-06	-

Table 8 Results for OLS regressions of log of income on smoking, drinking and sports, with controls but without the Health Change Dummies

VARIABLES	(1) Smoking (extensive)	(2) Smoking (intensive)	(3) Drinking (extensive)	(4) Drinking (intensive)	(5) Sports (extensive)	(6) Sports (intensive)
Log of income	0.00118	0.0255*	0.00603***	-0.0148	0.00355**	-0.0759***
	(0.00105)	(0.0151)	(0.00152)	(0.0111)	(0.00159)	(0.0109)
Age	0.00227***	-0.0173***	0.00276***	-0.0221***	0.00438***	0.00961***
	(0.000257)	(0.00375)	(0.000295)	(0.00225)	(0.000333)	(0.00222)
Female	-0.0381***	-0.501***	-0.136***	-1.975***	0.00826	-0.617***
	(0.00832)	(0.123)	(0.00828)	(0.0643)	(0.00966)	(0.0636)
Constant	0.324***	3.373***	0.640***	5.445***	0.758***	3.846***
	(0.0128)	(0.188)	(0.0144)	(0.109)	(0.0162)	(0.108)
Observations	29,412	29,412	29,412	29,412	29,412	29,412
Number of individuals	8,369	8,369	8,369	8,369	8,369	8,369
Marital Dummies	YES	YES	YES	YES	YES	YES
Health Change Dummies	NO	NO	NO	NO	NO	NO
Individual FE	NO	NO	NO	NO	NO	NO
Year FE	NO	NO	NO	NO	NO	NO

[vii]

[vii] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Smoking (extensive) is a binary variable indicating currently smoking, smoking (intensive) is the number of cigarettes, cigars and pipes smoked per day. Alcohol (extensive) is a binary variable indicating alcohol consumption in past seven days, alcohol (intensive) is the number of alcoholic drinks on the day with the most alcohol consumption in the last seven days. Sports (extensive) is a binary variable indicating engaging in sports activities, sports (intensive) is the number of hours of sports played per week. Marital dummies indicate the marital status.

Table 9 Results for FE regressions of log of income on smoking, drinking and sports, with controls but without the Health Change Dummies

VARIABLES	(1) Smoking (extensive)	(2) Smoking (intensive)	(3) Drinking (extensive)	(4) Drinking (intensive)	(5) Sports (extensive)	(6) Sports (intensive)
Log of income	-0.000133	0.00633	0.00601**	0.00877	-0.00372	-0.0457***
	(0.00125)	(0.0178)	(0.00246)	(0.0166)	(0.00227)	(0.0163)
Age	0.00467*	0.100**	0.000927	-0.0731*	-0.00201	-7.27e-07
	(0.00282)	(0.0402)	(0.00555)	(0.0374)	(0.00513)	(0.0369)
Female	0.316***	0.876	0.247	-0.0716	-0.0769	-0.371
	(0.109)	(1.558)	(0.215)	(1.450)	(0.199)	(1.430)
Constant	-0.145	-2.164	0.585**	6.780***	0.789***	3.282*
	(0.144)	(2.053)	(0.284)	(1.911)	(0.262)	(1.884)
Observations	29,412	29,412	29,412	29,412	29,412	29,412
R-squared	0.015	0.013	0.002	0.004	0.004	0.004
Number of individuals	8,369	8,369	8,369	8,369	8,369	8,369
Marital Dummies	YES	YES	YES	YES	YES	YES
Health Change Dummies	NO	NO	NO	NO	NO	NO
Individual FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

viii

[viii] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Smoking (extensive) is a binary variable indicating currently smoking, smoking (intensive) is the number of cigarettes, cigars and pipes smoked per day. Alcohol (extensive) is a binary variable indicating alcohol consumption in past seven days, alcohol (intensive) is the number of alcoholic drinks on the day with the most alcohol consumption in the last seven days. Sports (extensive) is a binary variable indicating engaging in sports activities, sports (intensive) is the number of hours of sports played per week. Marital dummies indicate the marital status.

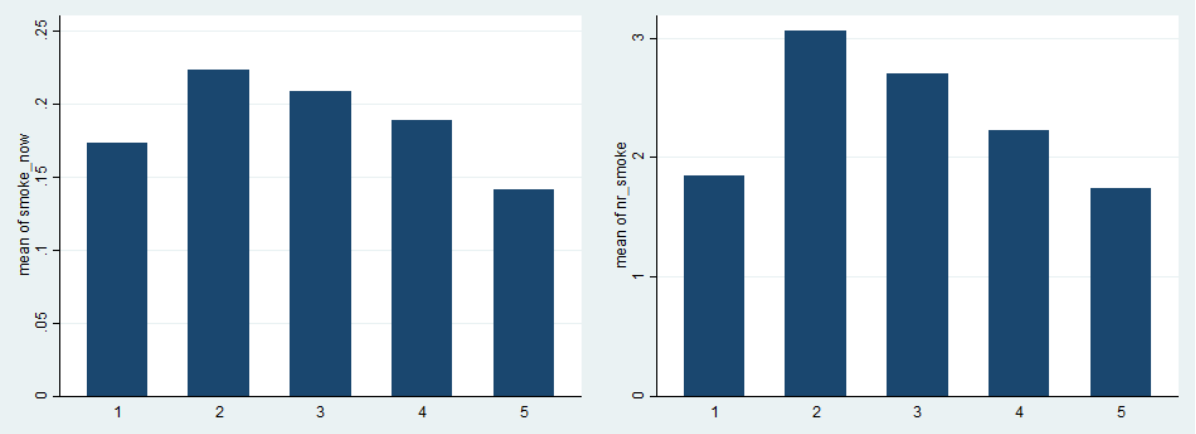

Figure 1 Mean of extensive margin of smoking (L) and intensive margin of smoking (R) by income quintile

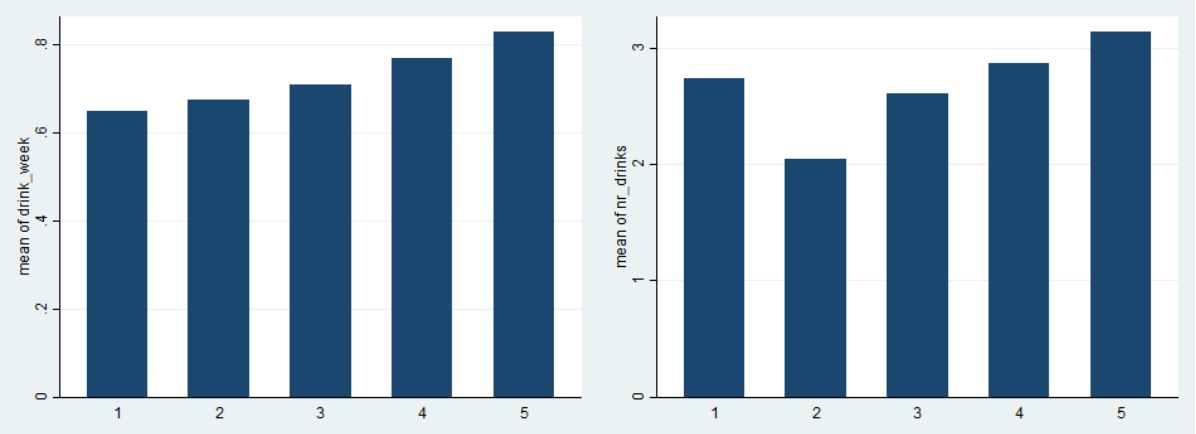

Figure 2 Mean of extensive margin of drinking (L) and intensive margin of drinking (R) by income quintile

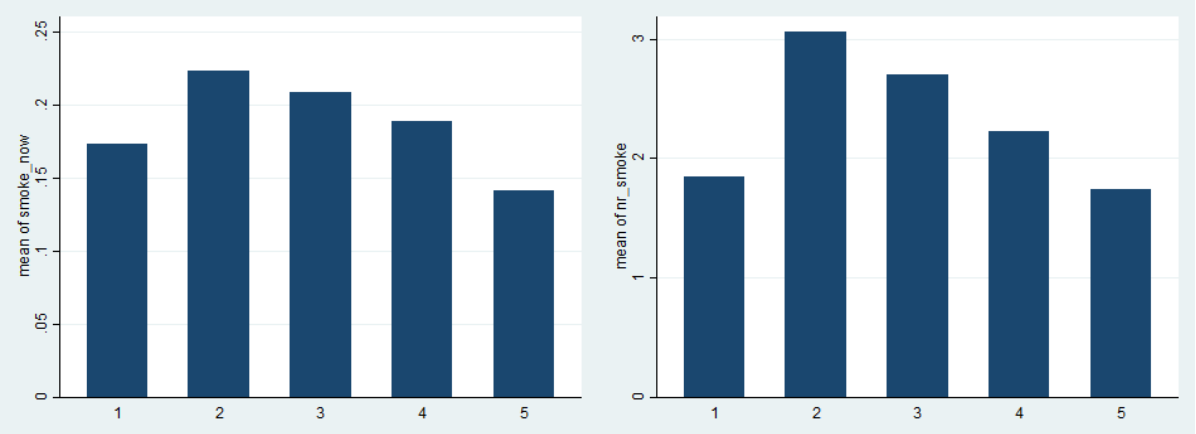

Figure 3 Mean of extensive margin of sports (L) and intensive margin of sports (R) by income quintile

Table 10 Regression results of log of income on the extensive margin of sports; without FE, with individual and year FE added separately and both FE

VARIABLES	(1) Sports (extensive)	(2) Sports (extensive)	(3) Sports (extensive)	(4) Sports (extensive)	(5) Sports (extensive)
Log of income	-0.00568***	0.00360**	0.00360**	-0.00364	-0.00356
	(0.00151)	(0.00159)	(0.00159)	(0.00227)	(0.00227)
Age		-0.00418***	-0.00402***	-0.00581***	-0.00196
		(0.000332)	(0.000345)	(0.000967)	(0.00512)
Gender		0.00835	0.00865	-0.0690	-0.0749
		(0.00963)	(0.00963)	(0.199)	(0.199)
Constant	0.591***	0.747***	0.747***	0.958***	0.782***
	(0.0107)	(0.0163)	(0.0166)	(0.114)	(0.262)
Observations	29,412	29,412	29,412	29,412	29,412
R-squared				0.006	0.006
Number of individuals	8,369	8,369	8,369	8,369	8,369
Marital Dummies	NO	YES	YES	YES	YES
Health Dummies	NO	YES	YES	YES	YES
Indiviudal FE	NO	NO	NO	YES	YES
Year FE	NO	NO	YES	NO	YES

[ix]

[ix] [ix] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Sports (extensive) is a binary variable indicating engaging in sports activities, sports (intensive) is the number of hours of sports played per week. Individual and year fixed effects are separately added.

Table 11 Results of regressions of income quintiles on smoking, extensive and intensive margin

VARIABLES	(1) Smoking (extensive)	(2) Smoking (extensive)	(3) Smoking (extensive)	(4) Smoking (intensive)	(5) Smoking (intensive)	(6) Smoking (intensive)
Second income quintile	0.00506	0.0159***	0.00309	0.182**	0.253***	0.0409
	(0.00600)	(0.00610)	(0.00673)	(0.0860)	(0.0875)	(0.0958)
Third income quintile	0.00422	0.0109	0.00456	0.162*	0.169*	0.0862
	(0.00667)	(0.00685)	(0.00796)	(0.0959)	(0.0985)	(0.113)
Fourth income quintile	-0.00577	0.000137	0.00762	0.0866	0.0713	0.226*
	(0.00705)	(0.00736)	(0.00882)	(0.102)	(0.106)	(0.125)
Fifth income quintile	-0.0297***	-0.0226***	0.000825	-0.195*	-0.216*	0.172
	(0.00793)	(0.00839)	(0.0102)	(0.114)	(0.121)	(0.146)
Age		-0.00224***	0.00449		-0.0169***	0.0984**
		(0.000257)	(0.00282)		(0.00375)	(0.0401)
Gender		-0.0468***	0.317***		-0.605***	0.878
		(0.00853)	(0.109)		(0.126)	(1.555)
Constant	0.204***	0.338***	-0.136	2.338***	3.560***	-2.070
	(0.00607)	(0.0123)	(0.144)	(0.0885)	(0.181)	(2.048)
Observations	29,412	29,412	29,412	29,412	29,412	29,412
R-squared			0.018			0.017
Number of individuals	8,369	8,369	8,369	8,369	8,369	8,369
Marital Dummies	NO	YES	YES	NO	YES	YES
Health Dummies	NO	YES	YES	NO	YES	YES
Indiviudal FE	NO	NO	YES	NO	NO	YES
Year FE	NO	NO	YES	NO	NO	YES

[x]

[x] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Smoking (extensive) is a binary variable indicating currently smoking, smoking (intensive) is the number of cigarettes, cigars and pipes smoked per day. Marital dummies indicate the marital status and health dummies indicate the change in health compared to the previous year.

Table 12 Results of regressions of income quintiles on alcohol consumption, extensive and intensive margin

VARIABLES	(1) Alcohol consumption (extensive)	(2) Alcohol consumption (extensive)	(3) Alcohol consumption (extensive)	(4) Alcohol consumption (intensive)	(5) Alcohol consumption (intensive)	(6) Alcohol consumption (intensive)
Second income quintile	0.0169*	-0.00489	0.00953	-0.476***	-0.205***	-0.157*
	(0.00945)	(0.00973)	(0.0132)	(0.0692)	(0.0695)	(0.0893)
Third income quintile	0.0489***	0.00875	0.0137	-0.223***	-0.231***	-0.122
	(0.00990)	(0.0103)	(0.0157)	(0.0738)	(0.0744)	(0.106)
Fourth income quintile	0.0945***	0.0343***	0.0112	-0.0621	-0.257***	-0.177
	(0.0100)	(0.0107)	(0.0173)	(0.0757)	(0.0779)	(0.117)
Fifth income quintile	0.142***	0.0649***	0.0223	0.126	-0.163*	-0.140
	(0.0109)	(0.0119)	(0.0201)	(0.0832)	(0.0873)	(0.136)
Age		0.00297***	0.000658		-0.0209***	-0.0744**
		(0.000296)	(0.00554)		(0.00226)	(0.0374)
Gender		-0.121***	0.251		-1.991***	-0.0499
		(0.00874)	(0.215)		(0.0677)	(1.449)
Constant	0.651***	0.651***	0.630**	2.863***	5.500***	7.026***
	(0.00753)	(0.0136)	(0.283)	(0.0585)	(0.104)	(1.909)
Observations	29,412	29,412	29,412	29,412	29,412	29,412
R-squared			0.004			0.005
Number of individuals	8,369	8,369	8,369	8,369	8,369	8,369
Marital Dummies	NO	YES	YES	NO	YES	YES
Health Dummies	NO	YES	YES	NO	YES	YES
Indiviudal FE	NO	NO	YES	NO	NO	YES
Year FE	NO	NO	YES	NO	NO	YES

[xi]

[xi] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Alcohol (extensive) is a binary variable indicating alcohol consumption in past seven days, alcohol (intensive) is the number of alcoholic drinks on the day with the most alcohol consumption in the last seven days. Marital dummies indicate the marital status and health dummies indicate the change in health compared to the previous year.

Table 13 Results of regressions of income quintiles on sports, extensive and intensive margin

VARIABLES	(1) Sports (extensive)	(2) Sports (extensive)	(3) Sports (extensive)	(4) Sports (intensive)	(5) Sports (intensive)	(6) Sports (intensive)
Second income quintile	-0.0575***	-0.0104	-0.00927	-0.499***	-0.321***	-0.0771
	(0.00962)	(0.00985)	(0.0122)	(0.0668)	(0.0687)	(0.0880)
Third income quintile	-0.0348***	0.0180*	-0.00853	-0.510***	-0.423***	-0.234**
	(0.0103)	(0.0106)	(0.0145)	(0.0709)	(0.0734)	(0.104)
Fourth income quintile	-0.0217**	0.0410***	-0.0138	-0.454***	-0.413***	-0.300***
	(0.0106)	(0.0112)	(0.0160)	(0.0723)	(0.0769)	(0.115)
Fifth income quintile	0.00944	0.0877***	-0.0112	-0.240***	-0.205**	-0.235*
	(0.0116)	(0.0125)	(0.0186)	(0.0792)	(0.0861)	(0.134)
Age		-0.00427***	-0.00196		-0.00909***	0.000215
		(0.000331)	(0.00512)		(0.00222)	(0.0368)
Gender		0.0322***	-0.0748		-0.584***	-0.354
		(0.0100)	(0.199)		(0.0666)	(1.428)
Constant	0.575***	0.737***	0.767***	2.578***	3.590***	3.113*
	(0.00821)	(0.0153)	(0.262)	(0.0554)	(0.102)	(1.881)
Observations	29,412	29,412	29,412	29,412	29,412	29,412
R-squared			0.006			0.006
Number of individuals	8,369	8,369	8,369	8,369	8,369	8,369
Marital Dummies	NO	YES	YES	NO	YES	YES
Health Dummies	NO	YES	YES	NO	YES	YES
Indiviudal FE	NO	NO	YES	NO	NO	YES
Year FE	NO	NO	YES	NO	NO	YES

[xii]

[xii] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Sports (extensive) is a binary variable indicating engaging in sports activities, sports (intensive) is the number of hours of sports played per week Marital dummies indicate the marital status and health dummies indicate the change in health compared to the previous year.

Table 14 Results of regressions of log of income on dependent variables, with interaction effects for gender and log of income

VARIABLES	(1) Smoking (extensive)	(2) Smoking (intensive)	(3) Alcohol (extensive)	(4) Alcohol (intensive)	(5) Sports (extensive)	(6) Sports (intensive)
Log of Income	0.000416	0.0401	0.00819*	0.0363	-0.00703	-0.101***
	(0.00242)	(0.0345)	(0.00476)	(0.0321)	(0.00440)	(0.0316)
Age	0.00451	0.0984**	0.000624	-0.0745**	-0.00196	0.000239
	(0.00282)	(0.0401)	(0.00554)	(0.0374)	(0.00512)	(0.0368)
Female	0.322***	1.135	0.267	0.146	-0.0994	-0.751
	(0.110)	(1.569)	(0.217)	(1.462)	(0.200)	(1.441)
Female*Log of Income (interaction effect)	-0.000943	-0.0487	-0.00312	-0.0380	0.00469	0.0757**
	(0.00281)	(0.0399)	(0.00552)	(0.0372)	(0.00510)	(0.0367)
Constant	-0.136	-2.223	0.591**	6.712***	0.802***	3.572*
	(0.145)	(2.057)	(0.284)	(1.917)	(0.263)	(1.889)
Observations	29,412	29,412	29,412	29,412	29,412	29,412
R-squared	0.018	0.017	0.004	0.005	0.006	0.006
Number of nomem_encr	8,369	8,369	8,369	8,369	8,369	8,369
Marital Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Health Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Indiviudal FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

[xiii]

[xiii] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Smoking (extensive) is a binary variable indicating currently smoking, smoking (intensive) is the number of cigarettes, cigars and pipes smoked per day. Alcohol (extensive) is a binary variable indicating alcohol consumption in past seven days, alcohol (intensive) is the number of alcoholic drinks on the day with the most alcohol consumption in the last seven days. Sports (extensive) is a binary variable indicating engaging in sports activities, sports (intensive) is the number of hours of sports played per week.

Table 15 Results for regressions of log of income on dependent variables, with interaction effects for age categories and log of income

VARIABLES	(1) Smoking (extensive)	(2) Smoking (intensive)	(3) Alcohol (extensive)	(4) Alcohol (extensive)	(5) Alcohol (extensive)	(6) Alcohol (intensive)	(7) Sports (extensive)	(8) Sports (intensive)
Log of Income	-0.00125	-0.0506	-0.00668	-0.00671	-0.00674	0.0207	-0.00643	-0.0106
	(0.00283)	(0.0402)	(0.00556)	(0.00556)	(0.00556)	(0.0375)	(0.00514)	(0.0370)
16-24 years old	0.0274	-0.778**	-0.275***	-0.255***	-0.270***	0.201	-0.0760	0.689*
	(0.0278)	(0.396)	(0.0547)	(0.0550)	(0.0548)	(0.369)	(0.0506)	(0.364)
25-34 years old	0.0871***	0.376	-0.193***	-0.194***	-0.194***	-0.672*	-0.0632	-0.314
	(0.0288)	(0.410)	(0.0566)	(0.0566)	(0.0566)	(0.382)	(0.0524)	(0.376)
45-54 years old	-0.000764	-0.548*	-0.0202	-0.0185	-0.0204	0.271	-0.00912	0.213
	(0.0228)	(0.325)	(0.0449)	(0.0449)	(0.0449)	(0.303)	(0.0416)	(0.299)
55-64 years old	0.00196	-0.405	0.000533	0.00376	0.000470	0.256	0.0431	0.618*
	(0.0249)	(0.354)	(0.0489)	(0.0489)	(0.0489)	(0.330)	(0.0452)	(0.325)
65 years old and older	-0.000211	-0.344	0.0125	0.0183	0.0127	0.177	0.0498	0.331
	(0.0332)	(0.473)	(0.0653)	(0.0653)	(0.0653)	(0.440)	(0.0604)	(0.434)
16-24 years old * Log of income	0.00466	0.141***	0.0273***	0.0243***	0.0270***	0.000245	0.00693	-0.0768*
	(0.00349)	(0.0496)	(0.00686)	(0.00691)	(0.00686)	(0.0463)	(0.00634)	(0.0456)
25-34 years old * Log of income	-0.00708*	-0.0349	0.0192**	0.0191**	0.0192**	0.0510	0.00696	0.0516
	(0.00394)	(0.0560)	(0.00774)	(0.00774)	(0.00774)	(0.0522)	(0.00716)	(0.0515)
45-54 years old * Log of income	0.00139	0.0655	0.00553	0.00552	0.00560	-0.0299	0.00343	-0.00355
	(0.00314)	(0.0447)	(0.00618)	(0.00618)	(0.00618)	(0.0417)	(0.00572)	(0.0411)
55-64 years old * Log of income	0.000495	0.0132	0.00218	0.00213	0.00224	-0.0177	-0.00226	-0.0362
	(0.00327)	(0.0465)	(0.00643)	(0.00643)	(0.00643)	(0.0434)	(0.00594)	(0.0427)
65 years old and older * Log of income	-0.00153	-0.0329	0.00116	0.00101	0.00121	-0.0111	-0.000205	0.0315
	(0.00437)	(0.0622)	(0.00860)	(0.00860)	(0.00860)	(0.0580)	(0.00795)	(0.0572)
Female	0.318***	0.893	0.248	0.236	0.245	-0.0545	-0.0768	-0.368
	(0.109)	(1.554)	(0.215)	(0.215)	(0.215)	(1.449)	(0.199)	(1.428)
Education (with diploma)					0.00462			
					(0.00393)			
Education (without diploma)				0.0199***				

				(0.00581)				
Constant	0.0713	3.168***	0.743***	0.670***	0.727***	3.244***	0.701***	2.734***
	(0.0607)	(0.864)	(0.119)	(0.121)	(0.120)	(0.806)	(0.110)	(0.794)
Observations	29,412	29,412	29,412	29,412	29,412	29,412	29,412	29,412
R-squared	0.020	0.018	0.006	0.007	0.006	0.005	0.006	0.007
Number of nomem_encr	8,369	8,369	8,369	8,369	8,369	8,369	8,369	8,369
Marital Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Health Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Indiviudal FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

[xiv]

[xiv] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Smoking (extensive) is a binary variable indicating currently smoking, smoking (intensive) is the number of cigarettes, cigars and pipes smoked per day. Alcohol (extensive) is a binary variable indicating alcohol consumption in past seven days, alcohol (intensive) is the number of alcoholic drinks on the day with the most alcohol consumption in the last seven days. Sports (extensive) is a binary variable indicating engaging in sports activities, sports (intensive) is the number of hours of sports played per week. Reference category for age is 35-44 year-olds.

Table 16 Results of regressions of log of income on smoking, extensive and intensive margin, on IACK subsample

VARIABLES	(1) Smoking (extensive)	(2) Smoking (extensive)	(3) Smoking (extensive)	(4) Smoking (intensive)	(5) Smoking (intensive)	(6) Smoking (intensive)
Log of Income	0.00469	-0.00295	0.0324	0.101	0.00834	0.0773
	(0.0145)	(0.0144)	(0.0291)	(0.167)	(0.167)	(0.242)
Age		0.000982	-0.00101		1.39e-05	0.0226
		(0.00294)	(0.138)		(0.0372)	(1.141)
Female		0.0538			0.946	
		(0.0504)			(0.693)	
Constant	0.143	0.168	0.170	1.072	2.054	0.406
	(0.102)	(0.170)	(5.332)	(1.183)	(2.116)	(44.19)
Observations	808	808	808	789	789	789
R-squared			0.013			0.032
Number of individuals	364	364	364	362	362	362
Marital Dummies	NO	YES	YES	NO	YES	YES
Health Dummies	NO	YES	YES	NO	YES	YES
Indiviudal FE	NO	NO	YES	NO	NO	YES
Year FE	NO	NO	YES	NO	NO	YES

[xv]

[xv] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Smoking (extensive) is a binary variable indicating currently smoking, smoking (intensive) is the number of cigarettes, cigars and pipes smoked per day. Sample is restricted to IACK subsample.

Table 17 Results of regressions of log of income on drinking, extensive and intensive margin, on IACK subsample

VARIABLES	(1) Alcohol consumption (extensive)	(2) Alcohol consumption (extensive)	(3) Alcohol consumption (extensive)	(4) Alcohol consumption (intensive)	(5) Alcohol consumption (intensive)	(6) Alcohol consumption (intensive)
Log of Income	-0.0132	3.97e-05	0.0646	0.206*	0.0566	0.256
	(0.0193)	(0.0193)	(0.0657)	(0.107)	(0.104)	(0.343)
Age		-0.0139***	-0.463		0.00833	2.390
		(0.00401)	(0.352)		(0.0217)	(1.835)
Female		0.135**			-1.972***	-
		(0.0646)			(0.350)	
		-0.0246			-0.619	
Constant	1.465***	1.815***	18.80	0.632	3.455***	-90.73
	(0.135)	(0.228)	(13.70)	(0.748)	(1.235)	(71.47)
Observations	664	664	664	664	664	664
R-squared			0.021			0.025
Number of nomem_encr	316	316	316	316	316	316
Marital Dummies	NO	YES	YES	NO	YES	YES
Health Dummies	NO	YES	YES	NO	YES	YES
Indiviudal FE	NO	NO	YES	NO	NO	YES
Year FE	NO	NO	YES	NO	NO	YES

[xvi]

[xvi] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Alcohol (extensive) is a binary variable indicating alcohol consumption in past seven days, alcohol (intensive) is the number of alcoholic drinks on the day with the most alcohol consumption in the last seven days. Sample is restricted to IACK subsample.

Table 18 Results of regressions of log of income on sports, extensive and intensive margin, on IACK subsample

VARIABLES	(1) Sports (extensive)	(2) Sports (extensive)	(3) Sports (extensive)	(4) Sports (intensive)	(5) Sports (intensive)	(6) Sports (intensive)
Log of Income	0.00365	0.00329	-0.114**	-0.103	-0.150*	-0.505***
	(0.0201)	(0.0208)	(0.0447)	(0.0845)	(0.0848)	(0.177)
Age		0.0104**	-0.0356		0.0629***	-0.426
		(0.00411)	(0.211)		(0.0169)	(0.836)
Female		-0.0477			-1.103***	-
		(0.0687)			(0.283)	
Constant	0.510***	0.0737	2.394	2.219***	0.463	19.77
	(0.141)	(0.240)	(8.178)	(0.593)	(0.984)	(32.40)
Observations	803	748	748	803	748	748
R-squared			0.067			0.121
Number of nomem_encr	354	339	339	354	339	339
Marital Dummies	NO	YES	YES	NO	YES	YES
Health Dummies	NO	YES	YES	NO	YES	YES
Indiviudal FE	NO	NO	YES	NO	NO	YES
Year FE	NO	NO	YES	NO	NO	YES

[xvii]

[xvii] Standard errors are in parentheses. Stars indicate significance level; ***p<0.01, **p<0.05, *p<0.1. Data is obtained from the LISS panel, for the period 2008-2015 (without 2014). Sports (extensive) is a binary variable indicating engaging in sports activities, sports (intensive) is the number of hours of sports played per week. Sample is restricted to IACK subsample.