# Can receiving a serious health update help individuals suffering from tobacco addiction? An examination into the Dutch population's behavioural reactions to health shocks.

**IBEB Bachelor Thesis, Final Draft**

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Department of Economics

Supervisor: Tom Van Ourti

Name: Andrea Mudd
Student number: 368908
E-mail address: mudd.andrea@gmail.com

***Abstract***: This research examines a panel of Dutch individuals' reactions to the diagnosis of a serious health condition in terms of smoking behaviour, analysing the presence of the effects of unstable steady states on addiction to tobacco products. The elements of addiction of participation elasticity and conditional demand elasticity are represented by a person being a smoker (or not) and the amount of cigarettes smoked by an individual, respectively. The effects of a health shock on smoking are empirically modeled focused on within-individual changes, and the primary outcome is that individuals who were smokers at the time of diagnosis have a slightly increased likelihood of halting their participation in smoking along with slightly decreasing their level of addiction to cigarettes. The effect of a diagnosis on the smoking behaviour of those not actively smoking is positive (individuals are more likely to pick up the habit), but statistically negligible. A diagnosis can thus be seen as an opportunity for targeted changes in behaviour for those suffering from tobacco addiction.

## Table of Contents

## INTRODUCTION

One of the most common human tasks is the decision to consume. The choices humans make regarding the goods, services, activities, and lifestyles they consume directly affect their health; in modern times, the consumption choices of the industrialised world have led to a high prevalence of self-induced ailments. While morbidity and mortality used to be primarily due to infectious diseases, they are now mostly related to chronic diseases (Cawley, Ruhm, 2011). Health (and death) is thus largely related to external modifiable risk factors that individuals may choose to consume in different quantities and with different frequencies.

One good that is highly consumed despite its being proven to be harmful to health and addictive is tobacco. Tobacco products have negative effects on nearly every organ system in the human body and strongly increase the risk of health crises such as stroke, heart attack, and many types of cancers (Center for Disease Control and Prevention, 2009). Despite these destructive effects being widely known since the middle of the twentieth century, 1 billion people continue to use tobacco products worldwide (Chaloupka, Warner, 2000). Tobacco has in fact been referred to as a public health plague by many entities and is expected to become the leading cause of death worldwide in the very near future (ITC Project, 2010; Chaloupka, Warner, 2000). In the Netherlands, 25% of the population 15 years and older smoked tobacco products and 20% reported being daily smokers in 2012 (WHO, 2013). The consumption of tobacco products has significantly decreased in the Netherlands since 1995 but remains higher than the Dutch Ministry of Health's goal of below 20% in 2010 (ITC Project, 2010). While this country is highly industrialised and has high income and education levels, the Dutch are reported to have strong pro-tobacco attitudes- only 22% of smokers have a 'negative' or 'very negative' opinion of smoking- and are not as respondent to the government's efforts to reduce smoking levels (messages on cigarette packages, taxes, bans in the hospitality industry) as policy-makers would like (ICT Project, 2010).

Since the Netherlands is a country with a high quality of education and substantial anti-smoking laws and regulations, addiction appears to be an important factor which keeps the Dutch smoking. Cigarettes contain the addictive substance nicotine, which temporarily relieves stress and produces a calming sensation (American Heart Association, 2011). Individuals may have difficulty quitting because of the physical symptoms of withdrawal from nicotine but also due to quitting costs such as the unpleasant time and effort it requires to alter one's behaviour. These extra quitting costs can help explain the inconsistency between tobacco smokers' desire to quit and their continued consumption of the product (Suranovic et al, 1999).

Smokers are consuming a good that is evidently harmful to their health and have difficulty quitting due to the product's addictive nature; this is a problem continually addressed by medical professionals, policy-makers, and the many smokers who are trying to quit.

For the purposes of this research, cigarettes will include both industrial cigarettes and loose rolling tobacco; the term cigarettes will also be interchanged with 'tobacco products', and the term 'smoker' will signify users of cigarettes as previously defined.

Tobacco is a convenient addictive good to use in empirical analyses due to its legality and thus the availability of data regarding its prices and quantities. However, this research could also give some insight into the effect of health shocks on the use of other, even more dangerous and harmful, addictive goods such as opiates and amphetamines. While the total elimination of the consumption of harmful addictive goods is quite unrealistic, investigating a way to help those who may desire to quit smoking or using other substances could bring about social benefits.

The economists Gary Becker and Kevin Murphy suggested that strenuous events in one's life could affect their demand for addictive goods and that life events could affect consumption similarly to changes in price (Becker, Murphy, 1988). A health shock, such as learning one is diagnosed with a serious chronic disease, might induce dramatic behaviour changes in response to this shock. Reactions to a serious diagnosis in relation to addictive goods could be to binge, meaning to consume a very large amount of the good, or abstain, completely halt consumption of the good (Chaloupka, Warner, 2000). With regard to smokers, a new diagnosis may lead to altered demand for tobacco products and even the ability to quit. Especially in the already highly-educated Dutch population, this alternative (albeit unfortunate) method of informing an individual of the harmful effects of addictive tobacco products through the diagnosis of a serious disease could potentially increase their likelihood of reducing or stopping their consumption of tobacco. On the other hand, a serious health update could also cause an individual to succumb to addiction due to the stressful event. The essential research question to investigate is therefore:

*How can health shocks in the form of a diagnosis alter individuals' behaviour with regard to the consumption of cigarettes in the Netherlands?*

The partial questions to be examined in order to reach an answer to the above question involve two facets of addiction. Addictive goods can be characterised by two types of price elasticity, participation elasticity and conditional demand elasticity (Chaloupka et al, 2003). The first refers to the extent to which the price influences the decision to be a smoker. Applied to this research, the diagnosis of a health condition can signify a change in the non-monetary price of consuming tobacco products, and the rate of participation whether or not a person consumes tobacco products. Conditional demand elasticity refers to the extent to which price influences the amount of cigarettes smoked. Examining the rate of participation and conditional demand following a diagnosis can give insight into which direction,  abstinence or binging, individuals lean towards when presented with a health shock.

Smoking behaviour will be classified into three groups. Current smokers are individuals who identify themselves as smokers at a given point in time, former smokers are those who report that they smoked at one time but that they have quit, and non-smokers are those who have never smoked cigarettes. The distinction is made because stressful life events such as a diagnosis are likely to affect individuals in each of these groups differently over time. The use of these groups is also useful in measuring the rate of participation and the movement of an individual from one group to another.

The behaviour of individuals who do not currently smoke will first be considered. Those who have never before smoked and receive a health shock might find themselves in a stressful and vulnerable position, and they could be more likely to pick up the habit of consuming addictive goods such as cigarettes. Indeed, coping with stress due to different types of difficult situations is often cited as a reason for which adults pick up the bad habit (Jacobs, 1997). On the other hand, since these individuals were not previously prone to tobacco addiction in their lives, it is also possible that a health shock will not affect their smoking behaviour at all.

Former smokers may be more vulnerable to smoking than non-smokers when receiving a health shock. Though not directly related to health updates, the 9/11 attacks in the United States were reportedly associated with around one million former American smokers resuming and maintaining their consumption of cigarettes, an impressive and costly effect of the strenuous national event (Medical News Today, 2013). For those who used to smoke but state that they have quit, the shock of the serious health update could induce them to binge and revert back to their old habit of smoking. Similarly to the non-smokers, a health shock might also not affect the former smokers' decision and efforts to quit their cigarette addiction.

For current smokers, receiving a health update could push them in either direction, towards abstinence or binging. The diagnosis of a serious health condition raises the implicit price of smoking so is likely to affect the severity of their addiction, illustrated by the quantity of the addictive good consumed. A diagnosis can be seen as a teachable moment, an opportunity for smoking cessation. For newly diagnosed lung and otolaryngeal cancer patients, cessation rates ranged from 40% to 96% versus 1% to 5% in the general population (Westmaas et al, 2015). These types of cancers are very closely linked to smoking, and this research aims to see whether this effect can be extended to serious diseases in general as the stressful event can be just as jarring, such as the effect of the 9/11 attacks. Smith et al found that smokers had a high perceived risk of, especially, smoking-related health shocks, and that they adjusted their longevity expectations accordingly (Smith et al, 2001). It is then also plausible that a serious health shock's decrease on an individual's perceived life expectancy might lead to that individual deciding to maintain or increase their consumption of cigarettes since they are likely to have a shortened life anyways due to their disease.

Individuals in all three groups of smoking behaviour might alter their rate of participation of smoking, and those who were smokers when they received the health shock might additionally change their conditional demand for cigarettes.

A review of the literature on the topics of the dynamics of smoking and addiction, different models of addiction, and the impact of stress on addiction and smoking in particular will first be presented. Subsequently, the data will be introduced and its application to the research question will be made clear. The proposed research process will be explained, followed by the presentation of the results of the empirical analyses. The interpretation of these results will follow, along with a conclusion and a discussion regarding policy recommendations and future research possibilities.

## LITERATURE REVIEW

### *Dimensions of addiction*

Alfred Marshall's discussion of the effects of addiction on demand laid out the framework for further development of the dimensions of addiction, which are used by economists in their analyses in current times. He wrote, "habits which have once grown up around the use of a commodity while its price is low, are not quickly abandoned when its price rises again" (Marshall, 1920). He uses the example of cotton during the first World War, and he refers to the concept of withdrawal and implicit price increases which add to increased monetary prices due to the development of habits or dependencies on the goods while their price is low.

The three basic dimensions of addiction which took root in Marshall's work are tolerance, withdrawal, and reinforcement (Chaloupka et al, 2003). Tolerance refers to the gradual adaptation to a certain level of consumption, meaning that over time a given level of current consumption becomes less satisfying compared to a higher past consumption. Applied to tobacco products, one would need to smoke more cigarettes today to achieve the same level of pleasure (utility) as yesterday. Withdrawal captures the physical and psychological discomforts or dissatisfaction brought about when current consumption of the addictive good is terminated; utility in fact decreases when current consumption is equal to zero. Common withdrawal symptoms of cigarette-smoking cessation include insomnia, irritability, depressed mood, difficulty concentrating, restlessness, decreased heart rate, increased appetite or weight gain (US Department of Health and Human Services, 1998). Reinforcement refers to the positive effects of habits, meaning that a greater current consumption of a good like cigarettes will raise its future consumption; likewise, greater past consumption raises the marginal utility of current consumption.

These dimensions of addiction are perceived as relevant for the consumption of cigarettes and will be used throughout this research. They capture the behaviours of smokers that need to be incorporated into economic models of addiction in order for these models to be representative and applicable to real-life situations. Until recently, economists disregarded the consumption of addictive goods as it was considered to be completely irrational behaviour (who would choose to smoke even one cigarette if they were fully informed and aware of the future consequences?). In the past half century addictive behaviour has been modeled according to economic principles based on different levels of rationality: imperfectly rational addiction models, myopic addiction models, and rational addiction models. These models will be briefly presented to demonstrate the differing methods of considering addiction in an economic setting and to show the logic and theory behind the hypotheses of this paper.

*Brief overview of imperfectly rational and myopic models*

Imperfectly rational models consider consumers of addictive goods to be naive. Elster describes addiction as an instance of weakness of will and as an exercise in self-control (Elster, 2000). Models which consider addiction to be an imperfectly rational behaviour have in common that short-run and long-run preferences are perceived as stable but inconsistent. This entails that the more farsighted part of one's personality might aim to achieve certain goals regarding the consumption of addictive goods while the shortsighted part sabotages these by indulging in addictive goods. Schelling describes this with relation to smoking as follows: "Everybody behaves like two people, one who wants clean lungs and long life and another who adores tobacco" (Schelling, 1978). These models describe the demand for addictive goods as depending solely on the current price (not past or future prices) and have not been applied empirically to cigarettes.

Myopic addiction models go one step further by including the consideration of the past when making the decision to consume an addictive good. These models acknowledge that current actions depend on past choices but ignore the impact of past and current choices on future decisions (Chaloupka et al, 2003). Past choices are modeled by addictive stock, the depreciated sum of all past consumption of the addictive good, introduced by Houthakker and Taylor (Houthakker, Taylor, 1966). A higher addictive stock translates into a higher level of addiction. Myopic models lead to asymmetric responses to changes in price and income when dealing with addictive goods; applied to cigarettes, smoking was found to respond to price decreases about twice as much as to increases (Young, 1983). According to these models, cigarettes are found to have small negative elasticities (around -0.4), meaning that when price rises, demand falls less than proportionately (Lewit, Coate, 1982). This negative inelasticity is thus characteristic of addictive goods as demand does not fall as much as it should when price goes up, suggesting that the habit, or addiction, is stronger than the influence of price. Myopic models do allow for the empirical interpretation of smoking behaviour.

*Overview of rational addiction model*

In economics, rationality implies that actors are responsive to incentives (Cawley, Ruhm, 2011). With regards to addiction, past, current, future prices and consumption are the main incentives to consume along with restricted access and various social pressures. The rational addiction model goes several steps further than the imperfectly rational and myopic models by accounting for the effect of future prices on current consumption, which also allows for the inclusion of the dimensions of addiction in the model's interpretation. They incorporate an individual's consideration of future consequences in their utility maximization during their life cycle; utility at a given point in time depends on the current consumption of the addictive good, the current consumption of all other non-addictive goods, and the stock of past consumption of the addictive good as previously defined (Becker, Murphy, 1988). The dimensions of addiction are included as restrictions to the model, nicely summarised by Becker, Grossman, and Murphy (Becker et al, 1991). Tolerance is defined as the marginal utility of the stock of past consumption being negative, showing that the more addicted one is, the less utility one gains for each extra unit and the more needed to attain the same level of utility as previously. In the framework of the rational addiction model, reinforcement requires that an increase in addictive stock lead to an increased marginal utility of current addictive behavior that is larger than the negative effect of a higher past consumption on the future harm of the greater current consumption. Plainly put, greater consumption of a good raises its future consumption. Withdrawal is defined as total utility falling when consumption of the addictive good stops (equals zero), showing the short-term displeasure associated with being deprived of the substance.

Several important hypotheses are deduced using this model. Being addicted is defined in this analysis as an increase in an individual's present consumption of the addictive good leading to a higher future consumption of the same good (Becker, Murphy, 1988). The notion that today's consumption be adjacent to tomorrow's and yesterday's consumptions, called adjacent complementarity, is closely related to reinforcement. This property of addictive consumption is the basis for the following deductions. Firstly, current consumption is inversely related to current price and also other periods' prices. This stems from the price elasticity of demand mentioned during the discussion of the myopic models along with adjacent complementarity, since a raised price in one period will not only negatively affect the demand in that period but in all others before and after. Another argument which follows from adjacent complementarity is that the long-run price elasticity is greater than short-run price elasticity, meaning that the long-run effect of a permanent price change will exceed its short-run effect (Chaloupka, Warner, 2000). The final important hypothesis derived from adjacent complementarity in the rational addiction model is that anticipated price changes have a larger effect on consumption than unanticipated ones along with permanent price changes having a larger effect on demand than temporary ones. This requires the actors to be able to anticipate and incorporate future price changes into their current decisions, which may be essential to this research with respect to the ability to take into account the future raised price of smoking once diagnosed with a serious health condition.

Though this research will not empirically apply the rational addiction model, many of its key components are relevant to the development of the hypotheses regarding the effect of an implicit price change on cigarette consumption. The models of addiction are especially useful for determining whether a good is addictive, and this has been done extensively with regard to cigarettes. The rational addiction model has previously been applied to smoking in several instances. Chaloupka was the first to use the model empirically, and he obtained long-run price elasticities of demand of smoking in the range of -0.27 to -0.48, a stronger response than was obtained using conventional demand functions (Chaloupka, 1991). Smoking was found by Becker, Grossman, and Murphy to be addictive in a non-myopic way but did not find sufficient evidence for the argument of full rationality when using state-level data (Becker et al, 1994). Several studies using aggregate-level data found that smoking behaviour is in line with the rational addiction theory (Chaloupka, Warner, 2000). This research considers cigarettes as addictive goods and attempts to determine the relationship between consumption decisions and newly acquired information related to health conditions so theoretically makes use of many facets of addiction models (in particular, the rational addiction model) without explicitly applying them to the data and empirical models.

*Unstable Steady States*

A key component of Becker and Murphy's model especially relevant to this research is the unstable steady state, which is a result of strong adjacent complementarity. Grossman describes this notion as a price change causing people to start or stop consuming the addictive good or to change their consumption drastically (Grossman, 1995). Contrary to a common occurrence of a steady state, steady-state consumption of addictive goods is unstable when the degree of addiction is strong, that is, when the complementarity between past and current consumption is strong.

Unstable steady states help to explain the phenomena of binging and quitting "cold turkey". This notion suggests that even a modest increase in cigarette prices might be a strong enough incentive to induce a smoker who happens to be near unstable steady states to cease (Grossman, 1995). Also suggested by Grossman is that permanent policies or price increases are more likely to have significant effects with regard to unstable steady states. Applied to this research, the diagnosis of a chronic condition can be viewed as a permanent implicit price change (as opposed to a temporary illness) that is more likely to have a significant effect on one's smoking behaviour.

Douglas wrote, in the context of the rational addiction model, that price changes have a stronger effect on cessation than on the initiation of smoking (Douglas, 1998). He does not, however distinguish between people who have never smoked and those who were former smokers and also refers to price as a monetary component combined with smoking regulations. For these reasons, this research makes use of his reasoning regarding smoking cessation but also postulates that non-smokers and former smokers could be initiated to the habit following an increase in price in the form of a serious health update.

Becker and Murphy stress the importance of considering unstable steady states in any analysis dealing with addiction (Becker, Murphy, 1988). They argue that unstable steady states are crucial to explain rational "pathological" addictions, meaning that an individual continues to increase his or her consumption of the addictive good over time despite being fully aware of the future. They are necessary to deal with "normal addictions" as well which can also involve temporary rapid increases in consumption. Put otherwise, unstable steady states explain why the same person can be heavily addicted at one point in time and yet lay off the addictive good at another. A strong statement made by the economists, individuals who are more heavily addicted quit more abruptly due to the presence of unstable steady states. They state that an individual will cease his or her consumption when a method of raising the long-term benefits of doing so sufficiently above the short-term costs of adjustment is found. Regarding binging, Becker and Murphy explain this behaviour by the inclusion of a second addictive stock with a different depreciation rate than the first and the occurring differences in degrees of complementarity and substitutability (they use the example of overeating: one stock represents eating capital and the other weight) . They argue that binges are the outcome of consistent maximisation over time that takes into account the effects of increased current consumption of the addictive good on both future addictive stocks.

The unstable steady states, namely those generated by important health updates, are the basis of this research's premise, which takes into account the cited authors' analyses of their potential to induce quitting, initiation, or a strong change in the amount of the good consumed.

### *Stressful events' effect on consumption and price of addictive goods*

Much research has been done on the topic of how significant stressful events affect the way that individuals choose to consume both addictive and non-addictive goods.

The behavioural responses to health shocks were studied in the optic of changes in diet upon the diagnosis of diabetes (Oster, 2014). This research found that although patients were closely informed about the benefits of changing their diet, individuals often placed a higher value on their preferred diet (lifestyle) than on their health. Also suggested by Oster is that behavioural changes may be of even larger benefit in the years prior to the diagnosis to prevent the onset of diabetes. Another finding from this research was that increases in 'good' foods were more difficult to sustain than decreases in 'bad' foods, suggesting that bad habits are more difficult to break than picking up additional habits which are beneficial to one's health. A study was also conducted concerning medical testing's effect on individuals' beliefs about their health. Oster et al found that individuals not yet tested (but at risk) for Huntington disease remain optimistic about their health and behave similarly to individuals who are not at risk with regard to retirement choices, for example (Oster et al, 2013). They found that individuals only adjust their behaviour once the test confirms that they carry the Huntington disease expansion, suggesting that individuals might also only change their smoking-related behaviour once they are diagnosed with a serious disease and can no longer believe that they are completely healthy.

Generally, several stress forms including childhood abuse, household dysfunction, parental divorce, negative life events, and acute and chronic stresses have been shown to affect smoking habits (Kassel et al, 2003). Stress is viewed as one of the most important reasons why people smoke and is also one of the most important causes of relapse. Kassel et al found that anywhere from 35% to 100% of smokers report that they lapsed due to some form of stress, based on several retrospective, self-reporting methodologies (Kassel et al, 2003). Though this is quite a wide range of response to stress, even 35% of former smokers relapsing due to stress is significant, and understanding this mechanism and possible ways to avoid or help this are imperative. A reason cited for the too-common resurgence in smoking to cope with stress is that drugs facilitate general mood regulation, providing short-term relief despite the long-term consequences of the substance abuse (Wills, Shiffman, 1985).

Michael Darden develops a model of lifetime smoking behaviour which takes health markers (blood pressure and cholesterol, for example) and what he refers to as health transitions (health exams, chronic health shocks, and transitioning to a health marker index above the 75th percentile in the population) into account (Darden, 2011). He argues that individuals use their endowment of information about their respective health markers to make choices related to smoking to maximise their utility. This knowledge of one's health markers enables a degree of awareness of future health transitions. The researcher distinguishes between heavy smokers, light smokers, and non-smokers. In terms of smoking habits in response to health updates, Darden finds that of those individuals who were considered heavy smokers prior to a chronic health shock, 70.24% and 47.06% were still smoking heavily one and three periods later, respectively. He also observed that some non-smokers at the time of a chronic health shock had taken up the habit three periods later, namely 2.83%. The trend in all smoking groups that many more individuals had quit three periods after a health transition and that a small but rising amount had taken up the habit three periods after a health transitions suggests that this behaviour change may not be directly linked to the diagnosis.

Research on whether smokers try to quit after a cancer diagnosis was performed and revealed that 70% of smokers continued smoking after two years of diagnosis and 57% continued after four years (Westmaas et al, 2015). These results suggest that a cancer diagnosis is indeed a motivator to fight the disease and improve habits related to addictive goods. This study considered all cancers and did not distinguish between those directly related to smoking and those not. The study does, however, give insight on the affect a serious diagnosis can have on smoking, as Westmaas stated: "we know a fair amount about smoking's effect on cancer but relatively less about cancer's effect on smoking" (Dotinga, 2015). A more specific study was recently conducted focused on the effect of bladder cancer diagnosis on smoking behaviour (Basset et al, 2012). The authors cite bladder cancer as the second most common smoking-related malignancy. They divided their sample into the three groups "never smokers", "former smokers", and "active smokers", similar to the three groups that this research will make use of. The researchers found that smokers with a new diagnosis were about five times more likely to quit than smokers in the general population (48% versus 10%). Interestingly, the researchers were able to ask the respondents about their reasons for quitting; diagnosis and advice from their urologist were the reasons most commonly cited. While bladder cancer is unequivocally linked to smoking, the respondents' awareness of this relationship was largely dependent on the frequent interaction with their urologist. Additionally, of the smokers who quit, 62% did so without any cessation aids, suggesting that the diagnosis has potential as a teaching moment and may be enough of an incentive to quit. This cross-sectional analysis isolates the impact of this serious diagnosis on smoking behaviour and has the advantage of having access to very detailed information from the respondents, especially with regard to the reasons for their smoking cessation.

Smith et al examine reactions to both directly smoking-related and not directly smoking-related health shocks on individuals' longevity expectations (Smith et al, 2001). The authors' main finding was that smokers react to smoking-related health shocks by negatively altering their longevity expectations but do not react comparably to general health shocks. Former smokers and non-smokers, however, reacted to a much wider range of health-related signals than smokers. Smith et al also consider exogenous health shocks to provide new information that is important for the respondents' perception of their health. They cite Fischloff, Bostrom, and Quadrel, who wrote that people make distinctions between their perception of the average population's risk of an adverse event and their personal likelihood of experiencing a similar outcome (Fischloff et al, 1993); this can help explain why smokers may need an adverse event such as a health shock to change their risk perception. Smith et al's study makes use of panel data, and this study will draw conclusions in a parallel fashion since the nature of the data assures that health shocks were experienced before each respondent reported their longevity expectation (or changed their smoking behaviour).

In sum, all of the cited authors find correlations between health shocks and behaviour related to the act of smoking and its consequences. A recurring theme in the literature is the use of diagnosis as an opportunity to inform individuals about the hazards of smoking, and personalized contact with medical professionals is seen as an effective way to accomplish this. The result that smokers react strongly to smoking-related diagnoses suggests that personal messages are able to induce life changes as they are relevant to their circumstances (Sloan et al, 2003). Stress can be considered a powerful trigger, and this trigger can cause cessation or initiation depending on the current smoking habits of the individual and on the nature of the stress along with the individual's environment (Medical News Today, 2013). In terms of methodology, separating individuals into groups based on their smoking habits is common and will also be done in this research, and some researchers also distinguished between smoking-related and not smoking-related diseases. The distinction between smoking-related and not smoking-related diseases (or any distinction between types of diseases) is not made here. This is because this research's focus is on the diagnosis being a stressful event for the individual, and this is thought to be independent of whether the disease is smoking-related or not. This research attempts to learn from these papers and to examine health shocks in a broader sense and their effect on actual smoking behaviour.

## DATA & METHODOLOGY

Details concerning the data and its use with regard to the research will be described below, followed by a brief outline of the principal theories used to test the hypotheses along with its specific methodological application to this research.

### *Data sources*

Panel data from households in the Netherlands will be used to perform this research. The data is obtained from the LISS panel (Longitudinal Internet Studies for the Social sciences)[1] which consists of 8000 individuals in 5000 households aged 16 and above. The panel is representative of the Dutch population as it is drawn randomly from the Statistics Netherlands population register. The longitudinal study consists of seven annual waves, running from November 2007 until December 2013, meaning the same individuals have completed the surveys each year for seven years. The panel is unbalanced as certain individuals did not complete the survey during each of the seven years.

[1] http://www.lissdata.nl/dataarchive/study_units/view/1

The relevant information is drawn primarily from the core study category 'Health' along with 'Background Variables'. The health section contains information regarding whether or not an individual has certain health conditions, many self-assessment questions dealing with health, the frequency with which an individual visits a physician, and detailed information about an individual's behaviour with regard to the consumption of tobacco products. The section including background variables gives insight into the participants' ages, household situations, civil statuses, monthly income, education level, and other common demographic indicators.

Using panel data is advantageous in this study since it allows for controlling for variables that are unobservable or unmeasurable (Baltagi, 2008). Important variables that change over time but not across entities are able to be controlled for using time dummies, namely the monetary price of tobacco, along with individual heterogeneity.[2] One disadvantage of this particular panel data is that many of the key variables used are self-reported and thus risk human error or researcher bias.

*Variables used in the analysis and modifications to the data*

Using the LISS panel data with the responses from the seven waves combined, the necessary variables were extracted and formulated. All modifications and analyses were conducted using the statistical software STATA 13.0. The dependent variables to test the respective hypotheses, amount of tobacco product use and whether or not tobacco products are used, is based on existing variables in the database. The daily amount of cigarettes smoked is an integer variable, ranging from 0 to 250 in this particular data. One observation was removed as it is humanly impossible to smoke 2000 cigarettes per day; this was classified as a mistake. Three other observations were removed based on the answers given regarding the number of visits to the physician in the past twelve months (responses above 400 were removed since it is also extremely improbable, if not impossible, to visit the doctor many more than one unique time each day during an entire year).

[2] Baltagi, Badi. (2008). *Econometric analysis of panel data*. John Wiley & Sons, 2008. Ch.1.

The amount of cigarettes consumed was used as it is given in the database. The variables indicating whether a person is a smoker, a former smoker, or a non-smoker were created from answers to survey questions regarding whether a person has ever smoked and, if so, if they are currently smoking. If a person answered "no" to the first question, the dummy variable for non-smokers takes a value of 1 and 0 otherwise. If a respondent answered "yes" to the first question but "No, I quit" to the second, the dummy variable for former smokers takes a value of 1 and 0 otherwise. Lastly, if a person answered "yes" to both questions, the dummy variable representing current smokers takes a value of 1 and 0 otherwise.

If an individual reported being a former smoker, the amount of cigarettes he or she used to smoke was still reported. However, this amount will not be considered when analysing whether current smokers who are diagnosed change the quantity of cigarettes they consume.

A crucial variable generated for the analyses is an indicator of diagnosis. This dummy variable takes the value of 1 if an individual reports that a physician told him or her during the last year that he or she suffers from one of the serious diseases mentioned in the survey which they did not report the previous year. The diseases taken into account include angina, a heart attack, high blood pressure, high cholesterol, a stroke, diabetes, chronic lung disease, asthma, arthritis, cancer, an ulcer, Parkinson's disease, cataract, a broken hip, another fracture, Alzheimer's disease, a benign tumor, or another serious disease not mentioned in the survey ("other"). To accomplish this, an indicator of the wave number was first created (ranging from 1 to 7 based on the date of the entry). Next, an indicator called 'Sick' was created, taking the value of 1 if a person responded "yes" to having any of the diseases listed above, and 0 otherwise. The data was then reshaped so that each individual's answers were in a single row for all seven waves, which also created a variable 'SickX' for each wave, with X between 1 and 7. The dummy variable indicating diagnosis was then generated, taking a value of 1 if an individual reported being sick in a given period but did not in the period prior, and 0 otherwise. A diagnosis is therefore possible starting in the second wave, as the health conditions respondents suffered from prior to 2007 are not known. The indicator for diagnosis along with the identifying variable were then isolated and saved as a new database, then merged with the original database to obtain an indicator of diagnosis over time for each individual.

An overview of the key variables and their definitions can be found in *Appendix A, Table 1*.

Control variables included to test for demographic heterogeneity and to control for spurious effects are gross monthly income, education (highest diploma earned), age, number of children living at the individual's home, civil status, number of visits to the doctor in the past year, and frequency of alcoholic drinks in the past (see *Appendix A, Table 1*). These variables were included based on their perceived effect on smoking behaviours, both from theoretical standpoints along with their relationship to the amount of cigarettes smoked by individuals in the panel seen with scatterplots. Indicators such as gender which do not change within individuals over time are not included since this research's focus is on within-individual changes.

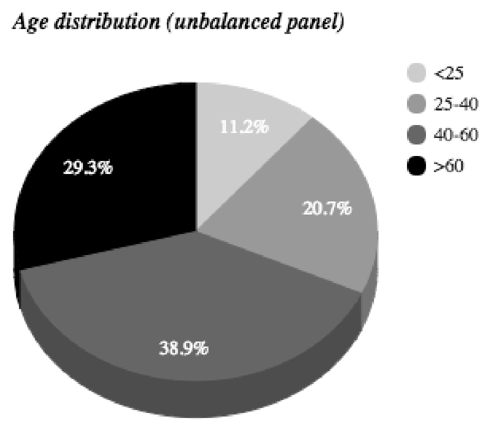
All analyses are conducted with the data set as panel data, with the identifying number of the respondent as the panel variable and the wave number indicator as the time variable.

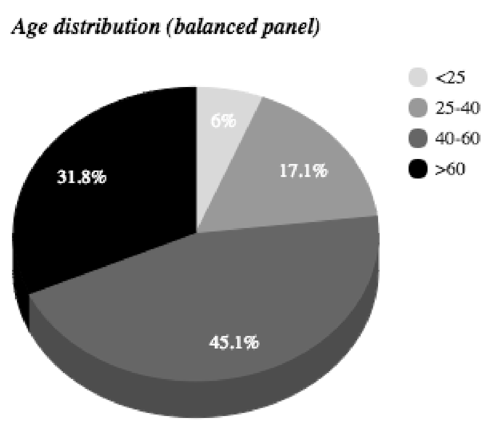
*Balancing the panel data*

In total, the original combined dataset had 40,272 unique observations. Upon closer look at the data, there is a high level of attrition; only 17,458 unique observations remain if the individuals are required to have participated in all seven waves of the survey. This represents a 57% decrease in the sample size, so it is unlikely that the selection of individuals in the unbalanced panel is random and selection bias is likely to be present.

Among important differences between the balanced and unbalanced panel are differences in age, number of diagnoses, and smoking habits. The difference in age distribution is illustrated in *Graph 1* on page 22. The balanced panel has an average age almost four years older than the unbalanced panel (52 and 48, respectively), which suggests that perhaps younger people are not motivated to continue filling in the survey year after year, among other reasons individuals could decide to discontinue their participation. In the balanced panel, 1,109 diagnoses take place during the seven waves compared to the 1,983 in the unbalanced panel. This represents a decrease of 44% which is less than proportional to the sample size decrease of 57%. The distribution of the diagnoses over time is similar in the two panels, with more individuals diagnosed during waves 2 and 3 than the later ones. The balanced panel is comprised of slightly less smokers and more former smokers, and both the balanced and unbalanced panels suffer from individuals giving negative answers to belonging to any of the groups (though for the balanced panel this number is very low: 18 out of 17,458 instances). Both the balanced and unbalanced panels also suffer from a slight increase in the number of non-smokers over time, which should theoretically be impossible as the question asked to respondents is whether he or she has *ever* smoked so should not be "yes" in one period and "no" in the next. Therefore, this could be due either to individuals misunderstanding the survey questions or wanting to be associated with being a non-smoker even though in previous years they may have consumed a small amount of tobacco products. Of note is that in the balanced panel the change over time of non-smokers is, on average, 0% so this error is quite small though still of importance (see *Appendix B, Table 1*).

<u>*Graph 1*: Difference in age distribution between balanced and unbalanced panels</u>




Due to the very high attrition rate in the unbalanced panel along with significant differences in the composition of the balanced panel compared to the unbalanced panel, the balanced panel will be used in the subsequent analyses. Summary statistics of the key variables from the balanced panel can be found in the following section, and those from the unbalanced panel can be found in *Appendix F*.

*Summary statistics of key variables*

*Table 1* and *Table 2* below show the descriptive statistics for the basic categorical and continuous demographic variables, respectively.

*Table 1*: Basic population demographics, categorical variables (balanced panel)

| Variable Name | Levels | Percentage | Frequency | Total Responses |
|---|---|---|---|---|
| Gender | Male | 48.09 | 8,395 | 17,458 |
| Civil Status | Married | 67.36 | 11,759 | 17,458 |
| | Separated | 0.41 | 72 | |
| | Divorced | 8.00 | 1,397 | |
| | Widow or Widower | 4.36 | 7610 | |
| | Never been married | 19.87 | 3,469 | |
| Highest level of education with diploma | Primary School | 4.12 | 719 | 17,458 |
| | VMBO (intermediate secondary school) | 26.62 | 4,647 | |
| | HAVO/VWO (higher secondary education) | 9.48 | 1,655 | |
| | MBO(intermediate vocational education) | 23.70 | 4,137 | |
| | HBO (higher vocational education) | 24.58 | 4,291 | |
| | WO (university) | 7.67 | 1,339 | |
| | Other | 2.67 | 467 | |
| | Not completed any education[1] | 1.04 | 182 | |
| | Not yet started any education[1] | 0.12 | 21 | |
| Frequency of drinks, last 12 months | Almost every day | 18.32 | 3,195 | 17,440[2] |
| | Five or six days per week | 6.46 | 1,127 | |
| | Three or four days per week | 12.13 | 2,116 | |
| | Once or twice a week | 22.55 | 3,933 | |
| | Once or twice a month | 13.38 | 2,334 | |
| | Once every two months | 7.15 | 1,247 | |
| | Once or twice a year | 9.07 | 1,581 | |
| | Not at all over the last 12 months | 10.93 | 1,907 | |

[1] These answer categories were changed in December 2008 and were no longer offered as options to respondents after that time (starting from wave 3).
[2] Drinking habits were not indicated in the survey in 18 instances. As this is a categorical variable, it is not possible to replace missing values with a default category.

*Table 2*: Basic population demographics, continuous variables (balanced panel)

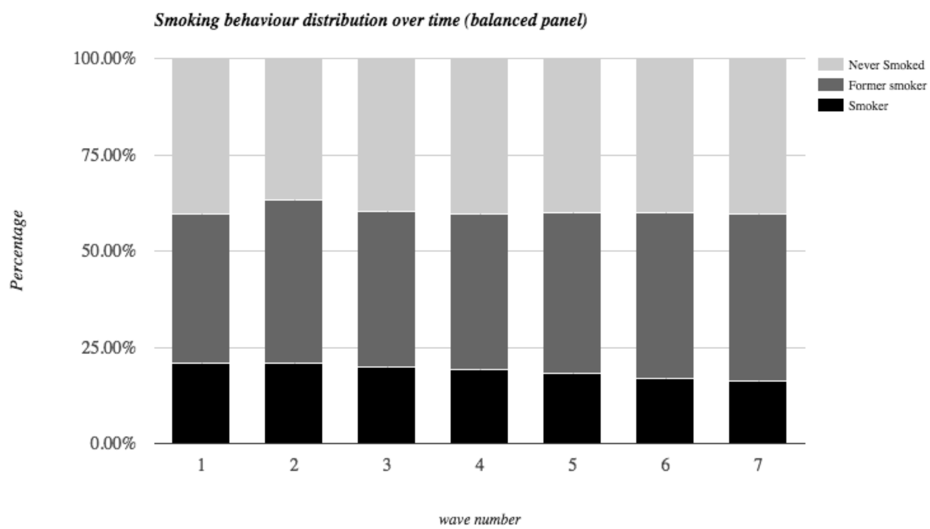
| | Number of Observations | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Age | 17,458 | 51.96368 | 14.85471 | 16 | 93 |
| Number of children | 17,458 | 0.7888074 | 1.102844 | 0 | 7 |
| Gross Personal Monthly Income (Euros) | 17,458 | 1,608.054 | 6,727.37 | -13 | 239,662 |
| Amount cigarettes smoked per day | 17,458[1] | 4.091362 | 8.110402 | 0 | 200 |
| Number of visits to the GP, last 12 months | 16,765[2] | 2.225649 | 3.627089 | 0 | 176 |

[1] Individuals who identified themselves as non-smokers did not answer this question, so their missing values were replaced by 0.
[2] Number of visits to the GP were not reported in the survey in 693 instances. As this is a categorical variable, it is not possible to replace missing values with a default category.

In general, the sample is representative of the Dutch population except for the mean age (39 in 2009), which could be due to reasons previously discussed for not participating in the survey or dropping out after a certain number of waves (CBS Statistics Netherlands, 2013). Overall the average number of cigarettes smoked per day is roughly 4, and among smokers this number is about 12; note that the amount of cigarettes smoked per day also encompasses former smoking habits due to the posing of the question (see *Appendix A, Table 1*). The individual who smokes the most cigarettes per day smokes 200. Descriptive statistics for each smoking behaviour group (smokers, former smokers, non-smokers) as reported in wave 1 are presented in *Appendix C* to give insight into differences between these initial groups.

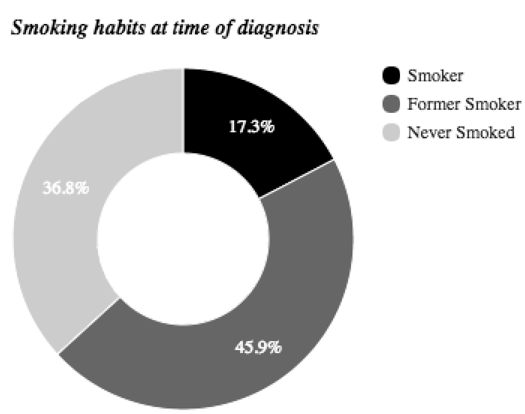
In the balanced panel, the number of non-smokers remained, on average, constant over time around 40% of the sample. However, the number of individuals identifying themselves as current smokers dropped by 23% between 2007 and 2013. It is supposed that many of the individuals who were no longer current smokers became former smokers, as this category's participation rose by 12% throughout the period. A graph showing this distribution over time can be seen below (*Graph 2*), and specific numbers of respondents and the changes over time can be seen in *Appendix B, Table 1*. Over the seven periods, the average percentage of smokers is 18.82%, slightly lower than the national percentage of daily smokers (WHO, 2013).

*Graph 2*: Smoking behaviour distribution over time (balanced panel)




Throughout the seven waves, 1,109 individuals reported having been diagnosed with one of the diseases considered (*Appendix B, Table 2*). More people became sick in waves 2, 3, and 4 compared to the later waves (70% of diagnoses occurred in these periods). Of those diagnosed, 63.2% had smoked at one time, and 17.3% were active smokers when diagnosed (see graph below, and *Appendix B, Table 3*).

*Graph 3*: Smoking habits at time of diagnosis (balanced panel)



The balanced panel is comprised of 2,494 individuals who participate consistently in the survey in each of the seven waves. While this total of 17,458 total observations can be seen in the majority of the variables, individuals did, in some cases, opt to not answer the questions concerning the frequency of drinks and of visits to the GP (18 and 693 individuals, respectively). This discrepancy can be seen in the notes to *Table 1* and *Table 2* above. Additionally, there were 18 total instances where individuals answered the survey questions concerning smoking habits in such a way that they supposedly belong to none of the smoking behaviour groups (see *Appendix B, Table 1*). Despite these discrepancies, there remain 2,494 unique individuals participating in the panel for the seven consecutive years. The panel remains balanced (though less strongly balanced), but the interpretation of the outcomes of the analyses based on this imperfect data will be less strong and able to be externally applied due to these errors and missing values in the responses.

The variables have been presented in detail, and the methods used to model the effect of a diagnosis on smoking behaviour will next be developed and explained.

*Methods used in analysis*

The research will be conducted using different types of regressions which each essentially act as demand functions for cigarettes depending on certain factors to be specified. All models will be based on the use of fixed-effects as the effects of interest vary over time (as opposed to between individuals). The choice of fixed versus random effects models will be tested for empirically, but the assumption is made for the purpose of specifying the models used.

To examine the effect of a health shock on the rate of participation of smoking, a logit model with fixed-effects will be used since the dependent variable is binary and the model requires fixed-effects. The model takes the following form:

$$S(t) = \alpha_0 + \alpha_1 D(t-1) + \alpha_2 S(t-1) + \alpha_3 D(t-1) * S(t-1) + \alpha_4 Z(t) + \alpha_5 T_i \qquad \textit{(eq.1)}$$

*S* and *D* are binary variables representing whether or not a person is a smoker in a given period and whether or not a person received a diagnosis in the given period, respectively. The period is represented by *t*. *Z* encompasses all of the control variables and $T_i$ is a time indicator (each wave is thus represented by a control variable, and i goes from 1 to 7). This model enables the testing of the joint effect of being a smoker and being diagnosed with a disease in a given period on being a smoker in the subsequent period. The main effects of being diagnosed (while not being a current smoker) and being a smoker (with no diagnosis) can also be estimated using this model and can give additional insight into the overall effects of the different factors involved in this analysis. The diagnosis and smoker dummies are first-lags in order to ensure that these events came before the outcome variable in time, along with allowing an individual sufficient time to react to the health update before observing a change in behaviour.

To determine the effect of a diagnosis on the level of consumption of cigarettes for smokers, the method of OLS regression with fixed-effects will be used. This method will estimate the unknown parameters in the model that regresses the consumption of the addictive good on the existence of a health shock and the control variables. The amount of cigarettes an individual reported smoking in a given period (*Q*) will serve as the dependant variable in this model. Along with the control variables included in *equation 1*, a variable representing the first difference between the current period's and the previous period's amount of cigarettes consumed is included ($\Delta Q$ = Q(t) - Q(t-1)). This variable is included as it takes into account the dimension of addiction of reinforcement: if an individual increases their consumption by a larger amount in the past, their current demand or need for the product will be larger. The model developed to test for the effect of a diagnosis on conditional demand for cigarettes is therefore the following:

$$Q(t) = \alpha_0 + \alpha_1 D(t-1) + \alpha_2 S(t-1) + \alpha_3 D(t-1) * S(t-1) + \alpha_4 \Delta Q + \alpha_5 Z(t) + \alpha_6 T_i \qquad \textit{(eq.2)}$$

All variables in *equation 2* are as previously defined.

Models with and without the interaction effect of a diagnosis and being a smoker will be estimated as to compare the joint effect of the two indicators with their separate main effects and to draw conclusions regarding the different groups of smoking behaviour. The outcomes of these regressions along with a presentation of actual changes in smoking behaviour from the balanced sample will next be presented, and they will give an outline of the existing relationship between health shocks in the form of diagnosis and smoking behaviour in this sample.

## RESULTS

A Hausman test was performed for each model to determine whether random or fixed effects were to be used in the analyses. For the OLS model, all control variables were included along with the explanatory variables indicating diagnosis and being a smoker, and the continuous variable indicating the amount of cigarettes smoked was used as the dependent variable to test for which model was most appropriate to use. For the logit model, the same explanatory variables were included with the binary variable indicating whether an individual is a smoker used as the dependent variable. The Hausman test results were that the fixed-effects model should be used for both the OLS and logit models[3]. Therefore, a fixed-effects model with a variance robust to heteroskedasticity was used to model the relationship between a diagnosis and the amount of cigarettes smoked[4] and a conditional fixed-effects logit model was used to model the relationship between a diagnosis and the rate of participation of smoking.

For the logit fixed-effects models, odds ratio output was used in order to facilitate interpretation. Since the slope of the relationship generated by a logit model is not constant (as it is with linear models where coefficients can be directly interpreted), odds ratios were used in all logit models to be able to more easily draw conclusions based on the results.

---

[3] p=0.0000 for both the OLS model and the logit model, so the null hypothesis that the difference in coefficients is not systematic can be rejected at the 95% confidence level and fixed-effects should be used.
[4] The Woolridge test for autocorrelation in panel data resulted in a Prob > F = 0.1647, meaning that no autocorrelation is present.

*Rate of Participation*

First, the outcomes of the conditional logit models with fixed-effects estimating the effect of a diagnosis on an increased likelihood of being a smoker are presented.

The number of observations in this model is of importance. Out of the 17,458 observations in the balanced sample, this model removes 13,042 which leaves 1,894 observations in the model (see *Table 3* below, and *Appendix E, Outputs 1 and 2*). These observations are removed when the dependent variable (being a smoker) does not change (always equals 0 or always equals 1) within an individual, meaning when an individual's smoking behaviour does not change throughout the 7 waves of the panel study. When this is the case, the coefficient of that individual's fixed-effect is infinite in magnitude so becomes a perfect predictor of the outcome. The conditional fixed-effects logit model thus includes observations of individuals who have changed their smoking habits at any point throughout the survey period.

In the model including only the lagged indicators of diagnosis and smoking, the model is, overall, significant (Prob>chi2 = 0.0000). Being a smoker in a given period raises the odds of being a smoker in the following period versus not being a smoker by about 1.87 times, given that all other variables are held constant. This result is highly statistically significant at the 1% level. However, receiving a diagnosis in a given period does not significantly alter the odds of being a smoker in the following period, and the effect size is relatively small (a rise in odds of 1.1 times).

When the interaction between a diagnosis and being a smoker is included, the effect of a diagnosis on the rate of participation changes depending on if an individual is a smoker or not. For individuals receiving a diagnosis and being a smoker in a given period, the odds of being a smoker in the following period versus not being a smoker are the combined odds of the main effect of receiving a diagnosis in a given period and of the interaction between being a smoker and being diagnosed. Adding together the natural logarithms of the odds ratios of the two coefficients representing the mentioned variables below then again transforming it into the odds ratio leads to odds about 0.79 times lower of being a smoker in the following period versus not[5] (these coefficients can be seen in *Table 3, equation (2)* on page 31). Therefore, smokers who receive a diagnosis in a given period have only slightly lowered odds of being a smoker in the following period. This effect is statistically significant at a level between 13.2% and 10.1% based on the significance of the two respective coefficients. Considering the odds ratio of the interaction term alone, individuals who smoke and are diagnosed have lowered odds of reacting to a diagnosis than individuals who do not smoke ($\beta = 0.48$, approximately), meaning that smokers will change their behaviour less strongly than non-smokers in response to a diagnosis.

A slightly stronger effect than in the model without the interaction, individuals who smoke in a given period have increased odds of 1.97 times of being a smoker versus not in the following period (p-value = 0.000), all else equal. For individuals who were not smoking (both former smokers and non-smokers) but received a diagnosis in a given period, the odds of being a smoker versus not in the following period are 1.64 times higher, though this effect is only statistically significant at the 13.2% level.

The models with and without the inclusion of the joint effect of a diagnosis and smoking (outcomes (2) and (1) in *Table 3* below, respectively) have very similar outcomes regarding the control variables. Age, number of kids in the household, civil status, and gross personal monthly income have very small negative effects on the odds of being a smoker versus not in a given period, all else equal, though none of these effects is statistically significant. In both models, one additional visit to the GP in the past twelve months in a given year (period) decreases the odds of being a smoker in the same period versus not by 0.86 times, significant at the 1% level in both models. Education is the only control variable with a positive (albeit small) effect on the likelihood of being a smoker; an increase in the level of education by one (ex. from MBO to HBO) leads to odds of being a smoker versus not in a given period being 1.05 times higher, though these effects are not statistically significant. Lastly, an increase in the frequency of drinks

---

[5] ln(1.640273) + ln(0.4802313) = -0.23862472474; exp(-0.23862472474) = 0.78771043515.

over the past twelve months (ex. from “five or six days per week” to “almost every day”) leads to decreased odds of being a smoker versus not of 0.83 times, significant at a level of 1%.

Table 3: Logit model with fixed-effects outcome, without (1) and with (2) interaction term

***Conditional Logit models with Fixed-Effects (Odds Ratio output)***

| ***Smoker (t): dependent variable*** | **(1)** | **(2)** |
|---|---|---|
| ***Number of Observations*** | 1894 | 1894 |
| ***Groups*** | 316 | 316 |
| ***Diagnosis (t-1)*** | 1.102551<br>(0.42) | 1.640273<br>(1.51) |
| ***Smoker (t-1)*** | 1.872301<br>(5.13)*** | 1.968004<br>(5.36)*** |
| ***Diagnosis (t-1) * Smoker (t-1)*** | | 0.4802313<br>(-1.64)* |
| ***GP visits, last 12 months*** | 0.8632646<br>(-4.51)*** | 0.8630948<br>(-4.50)*** |
| ***Age*** | 0.9707168<br>(-0.35) | 0.9689168<br>(-0.38) |
| ***Number of kids in household*** | 0.9848418<br>(-0.10) | 0.9822932<br>(-0.11) |
| ***Civil status*** | 0.9759639<br>(-0.16) | 0.9702157<br>(-0.20) |
| ***Gross personal monthly income (euros)*** | 0.9999433<br>(-0.95) | 0.9999473<br>(-0.87) |
| ***Education*** | 1.050201<br>(0.44) | 1.045239<br>(-0.40) |
| ***Frequency of drinks, last 12 months*** | 0.8260807<br>(-2.81)*** | 0.8280565<br>(-2.76)*** |
| ***Prob>chi2*** | 0.0000 | 0.0000 |

*Notes: The first (1) Regression does not include the joint effect of being a smoker and receiving a diagnosis in period (t-1). z-values are in parentheses. The symbols (***), (**), (*) Indicate a significance at the 1%, 5% and 10% levels respectively.*

In sum, receiving a diagnosis and being a smoker in a given period slightly lowers the odds of being a smoker in the following period, all else equal. Being a smoker without receiving a diagnosis in a given period strongly increases the odds of also being a smoker in the following period, all else equal. Receiving a diagnosis while not being a smoker in a given period increases the odds of being a smoker in the following period given all other variables are held constant, though this effect is not very significant statistically. It can be said that only about 1 out of 5 smokers who receive a diagnosis would change their rate of participation to smoking, and the effect of a diagnosis on individuals who are not smokers is not statistically conclusive.

### *Conditional Demand*

The effects of a diagnosis on the amount of cigarettes smoked by individuals will be presented next. In these OLS models with fixed-effects, 2,522 observations out of the 17,458 in the balanced panel are not included (see *Table 4* on page 33 and *Appendix E, Outputs 3 and 4*). It is postulated that these observations were removed due to logistic reasons or due to collinearity, meaning that a small number of missing values were present or one of the explanatory variables was constant within individuals over the 7 waves of the panel. Upon examination of the data, indeed 18 people failed to report their drinking frequency and 693 people failed to report their visits to the GP (see *Table 1* and *Table 2* on pages 23 and 24). Since these explanatory variables are of significance in the various models and are not able to be modified by making these people belong to a default category (as the reason for not answering is unknown), the variables were left as is. The models were also run without these control variables, but the number of observations did not rise significantly; it is therefore postulated that the observations that are not included are due to collinearity in some of the explanatory variables.

The models without and with the interaction of being diagnosed and being a smoker in given period (models (1) and (2) in *Table 4* on page 33, respectively) are very similar in their outcomes. In both models, receiving a diagnosis (irrespective of smoking behaviour) alone has no acknowledgable effect on the amount of cigarettes smoked; the effect size is close to 0 in both models and is not statistically significant. Being a smoker in a given period leads to about 4.5 additional cigarettes smoked in the following period (p-value = 0.000) in model (1) and to about 4.6 in model (2) (p-value = 0.000). The joint effect of being a smoker and receiving a diagnosis in a given period leads to a decrease by 0.5 in the amount of cigarettes smoked in the following period which is, similar to the logit model, a combination of the main effect of receiving a diagnosis and the interaction effect of being a smoker and receiving a diagnosis[6]. The interaction effect is significant at the 5% level. Therefore, the effect of a diagnosis on an individual who is not a smoker in a given period (either a former smoker or a non-smoker) is negligible as previously mentioned; receiving a diagnosis in a given period while not being a smoker leads to an increase of about 0.01 cigarettes smoked in the following period, an effect that is not statistically significant.

The effects of the control variables on the amount of cigarettes smoked are very different from those on the rate of participation, but the effects are very similar between models (1) and (2). The effects of the annual number of visits to the GP, gross personal monthly income, and the frequency of drinking are negligible as their coefficients are extremely close to zero in both models and they are not statistically significant. The number of kids in the household has a small negative but statistically insignificant effect on the amount of cigarettes smoked ($\beta$ is approximately -0.15 in both models). In these OLS models with fixed-effects, age, civil status, and education are the explanatory variables with significant effects on the amount of cigarettes smoked. In both models (1) and (2), one additional year of age leads to about one-fifth additional cigarette smoked ($\beta$ is approximately 0.20), an effect significant at the 1% level. A change in civil status (ex. going from "married" to "separated") leads to a decrease in conditional demand for tobacco by about 0.23 cigarettes, significant at the 5% level. Finally, an increase by one level of education results in around 0.12 additional cigarettes smoked, significant at the 10% level. Even these significant control effects are quite small, but taken together can result in a substantial change in smoking behaviour.

[6] 0.0939785 + -0.6027239 = -0.5087454.

_Table 4_: OLS fixed-effects model outcome, without (1) and with (2) interaction term

| | ***OLS Fixed-effects model*** | |
|---|---|---|
| ***Amount of cigarettes smoked (t): dependent variable*** | **(1)** | **(2)** |
| ***Number of Observations*** | 14,936 | 14,936 |
| ***Groups*** | 2,494 | 2,494 |
| ***Diagnosis (t-1)*** | -0.0140264 | 0.0939785 |
| | (-0.13) | (0.80) |
| ***Smoker (t-1)*** | 4.541986 | 4.576646 |
| | (8.54)*** | (8.58)*** |
| ***Diagnosis (t-1) * Smoker (t-1)*** | | -0.6027239 |
| | | (0.026)** |
| ***Amount cigarettes (t) - Amount cigarettes (t-1)*** | 0.4410859<br>(19.50)*** | 0.4411283<br>(19.49)*** |
| ***GP visits, last 12 months*** | -0.0002931 | 0.0002024 |
| | (-0.03) | (0.02) |
| ***Age*** | 0.1982473 | 0.1976504 |
| | (11.61)*** | (11.56)*** |
| ***Number of kids in household*** | -0.1447951 | -0.1456907 |
| | (-1.29) | (-1.29) |
| ***Civil status*** | -0.2288229 | -0.2301092 |
| | (-2.24)** | (-2.26)** |
| ***Gross personal monthly income (euros)*** | -8.05e-07<br>(-0.18) | -9.19e-07<br>(-0.64) |
| ***Education*** | 0.1189391 | 0.1173714 |
| | (1.92)* | (1.90)* |
| ***Frequency of drinks, last 12 months*** | -0.0319251<br>(-0.66) | -0.0310764<br>(-0.64) |
| ***Within R-square*** | 0.6717 | 0.6719 |

*Notes: The first (1) Regression does not include the joint effect of being a smoker and receiving a diagnosis in period (t-1). z-values are in parentheses. The symbols (***), (**), (*) Indicate a significance at the 1%, 5% and 10% levels respectively.*

To summarize, the effect of a diagnosis of an individual who is not a smoker in a given period on the amount of cigarettes he or she consumes in the following period is negligible. Being a smoker who does not receive a diagnosis in a given period leads to about four and a half additional cigarettes smoked in the following year, soliciting the dimensions of addiction of tolerance and reinforcement. Lastly, the effect of a smoker who gets diagnosed in a given period on his or her conditional demand for tobacco is a slight decrease (of about 0.5) in the amount of cigarettes consumed in the following period, suggesting that a health shock in the form of diagnosis on smokers can indeed act as an incentive to slow their consumption of tobacco products.

*Tabulations of actual smoking behaviour after diagnosis*

As an additional insight into the changes in behaviour of individuals who were diagnosed during the seven waves of the survey, tabulations of each group of (non-)smokers were conducted to inform regarding how these individuals' smoking behaviour actually changed one period after their diagnosis. This short analysis also allows for the isolated examination of the behaviour of former smokers.

*Table 5* below shows the smoking behaviour of diagnosed individuals at the time of their diagnosis and in the following period. All three groups tend to largely maintain their behaviour in the period after a diagnosis. *Graph 1 in Appendix D* illustrates the extent to which individuals retain the same smoking behaviour after a health shock. However, there are some changes in behaviour worth noting. Those who were smokers at the time of their diagnosis, for the most part, remain smokers, though 17 individuals (9.6%) report to have quit. Some former smokers take up their old habit (4.6%), and strangely 15 individuals then report to be non-smokers one period post-diagnosis. This could either be due to respondents not wanting to be associated with smoking, their old habit, or due to forgetfulness of their answers in previous years; regardless, this is an inconsistency that speaks to how the survey responses were collected. Only two individuals took up the habit of smoking the year after they were diagnosed, and 24 (6.3%) reported being former smokers; this answer could mean that individuals picked up the habit and proceeded to drop it within one year or could be another instance of researcher bias.

*Table 5*: Smoking behaviour changes in the period following a diagnosis

| Behaviour during period of diagnosis (t-1) | Behaviour one period after diagnosis (t) | | | |
|---|---|---|---|---|
| | Smoking | Former smoker | Never smoked | |
| Smoking | 158 | 17 | 2 | 177 |
| Former smoker | 21 | 425 | 15 | 461 |
| Never smoked | 2 | 24 | 354 | 380 |
| | | | TOTAL | 1,018[1] |

*Note: This table does not take into account the changes in smoking habits that the 90 individuals diagnosed during the 7th wave of the survey may have undergone due to the lack of data in the following period.*
[1] One individual did not report his or her smoking habit the period after his or her diagnosis.

Overall, smokers are the group of individuals who change their smoking behaviour the most frequently (in terms of percentage) after a health shock in this sample. Since the number of diagnoses is not extremely large in this balanced sample, analyses of the changes in smoking behaviour over longer periods of time further limit the number of diagnoses considered. These analyses were also conducted to observe individuals' behaviour changes from one period before a diagnosis to one and two periods after a diagnosis. The results can be seen in *Appendix D, Table 1* and show the same overall trend that individuals largely continue to maintain their smoking habits over time. Smokers are, again, those who most often change their smoking behaviour: 18.5% and 22.3% of individuals who were smokers one period before their diagnoses reportedly quit smoking one and two periods after their diagnoses, respectively. A small percentage of former smokers take up the habit, and very few non-smokers begin smoking after a diagnosis. It is important to note that these observations do not account for any external factors and are simply an account of individuals' smoking behaviours before and after a diagnosis; they do not aim to imply that the diagnosis caused this behaviour but simply serve as an additional point of information.

These observed changes in behaviour do, to some extent, echo the results obtained by the econometric models. Based on the models, smokers who receive a diagnosis are only slightly likely to change behaviour though this change is statistically significant. The tabulations of behaviour in the panel data reiterate that though some changes do take place, these remain small in terms of both the rate of participation and the conditional demand.

## CONCLUSION

The Netherlands is a country where smokers have a strong pro-tobacco attitude, and smoking cigarettes is a tough habit to kick due to its tolerance, reinforcement, and withdrawal characteristics. Both monetary and implicit prices play a role in the incentives for people to begin, continue, or quit their consumption of tobacco products. A serious health shock was seen as an event able to move individuals to or near an unstable steady state, an element of Becker and Murphy's rational addiction model that helps explain why people tend to binge and quit "cold turkey" with regard to the consumption of addictive goods. This research aimed to add to the existing literature on the topic of how individuals respond to health changes in terms of their behaviours by examining the effect of a diagnosis on smoking behaviour for the three groups of smokers, former smokers, and non-smokers. The rate of participation of smoking, meaning whether or not an individual smokes, was modeled using conditional fixed-effects logit models and the conditional demand for cigarettes, the amount of cigarettes smoked, was modeled using OLS models with fixed-effects. Balanced panel data collected by LISS included seven annual waves of responses that were used in the analysis.

One main finding is that individuals tend to maintain their behaviour over time; this result was found in both the logit and OLS models along with tabulations of the data. Former smokers were more likely to pick up their old habit than non-smokers were to begin smoking based on the behaviour of those diagnosed, but receiving a diagnosis while not being a smoker did not significantly affect either the rate of participation or the conditional demand for cigarettes. Smokers who receive a diagnosis in a given period have slightly lowered odds of being a smoker in the following period holding all other effects constant, meaning that a diagnosis can indeed serve as an incentive to change their smoking behaviour with regard to the rate of participation in some cases. A smoker who gets diagnosed also decreases the amount of cigarettes consumed in the following period by about 0.5 cigarettes, meaning that a diagnosis also decreases the conditional demand for cigarettes for smokers.

Interestingly, both the number of visits to the GP and the frequency of alcoholic drinks significantly decrease the odds of being a smoker in the next period (so affect the rate of participation), but do not have any notable effect on the conditional demand. Age, civil status, and education level significantly affect the amount of cigarettes smoked in the following period. Civil status is difficult to interpret since the categories are not necessarily ranked, but both an increase in age and education level lead to slightly more cigarettes smoked. The effect of age brings to mind the dimension of addiction of tolerance, as one may need to smoke more cigarettes to gain the same level of satisfaction over time.

In the case of current smokers, a correlation is present between diagnosis and the consumption of the addictive good. The panel structure and the construction of the models also ensure that time-order is maintained, meaning that the diagnosis and previous smoking behaviour are ensured to have come before any change in smoking behaviour. While fixed-effects do control for many external factors changing over time, the indicators concerning the explanatory power of the models suggest that other factors may be crucial to understanding the full effect of a diagnosis on the consumption of cigarettes. Therefore, the observed effects are not causal but give an indication that diagnosis can serve as a sort of trigger in individuals' choice to binge or abstain from smoking and do provide insight as to opportunities for positive changes in behaviour.

## LIMITATIONS & IDEAS FOR FUTURE RESEARCH

The limitations of this research concern both its internal and external validity and are comprised of, among others, issues such as the data and the analytical approaches used.

The data being in panel form was beneficial for observing individual changes over time while controlling for external effects, but its high level of attrition led to challenges in its interpretation. The most clear issue with the attrition is that individuals may have dropped out of the panel due to death from a disease, due to serious illness, or due to completely unrelated reasons. Since this distinction was impossible, no assumptions could be made and these individuals were disregarded in the main analyses to keep the results consistent and improve accuracy at the expense of disregarding potentially important factors in the analyses. The internal validity of the LISS questionnaire is not ideal, as it was observed that some answers to questions were inconsistent over time (ex. the same individual reporting to have been a smoker in one year and to have never smoked in the next or not reporting to belonging to any of the smoking behaviour groups). This is surely due to inattention of respondents or lack of supervision by the data collectors, and the main subject of the survey was not smoking behaviour but health in general so little attention may have been paid to the questions that were crucial to this research.

The steps and methods used in the analysis are also sub-ideal, mainly due to a lack of expertise of the researcher in advanced econometric methods and statistical software. Firstly, a number of observations were dropped from both the conditional logit and OLS models. Though reasons for this were speculated based on theory and observation of the data, the actual reason for the incompleteness of the sample is not known. This missing piece of the puzzle hinders the external validity of the results and the soundness of the interpretation of the models. Another issue with the quantitative analysis is that this research was unable to ensure that the onset of a disease was exogenous, meaning that the diagnosis could not be separated from smoking behaviour. Put otherwise, the diagnosis itself could have been due to previous smoking behaviours, which would affect an individual's ability or motivation to change their behaviour after such a diagnosis took place.

In conducting future research on this topic, gaining insight as to *why* individuals smoke or stop smoking could be beneficial and could add to the understanding of changes in behaviour, as was done by Basset et al in their study of bladder cancer's effect on smoking (Basset et al, 2012). Using a completely different method of analysis may also lead to more complete results; analysing the data in wide form (rather than longitudinal form, as was done in this research) would allow for modeling changes in smoking behaviour from a number of periods before diagnosis to a number of periods after the diagnosis along with a closer analysis of the specific behaviour of former smokers. Analysing the effects of individual diseases on different groups of people would also allow for a deeper understanding of the mechanism of unstable steady states, but this was outside the scope of this research.

With a budget and a longer period of time to conduct this research, a survey specific to health changes and smoking behaviour could be conducted. This research would entail observing individuals in a medical setting (just before or after an appointment with a medical professional) and in a casual setting (at home, walking around town) and focusing on observing a large number of individuals who were diagnosed and healthy belonging to all three groups of smokers. The health changes and smoking behaviour of the individuals would be recorded by the researchers as to avoid researcher bias and missing answers to important questions along with inconsistencies. Conducting a survey solely pertinent to the questions addressed by this research might lead to a more complete sample and more strongly correlated results.

## POLICY RECOMMENDATIONS

The outcomes of this research point to health shocks as an opportunity, especially for individuals smoking at the time they were diagnosed, to alter their smoking behaviour in a direction beneficial to their health. Since small effect exist in this sample, these can be taken advantage of and extended by health professionals and policy makers. At an individual level, medical professionals can make use of the increased points of contact that take place around the time of diagnosis to thoroughly and clearly inform their patients about the elevated health risks of smoking and about the opportunity to improve their future health. Since the behavioural changes caused by a health shock already take place without evidence of direct implication of medical professionals, a more activated role could only be beneficial to patients at little to no extra cost; more personalized advice and information could push individuals to their unstable steady state and cause more people to quit "cold turkey". Based on the significant effect of the number of GP visits on whether or not a person remains or becomes a smoker one period after a diagnosis suggests that this approach may be particularly effective in addressing the rate of participation.

At the national level, more specific educational programs linking tobacco addiction to some of the serious diseases addressed in this research could be beneficial to students at all levels of education and especially in younger classrooms. This recommendation is based on the results that the conditional demand for cigarettes rises, though only slightly, with an increased education level and with a higher age.

Anti-tobacco groups and the government can make use of the insights from this research to target specific groups such as current smokers who have recently been diagnosed with serious diseases to develop more effective and efficient campaigns to aid in these individuals fighting their addictions and eventually saving costs in health care and lost productivity for the population as a whole.

## APPENDIX A: Variable description

*Table 1*: Variable descriptions and uses in analysis

| Variable Name | Survey Question | Indicator in equations | Used in Transformation to new variable name: |
|---|---|---|---|
| nomem_encr | Number of the respondent encrypted | | ** used as identifier |
| wave | Year and month of the field work period | *T* | ‘wavenum' |
| ch07a80-ch07a98; ch08b80-ch08b98; ch09c80-ch09c98; ch10d80-ch10d98; ch11e80-ch11e98; ch12f80-ch12f98; ch13g80-ch13g98 | Has a physician told you this last year that you suffer from one of the following diseases / problems? More than one answer possible. Diseases include angina, heart attack, high blood pressure, high cholesterol, stroke, diabetes, chronic lung disease, asthma, arthritis, cancer, ulcer, Parkinson’s, cataract, broken hip, another fracture, Alzheimer’s, benign tumor, or "other". | *D* | ‘sick' and 'diagnosis' |
| ch07a125-ch13g125 | Have you ever smoked? | *S; N* | ‘never_smoke' and ‘smoker’ |
| ch07a126-ch13g126 | Do you smoke now? | *S; F* | ‘smoker' and 'former_smoker' |
| ch07a130-ch13g130 | How many cigarettes (including rolling tobacco) [did/do] you smoke per day? | *Q* | ‘amount_cigarettes' |
| ch07a133-ch13g133 | Now think of all the sorts of drink that exist. How often did you have a drink containing alcohol over the last 12 months? | *Z* | ‘freq_drink_last12months' |
| ch07a206-ch13g206 | How often did you use the following health services over the past 12 months? Family Physician: | *Z* | ‘gpvisit_last12months' |
| leeftijd | Age of the household member | *Z* | ‘age' |
| aantalki | Number of living-at-home children in the household, children of the household head or his/her partner | *Z* | ‘num_kids' |
| burgstat | Civil status | *Z* | ‘civilstatus' |
| brutoink | Personal gross monthly income in Euros | *Z* | ‘gross_income' |
| oplmet | Highest level of education with diploma | *Z* | ‘education' |

**APPENDIX B: Descriptive statistics, health-related variables (balanced panel)**

*Table 1*: Smoking habits over time

| Wave | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Total |
|---|---|---|---|---|---|---|---|---|
| Smoker = 1 | 517 | 518 | 495 | 480 | 451 | 420 | 401 | 3,282 |
| Former_smoker = 1 | 966 | 1,059 | 1,011 | 1,005 | 1,042 | 1,073 | 1,088 | 7,244 |
| Never_smoke = 1 | 1,005 | 917 | 988 | 1,004 | 999 | 998 | 1,003 | 6,914 |
| Total | 2,488 | 2,494 | 2,494 | 2,489 | 2,492 | 2,491 | 2,492 | 17,440[1] |

| Percentage per wave | 1 | 2 | 3 | 4 | 5 | 6 | 7 | % change 1-7 |
|---|---|---|---|---|---|---|---|---|
| Smoker = 1 | 20.78% | 20.77% | 19.85% | 19.28% | 18.10% | 16.86% | 16.09% | -23% |
| Former_smoker = 1 | 38.83% | 42.46% | 40.54% | 40.38% | 41.81% | 43.08% | 43.66% | 12% |
| Never_smoke = 1 | 40.39% | 36.77% | 39.62% | 40.34% | 40.09% | 40.06% | 40.25% | 0% |
| Total | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | |

[1]This number indicates the individuals who reported belonging to any of the three smoking-behaviour groups, meaning that individuals reported not belonging to all of the groups in 18 unique instances.

*Table 2*: Diagnosis over time

| Wave | 1 | 2 | 3 | 4 | 5 | 6 | 7 | TOTAL |
|---|---|---|---|---|---|---|---|---|
| Number of diagnoses | -- | 209 | 366 | 191 | 143 | 110 | 90 | 1,109 |
| Percentage of sample in wave *i* | -- | 8.38% | 14.68% | 7.66% | 5.73% | 4.41% | 3.61% | |

* Note: these numbers were obtained following the method described on page 20 of the Data & Methodology portion of the text.

*Table 3*: Smoking habits at time of diagnosis

| Smoking Habits when diagnosed | Number | Percentage of all diagnoses |
|---|---|---|
| Smoker | 192 | 17.31% |
| Former Smoker | 509 | 45.90% |
| Never Smoked | 408 | 36.79% |
| Total | 1,109 | |

**APPENDIX C: Descriptive statistics for each smoking behaviour group at wave 1, based on smoking behaviour during wave 1 (balanced panel; 2,494 individuals in total)**

*Smokers during wave 1*

*Table 1*: Basic population demographics, categorical variables (smokers)

| Variable Name | Levels | Percentage | Frequency | Total Responses |
|---|---|---|---|---|
| Gender | Male | 49.32 | 255 | 517 |
| Civil Status | Married | 57.45 | 297 | 517 |
| | Separated | 0.39 | 2 | |
| | Divorced | 14.51 | 75 | |
| | Widow or Widower | 3.29 | 17 | |
| | Never been married | 24.37 | 126 | |
| Highest level of education with diploma | Primary School | 6.77 | 35 | 517 |
| | VMBO (intermediate secondary school) | 31.91 | 165 | |
| | HAVO/VWO (higher secondary education) | 8.70 | 45 | |
| | MBO(intermediate vocational education) | 23.40 | 121 | |
| | HBO (higher vocational education) | 18.38 | 95 | |
| | WO (university) | 4.64 | 24 | |
| | Not completed any education[1] | 1.55 | 8 | |
| | Not yet started any education[1] | 4.64 | 24 | |
| Frequency of drinks, last 12 months | Almost every day | 26.89 | 139 | 517 |
| | Five or six days per week | 5.42 | 28 | |
| | Three or four days per week | 12.57 | 65 | |
| | Once or twice a week | 23.21 | 120 | |
| | Once or twice a month | 11.41 | 59 | |
| | Once every two months | 4.45 | 23 | |
| | Once or twice a year | 8.70 | 45 | |
| | Not at all over the last 12 months | 7.35 | 38 | |

[1] These answer categories were changed in December 2008 and were no longer offered as options to respondents after that time (starting from wave 3).

*Table 2*: Basic population demographics, continuous variables (smokers)

| | Number of Observations | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Age | 517 | 47.58414 | 13.80776 | 16 | 81 |
| Number of children | 517 | 0.7833656 | 1.106479 | 0 | 7 |
| Gross Personal Monthly Income (Euros) | 517 | 2,472.135 | 16,073.24 | -13 | 233,300 |
| Amount cigarettes smoked per day | 517 | 12.92456 | 8.294957 | 0 | 50 |
| Number of visits to the GP, last 12 months | 364[1] | 3.090659 | 3.164894 | 1 | 25 |

[1] Number of visits to the GP were not reported in the survey by individuals who were smokers in the first period in 153 instances. As this is a categorical variable, it is not possible to replace missing values with a default category.

*Former smokers during wave 1*

*Table 3*: Basic population demographics, categorical variables (former smokers)

| Variable Name | Levels | Percentage | Frequency | Total Responses |
|---|---|---|---|---|
| Gender | Male | 52.59 | 508 | 966 |
| Civil Status | Married | 77.23 | 746 | 966 |
| | Separated | 0.21 | 2 | |
| | Divorced | 7.04 | 68 | |
| | Widow or Widower | 4.45 | 43 | |
| | Never been married | 11.08 | 107 | |
| Highest level of education with diploma | Primary School | 4.66 | 45 | 966 |
| | VMBO (intermediate secondary school) | 27.64 | 267 | |
| | HAVO/VWO (higher secondary education) | 7.87 | 76 | |
| | MBO(intermediate vocational education) | 23.81 | 230 | |
| | HBO (higher vocational education) | 25.26 | 244 | |
| | WO (university) | 6.31 | 61 | |
| | Not completed any education[1] | 3.21 | 31 | |
| | Not yet started any education[1] | 1.24 | 12 | |
| Frequency of drinks, last 12 months | Almost every day | 27.54 | 266 | 966 |
| | Five or six days per week | 7.87 | 76 | |
| | Three or four days per week | 15.42 | 149 | |
| | Once or twice a week | 22.05 | 213 | |
| | Once or twice a month | 9.21 | 89 | |
| | Once every two months | 4.45 | 43 | |
| | Once or twice a year | 7.14 | 69 | |
| | Not at all over the last 12 months | 6.31 | 61 | |

[1] These answer categories were changed in December 2008 and were no longer offered as options to respondents after that time (starting from wave 3).

*Table 4* Basic population demographics, continuous variables (former smokers)

| | Number of Observations | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Age | 966 | 54.37578 | 12.30647 | 16 | 86 |
| Number of children | 966 | 0.6097308 | 0.948589 | 0 | 5 |
| Gross Personal Monthly Income (Euros) | 966 | 2,043.202 | 10,060.82 | -13 | 219,107 |
| Amount cigarettes smoked per day[1] | 966 | 13.61801 | 10.88499 | 0 | 100 |
| Number of visits to the GP, last 12 months | 738[2] | 3.188347 | 2.873833 | 1 | 25 |

[1] For former smokers, this number represents the amount of cigarettes individuals used to smoke (refer to *Appendix A, Table 1*)
[2] Number of visits to the GP were not reported in the survey by former smokers in period 1 in 228 instances. As this is a categorical variable, it is not possible to replace missing values with a default category.

*Non-smokers during wave 1*

*Table 5*: Basic population demographics, categorical variables (non-smokers)

| Variable Name | Levels | Percentage | Frequency | Total Responses |
|---|---|---|---|---|
| Gender | Male | 42.89 | 431 | 1,005 |
| Civil Status | Married | 62.09 | 624 | 1,005 |
| | Separated | 0.50 | 5 | |
| | Divorced | 4.88 | 49 | |
| | Widow or Widower | 3.08 | 31 | |
| | Never been married | 29.45 | 296 | |
| Highest level of education with diploma | Primary School | 2.89 | 29 | 1,005 |
| | VMBO (intermediate secondary school) | 23.28 | 234 | |
| | HAVO/VWO (higher secondary education) | 11.64 | 117 | |
| | MBO(intermediate vocational education) | 22.39 | 225 | |
| | HBO (higher vocational education) | 24.58 | 247 | |
| | WO (university) | 8.56 | 86 | |
| | Not completed any education[1] | 2.99 | 30 | |
| | Not yet started any education[1] | 3.68 | 37 | |
| Frequency of drinks, last 12 months | Almost every day | 8.26 | 83 | 1,005 |
| | Five or six days per week | 4.18 | 42 | |
| | Three or four days per week | 8.86 | 89 | |
| | Once or twice a week | 23.48 | 236 | |
| | Once or twice a month | 17.41 | 175 | |
| | Once every two months | 9.75 | 98 | |
| | Once or twice a year | 12.74 | 128 | |
| | Not at all over the last 12 months | 15.32 | 154 | |

[1] These answer categories were changed in December 2008 and were no longer offered as options to respondents after that time (starting from wave 3).

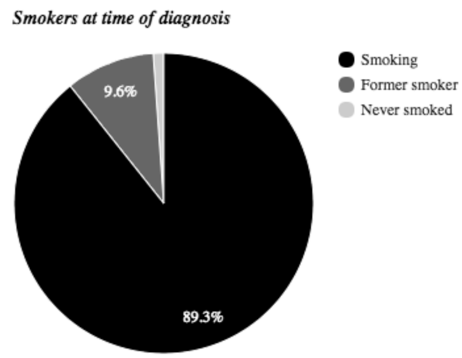
*Table 6*: Basic population demographics, continuous variables (non-smokers)

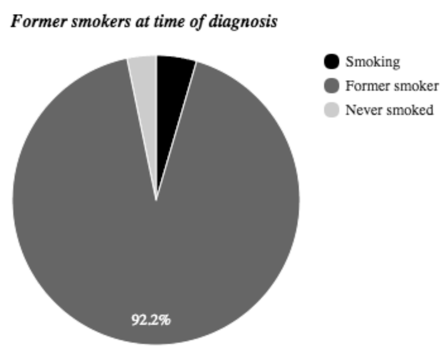
| | Number of Observations | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Age | 1,005 | 44.65373 | 15.64476 | 16 | 87 |
| Number of children | 1,005 | 1.057711 | 1.210287 | 0 | 7 |
| Gross Personal Monthly Income (Euros) | 1,005 | 1,423.709 | 4,400.92 | -13 | 113,179 |
| Amount cigarettes smoked per day | 1,005 | 0 | 0 | 0 | 0 |
| Number of visits to the GP, last 12 months | 727[1] | 3.070151 | 3.028201 | 1 | 40 |

[1] Number of visits to the GP were not reported in the survey by non-smokers in period 1 in 278 instances. As this is a categorical variable, it is not possible to replace missing values with a default category.

**APPENDIX D: Tabulations of actual smoking behaviour (balanced panel)**

<u>*Graph 1*: Smoking behaviour changes in the period following a diagnosis for each group</u>



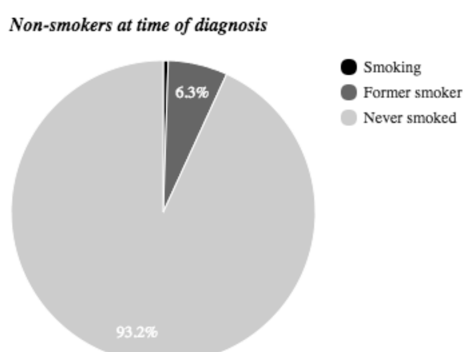



*Table 1*: Smoking behaviour changes from one period before diagnosis to one and two periods after diagnosis, respectively

| Behaviour one period before diagnosis (t-1) | Behaviour one period after diagnosis (t+1) | | | Behaviour two periods after diagnosis (t+2) | | |
|---|---|---|---|---|---|---|
| | Smoking | Former smoker | Never smoked | Smoking | Former smoker | Never smoked |
| Smoking | 157 | 36 | 2 | 134 | 39 | 2 |
| Former smoker | 17 | 417 | 27 | 14 | 372 | 26 |
| Never smoked | 3 | 7 | 350 | 2 | 12 | 306 |
| | | TOTAL | 1,016[1] | | TOTAL | 907[2] |

[1] Responses in wave 7 are not included (90) and 3 people did not answer about smoking habits in either the period before or in the period after diagnosis.
[2] Responses in waves 2 and 7 are not included (209 and 90) and one person did not answer about their smoking habits either one period before or two periods after diagnosis.

## APPENDIX E: STATA output (balanced panel)

*Output 1*: Conditional logit model with fixed-effects, without interaction

```
. xtlogit smoker L1.diagnosis L1.smoker $x i.wavenum, fe or
note: multiple positive outcomes within groups encountered.
note: 2178 groups (13042 obs) dropped because of all positive or
      all negative outcomes.

Iteration 0:   log likelihood = -670.91418
Iteration 1:   log likelihood = -629.70044
Iteration 2:   log likelihood = -629.11243
Iteration 3:   log likelihood = -628.86065
Iteration 4:   log likelihood = -628.84046
Iteration 5:   log likelihood = -628.84046

Conditional fixed-effects logistic regression   Number of obs      =      1894
Group variable: nomem_encr                      Number of groups   =       316

                                                Obs per group: min =         5
                                                               avg =       6.0
                                                               max =         6

                                                LR chi2(14)        =    206.68
Log likelihood  = -628.84046                    Prob > chi2        =    0.0000
```

| smoker | OR | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| diagnosis | | | | | | |
| L1. | 1.102551 | .2569052 | 0.42 | 0.675 | .6983297 | 1.740753 |
| smoker | | | | | | |
| L1. | 1.872301 | .2289104 | 5.13 | 0.000 | 1.473352 | 2.379277 |
| gpvisit_last12months | .8632646 | .0281351 | -4.51 | 0.000 | .8098452 | .9202077 |
| age | .9707168 | .081484 | -0.35 | 0.723 | .8234569 | 1.144311 |
| num_kids | .9848418 | .158127 | -0.10 | 0.924 | .7189469 | 1.349075 |
| civilstatus | .9759639 | .1513116 | -0.16 | 0.875 | .7202199 | 1.322521 |
| gross_income | .9999433 | .0000596 | -0.95 | 0.341 | .9998264 | 1.00006 |
| education | 1.050201 | .1170437 | 0.44 | 0.660 | .8441255 | 1.306586 |
| freq_drink_last12months | .8260807 | .0561817 | -2.81 | 0.005 | .72299 | .9438711 |
| wavenum | | | | | | |
| 3 | .7047045 | .141423 | -1.74 | 0.081 | .4755371 | 1.044311 |
| 4 | .5976429 | .1509257 | -2.04 | 0.042 | .3643205 | .9803922 |
| 5 | .4053664 | .1269679 | -2.88 | 0.004 | .2193999 | .7489606 |
| 6 | .2763131 | .1043406 | -3.41 | 0.001 | .1318175 | .5792019 |
| 7 | .2267341 | .1030951 | -3.26 | 0.001 | .0929991 | .5527833 |

*Output 2*: Conditional logit model with fixed-effects, with interaction

```
. xtlogit smoker L1.diagnosis#L1.smoker $x i.wavenum, fe or
note: multiple positive outcomes within groups encountered.
note: 2178 groups (13042 obs) dropped because of all positive or
      all negative outcomes.

Iteration 0:   log likelihood = -670.00964
Iteration 1:   log likelihood = -628.37009
Iteration 2:   log likelihood = -627.76619
Iteration 3:   log likelihood = -627.55436
Iteration 4:   log likelihood = -627.53687
Iteration 5:   log likelihood = -627.53687

Conditional fixed-effects logistic regression   Number of obs      =      1894
Group variable: nomem_encr                      Number of groups   =       316

                                                Obs per group: min =         5
                                                               avg =       6.0
                                                               max =         6

                                                LR chi2(15)        =    209.29
Log likelihood  = -627.53687                    Prob > chi2        =    0.0000
```

| smoker | OR | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| L.diagnosis#L.smoker | | | | | | |
| 0 1 | 1.968004 | .2486778 | 5.36 | 0.000 | 1.536271 | 2.521065 |
| 1 0 | 1.640273 | .538348 | 1.51 | 0.132 | .8620777 | 3.120943 |
| 1 1 | 1.550217 | .485277 | 1.40 | 0.161 | .8393325 | 2.863196 |
| gpvisit_last12months | .8630948 | .0282556 | -4.50 | 0.000 | .8094541 | .9202901 |
| age | .9689168 | .0815869 | -0.38 | 0.708 | .8215078 | 1.142776 |
| num_kids | .9822932 | .1584455 | -0.11 | 0.912 | .716046 | 1.347539 |
| civilstatus | .9702157 | .1498859 | -0.20 | 0.845 | .7167515 | 1.313312 |
| gross_income | .9999473 | .0000607 | -0.87 | 0.385 | .9998282 | 1.000066 |
| education | 1.045239 | .1169151 | 0.40 | 0.692 | .8394683 | 1.301447 |
| freq_drink_last12months | .8280565 | .056639 | -2.76 | 0.006 | .7241656 | .946852 |
| wavenum | | | | | | |
| 3 | .7127652 | .1436854 | -1.68 | 0.093 | .4801246 | 1.05813 |
| 4 | .5970193 | .1509608 | -2.04 | 0.041 | .3637104 | .9799886 |
| 5 | .4040375 | .1269106 | -2.89 | 0.004 | .2183003 | .7478062 |
| 6 | .2760704 | .1045658 | -3.40 | 0.001 | .1314057 | .5799963 |
| 7 | .2257398 | .1029637 | -3.26 | 0.001 | .0923338 | .5518936 |

*Output 3*: OLS model with fixed-effects, without interaction

```
. xtreg amount_cigarettes L1.diagnosis L1.smoker D.amount_cigarettes $x i.wavenum, fe r

Fixed-effects (within) regression               Number of obs      =     14936
Group variable: nomem_encr                      Number of groups   =      2494

R-sq:  within  = 0.6717                         Obs per group: min =         4
       between = 0.1402                                        avg =       6.0
       overall = 0.3486                                        max =         6

                                                F(15,2493)         =   1898.96
corr(u_i, Xb)  = -0.1554                        Prob > F           =    0.0000

                                   (Std. Err. adjusted for 2494 clusters in nomem_encr)
```

| amount_cigarettes | Coef. | Robust Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| diagnosis | | | | | | |
| L1. | -.0140264 | .1072455 | -0.13 | 0.896 | -.2243259 | .1962731 |
| smoker | | | | | | |
| L1. | 4.541986 | .5319949 | 8.54 | 0.000 | 3.498788 | 5.585183 |
| amount_cigarettes | | | | | | |
| D1. | .4410859 | .0226217 | 19.50 | 0.000 | .3967267 | .4854451 |
| gpvisit_last12months | -.0002931 | .0097748 | -0.03 | 0.976 | -.0194607 | .0188745 |
| age | .1982473 | .0170763 | 11.61 | 0.000 | .1647621 | .2317324 |
| num_kids | -.1447951 | .1125622 | -1.29 | 0.198 | -.3655202 | .07593 |
| civilstatus | -.2288229 | .1020316 | -2.24 | 0.025 | -.4288983 | -.0287474 |
| gross_income | -8.05e-07 | 4.45e-06 | -0.18 | 0.857 | -9.53e-06 | 7.92e-06 |
| education | .1189391 | .0618204 | 1.92 | 0.054 | -.0022855 | .2401637 |
| freq_drink_last12months | -.0319251 | .0487136 | -0.66 | 0.512 | -.1274484 | .0635982 |
| wavenum | | | | | | |
| 3 | .406279 | .241512 | 1.68 | 0.093 | -.0673058 | .8798637 |
| 4 | -.4004104 | .0593042 | -6.75 | 0.000 | -.516701 | -.2841197 |
| 5 | -2.917024 | .1631589 | -17.88 | 0.000 | -3.236964 | -2.597083 |
| 6 | -3.220929 | .1756045 | -18.34 | 0.000 | -3.565274 | -2.876583 |
| 7 | -3.532544 | .1602268 | -22.05 | 0.000 | -3.846736 | -3.218353 |
| _cons | -5.521183 | .9399158 | -5.87 | 0.000 | -7.364279 | -3.678087 |
| sigma_u | 5.4540375 | | | | | |
| sigma_e | 3.082469 | | | | | |
| rho | .75790916 | (fraction of variance due to u_i) | | | | |

*Output 4*: OLS model with fixed-effects, with interaction

```
. xtreg amount_cigarettes L1.diagnosis##L1.smoker D.amount_cigarettes $x i.wavenum, fe r

Fixed-effects (within) regression               Number of obs      =     14936
Group variable: nomem_encr                      Number of groups   =      2494

R-sq:  within  = 0.6719                         Obs per group: min =         4
       between = 0.1406                                        avg =       6.0
       overall = 0.3491                                        max =         6

                                                F(16,2493)         =   1779.06
corr(u_i, Xb)  = -0.1541                        Prob > F           =    0.0000

                           (Std. Err. adjusted for 2494 clusters in nomem_encr)
------------------------------------------------------------------------------
                        |               Robust
      amount_cigarettes |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------------+-----------------------------------------------------
            L.diagnosis |
                      1 |   .0939785   .1180357     0.80   0.426    -.1374795    .3254365
                        |
               L.smoker |
                      1 |   4.576646   .5336667     8.58   0.000      3.53017    5.623121
                        |
     L.diagnosis#L.smoker |
                    1 1 |  -.6027239   .2713658    -2.22   0.026    -1.134849   -.0705984
                        |
      amount_cigarettes |
                    D1. |   .4411283   .0226295    19.49   0.000     .3967538    .4855028
                        |
    gpvisit_last12months |   .0002024    .009897     0.02   0.984    -.0192049    .0196096
                    age |   .1976504   .0170938    11.56   0.000     .1641309      .23117
               num_kids |  -.1456907   .1125352    -1.29   0.196    -.3663627    .0749813
            civilstatus |  -.2301092   .1020197    -2.26   0.024    -.4301612   -.0300571
           gross_income |  -9.19e-07   4.59e-06    -0.20   0.841    -9.92e-06    8.08e-06
              education |   .1173714   .0618978     1.90   0.058     -.004005    .2387478
  freq_drink_last12months |  -.0310764   .0487722    -0.64   0.524    -.1267145    .0645617
                        |
                wavenum |
                      3 |   .4063718     .24163     1.68   0.093    -.0674443     .880188
                      4 |   -.398223   .0593134    -6.71   0.000    -.5145317   -.2819144
                      5 |  -2.914889   .1632414   -17.86   0.000    -3.234992   -2.594786
                      6 |  -3.218343   .1757255   -18.31   0.000    -3.562926    -2.87376
                      7 |  -3.529725   .1603622   -22.01   0.000    -3.844182   -3.215269
                        |
                  _cons |  -5.493339   .9408566    -5.84   0.000    -7.338279   -3.648398
------------------------+-----------------------------------------------------
                sigma_u |  5.4504615
                sigma_e |  3.0820184
                    rho |  .75772208   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

## APPENDIX F: Descriptive statistics (unbalanced panel)

*Table 1a*: Basic population demographics (categorical variables)

| Variable Name | Levels | Percentage | Frequency | Total Responses |
|---|---|---|---|---|
| Gender | Male | 46.26 | 18,836 | 40,772 |
| Civil Status | Married | 58.36 | 23,766 | 40,772 |
| | Separated | 0.36 | 147 | |
| | Divorced | 8.38 | 3,413 | |
| | Widow or Widower | 4.67 | 1,900 | |
| | Never been married | 28.23 | 11,496 | |
| Highest level of education with diploma | Primary School | 5.5 | 2,240 | 40,772 |
| | VMBO (intermediate secondary school) | 25.73 | 10,476 | |
| | HAVO/VWO (higher secondary education) | 10.83 | 4,411 | |
| | MBO(intermediate vocational education) | 22.81 | 9,287 | |
| | HBO (higher vocational education) | 22.22 | 9,047 | |
| | WO (university) | 8.07 | 3,287 | |
| | Other | 3.1 | 1,264 | |
| | Not completed any education[1] | 1.55 | 631 | |
| | Not yet started any education[1] | 0.19 | 79 | |
| Frequency of drinks, last 12 months | Almost every day | 16.14 | 6,546 | 40,559[2] |
| | Five or six days per week | 5.76 | 2,338 | |
| | Three or four days per week | 11.71 | 4,748 | |
| | Once or twice a week | 24.75 | 10,040 | |
| | Once or twice a month | 14.25 | 5,778 | |
| | Once every two months | 7.25 | 2,939 | |
| | Once or twice a year | 8.84 | 3,584 | |
| | Not at all over the last 12 months | 11.31 | 4,586 | |

[1]These answer categories were changed in December 2008 and were no longer offered as options to respondents after that time (starting from wave 3).

[2] Drinking habits were not indicated in the survey in 213 instances. As this is a categorical variable, it is not possible to replace missing values with a default category.

*Table 1b:* Basic population demographics (continuous variables)

| | Number of Observations | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Age | 40,722 | 48.32449 | 17.02313 | 16 | 97 |
| Number of children | 40,722 | 0.8876283 | 1.157559 | 0 | 7 |
| Gross Personal Monthly Income (Euros) | 40,722 | 1,480.227 | 6,539.45 | -13 | 552,000 |
| Amount cigarettes smoked per day | 40,722 | 4.476941 | 8.280691 | 0 | 250 |
| Number of visits to the GP, last 12 months | 38,711[1] | 2.325411 | 4.000109 | 0 | 248 |

[1] The number of visits to the GP was not indicated in the survey in 12,061 instances. As this is a categorical variable, it is not possible to replace missing values with a default category.

*Table 2*: Smoking habits of participants over time

| Wave | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Smoker = 1 | 1,545 | 1,317 | 1,296 | 1,197 | 1,015 | 1,087 | 905 | |
| Former_smoker = 1 | 2,353 | 2,329 | 2,264 | 2,081 | 1,907 | 2,199 | 2,123 | |
| Never_smoke = 1 | 2,750 | 2,29 | 2,533 | 2,411 | 2,137 | 2,476 | 2,338 | |
| Total | 6,648 | 5,944 | 6,093 | 5,689 | 5,059 | 5,762 | 5366 | 40,561[1] |

| Percentage per wave | 1 | 2 | 3 | 4 | 5 | 6 | 7 | % change 1-7 |
|---|---|---|---|---|---|---|---|---|
| Smoker = 1 | 23.24% | 22.16% | 21.27% | 21.04% | 20.06% | 18.86% | 16.87% | -27% |
| Former_smoker = 1 | 35.39% | 39.18% | 37.16% | 36.58% | 37.70% | 38.16% | 39.56% | 12% |
| Never_smoke = 1 | 41.37% | 38.66% | 41.57% | 42.38% | 42.24% | 42.97% | 43.57% | 5% |
| Total | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | |

[1]This number indicates the individuals who reported belonging to any of the three smoking-behaviour groups, meaning that individuals reported not belonging to all of the groups in 211 unique instances.

*Table 3*: Diagnosis over time

| Wave | 1 | 2 | 3 | 4 | 5 | 6 | 7 | TOTAL |
|---|---|---|---|---|---|---|---|---|
| Number of diagnoses | -- | 500 | 690 | 317 | 226 | 160 | 90 | 1,983 |
| Percentage of sample in wave *i* | -- | 8.41% | 11.32% | 5.57% | 4.47% | 2.78% | 1.68% | |

* Note: these numbers were obtained following the method described on page 20 of the Data & Methodology portion of the text.

*Table 4*: Smoking habits at time of diagnosis

| Smoking Habits when diagnosed | Number | Percentage of all diagnoses |
|---|---|---|
| Smoker | 383 | 19.31% |
| Former Smoker | 862 | 43.47% |
| Never Smoked | 737 | 37.17% |
| Total | 1982[1] | |

[1]One person who was diagnosed did not report their smoking habits, resulting in a difference of 1 between this number and the total number of diagnoses in *Table 2* (above).

**APPENDIX G: Regression results (unbalanced panel)**

Note: The Hausman test was conducted for both the OLS and logit models to decide whether fixed or random effects were appropriate to use, the result being a p-value of 0.000 in both cases. Therefore, a conditional logit model with fixed-effects and an OLS model with fixed-effects were used.

*Table 1*: Logit model with fixed-effects outcome, without (1) and with (2) interaction term

***Conditional Logit models with Fixed-Effects (Odds Ratio output)***

| *Smoker (t): dependent variable* | (1) | (2) |
|---|---|---|
| *Number of Observations* | 3060 | 3060 |
| *Groups* | 640 | 640 |
| *Diagnosis (t-1)* | 0.9114319<br>(-0.53) | 1.369049<br>(1.21) |
| *Smoker (t-1)* | 1.261922<br>(2.52)** | 1.323411<br>(2.94)*** |
| *Diagnosis (t-1) * Smoker (t-1)* | | 0.4968539<br>(-2.04)** |
| *GP visits, last 12 months* | 0.9698276<br>(-1.75)* | 0.9699727<br>(-1.76)* |
| *Age* | 0.96944<br>(-0.38) | 0.9692067<br>(-0.38) |
| *Number of kids in household* | 0.9635496<br>(-0.31) | 0.9632566<br>(-0.31) |
| *Civil status* | 0.9687183<br>(-0.32) | 0.970802<br>(-0.30) |
| *Gross personal monthly income (euros)* | 0.9999832<br>(-0.44) | 0.9999845<br>(-0.44) |
| *Education* | 1.093321<br>(1.08) | 1.092332<br>(1.06) |
| *Frequency of drinks, last 12 months* | 0.8092723<br>(-4.24)*** | 0.8094449<br>(-4.22)*** |
| *Prob>chi2* | 0.0000 | 0.0000 |

*Notes: The first (1) Regression does not include the joint effect of being a smoker and receiving a diagnosis in period (t-1). z-values are in parentheses. The symbols (***), (**), (*) Indicate a significance at the 1%, 5% and 10% levels respectively.*

*Table 2*: OLS fixed-effects model outcome, without (1) and with (2) interaction term

***OLS Fixed-effects model***

| ***Amount of cigarettes smoked (t): dependent variable*** | **(1)** | **(2)** |
|---|---|---|
| ***Number of Observations*** | 28,652 | 28,652 |
| ***Groups*** | 8,165 | 8,165 |
| ***Diagnosis (t-1)*** | 0.0960304<br>(1.14) | 0.205403<br>(2.21)** |
| ***Smoker (t-1)*** | 4.125557<br>(11.93)*** | 4.158587<br>(11.9)*** |
| ***Diagnosis (t-1) * Smoker (t-1)*** | | -0.5639092<br>(-2.62)*** |
| ***Amount cigarettes (t) - Amount cigarettes (t-1)*** | 0.419032<br>(26.64)*** | 0.4190824<br>(26.64)*** |
| ***GP visits, last 12 months*** | 0.0031393<br>(0.55) | 0.0034166<br>(0.59) |
| ***Age*** | 0.2007219<br>(9.32)*** | 0.2003898<br>(9.24)*** |
| ***Number of kids in household*** | -0.0957242<br>(-1.16) | -0.0961984<br>(-1.17) |
| ***Civil status*** | -0.1053332<br>(-1.52) | -0.1045988<br>(-1.51) |
| ***Gross personal monthly income (euros)*** | 8.89e-08<br>(0.02) | -1.44e-08<br>(-0.00) |
| ***Education*** | 0.1555526<br>(2.51)** | 0.1546212<br>(2.50)** |
| ***Frequency of drinks, last 12 months*** | -0.051578<br>(-1.46) | -0.0510583<br>(-1.45) |
| ***Within R-square*** | 0.6543 | 0.6544 |

*Notes: The first (1) Regression does not include the joint effect of being a smoker and receiving a diagnosis in period (t-1). z-values are in parentheses. The symbols (***), (**), (*) Indicate a significance at the 1%, 5% and 10% levels respectively.*

## APPENDIX H: STATA output (unbalanced panel)

*Output 1*: Conditional logit model with fixed-effects, without interaction

```
. xtlogit smoker L1.diagnosis L1.smoker $x i.wavenum, fe or
note: multiple positive outcomes within groups encountered.
note: 7525 groups (25592 obs) dropped because of all positive or
      all negative outcomes.

Iteration 0:   log likelihood = -1120.4352
Iteration 1:   log likelihood = -1059.1724
Iteration 2:   log likelihood = -1058.8446
Iteration 3:   log likelihood = -1058.8221
Iteration 4:   log likelihood = -1058.8214
Iteration 5:   log likelihood = -1058.8214

Conditional fixed-effects logistic regression   Number of obs      =      3060
Group variable: nomem_encr                      Number of groups   =       640

                                                Obs per group: min =         2
                                                               avg =       4.8
                                                               max =         6

                                                LR chi2(14)        =    219.24
Log likelihood  = -1058.8214                    Prob > chi2        =    0.0000
```

| smoker | OR | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| diagnosis | | | | | | |
| L1. | .9114319 | .1580974 | -0.53 | 0.593 | .6487455 | 1.280484 |
| smoker | | | | | | |
| L1. | 1.261922 | .1165448 | 2.52 | 0.012 | 1.052979 | 1.512325 |
| gpvisit_last12months | .9698276 | .016953 | -1.75 | 0.080 | .937163 | 1.003631 |
| age | .96944 | .0787565 | -0.38 | 0.702 | .826742 | 1.136768 |
| num_kids | .9635496 | .1170949 | -0.31 | 0.760 | .7593329 | 1.222689 |
| civilstatus | .9687183 | .0957925 | -0.32 | 0.748 | .7980421 | 1.175897 |
| gross_income | .9999832 | .0000379 | -0.44 | 0.658 | .999909 | 1.000058 |
| education | 1.093321 | .0905206 | 1.08 | 0.281 | .9295505 | 1.285944 |
| freq_drink_last12months | .8092723 | .0403437 | -4.24 | 0.000 | .7339402 | .8923365 |
| wavenum | | | | | | |
| 3 | .792707 | .1256134 | -1.47 | 0.143 | .5810723 | 1.081422 |
| 4 | .6309178 | .1362053 | -2.13 | 0.033 | .4132489 | .9632387 |
| 5 | .4642485 | .1315393 | -2.71 | 0.007 | .2664236 | .8089625 |
| 6 | .3096418 | .109464 | -3.32 | 0.001 | .1548618 | .61912 |
| 7 | .2357842 | .1016301 | -3.35 | 0.001 | .1013027 | .5487926 |

*Output 2*: Conditional logit model with fixed-effects, with interaction

```
. xtlogit smoker L1.diagnosis##L1.smoker $x i.wavenum, fe or
note: multiple positive outcomes within groups encountered.
note: 7525 groups (25592 obs) dropped because of all positive or
      all negative outcomes.

Iteration 0:   log likelihood =  -1118.462
Iteration 1:   log likelihood = -1057.1274
Iteration 2:   log likelihood = -1056.8065
Iteration 3:   log likelihood = -1056.7869
Iteration 4:   log likelihood = -1056.7863
Iteration 5:   log likelihood = -1056.7863

Conditional fixed-effects logistic regression   Number of obs     =      3060
Group variable: nomem_encr                      Number of groups  =       640

                                                Obs per group: min =        2
                                                               avg =      4.8
                                                               max =        6

                                                LR chi2(15)       =    223.31
Log likelihood  = -1056.7863                    Prob > chi2       =    0.0000

------------------------------------------------------------------------------------
              smoker |        OR   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------------+--------------------------------------------------------------
         L.diagnosis |
                  1  |  1.369049   .3563587     1.21   0.228      .821967    2.280257
                     |
            L.smoker |
                  1  |  1.323411   .1261957     2.94   0.003      1.09781    1.595373
                     |
L.diagnosis#L.smoker |
                1 1  |  .4968539   .1706197    -2.04   0.042     .2534697    .9739384
                     |
  gpvisit_last12months |  .9699727   .0168465    -1.76   0.079     .9375099     1.00356
                 age |  .9692067   .0787572    -0.38   0.700     .8265102     1.13654
            num_kids |  .9632566   .1178062    -0.31   0.760     .7579492    1.224176
         civilstatus |   .970802   .0957518    -0.30   0.764     .8001571    1.177839
        gross_income |  .9999845   .0000356    -0.44   0.663     .9999146    1.000054
           education |  1.092332    .090628     1.06   0.287     .9283949    1.285218
freq_drink_last12months |  .8094449   .0405316    -4.22   0.000     .7337781    .8929143
                     |
             wavenum |
                  3  |  .7973725   .1267352    -1.42   0.154     .5839429     1.08881
                  4  |  .6311945    .136302    -2.13   0.033     .4133827    .9637718
                  5  |  .4600609    .130462    -2.74   0.006     .2638975    .8020387
                  6  |  .3093788    .109421    -3.32   0.001     .1546812    .6187901
                  7  |  .2346257   .1011782    -3.36   0.001     .1007651    .5463126
------------------------------------------------------------------------------------
```

*Output 3*: OLS model with fixed-effects, without interaction

Note: Testing for serial correlation of the errors of the unbalanced data revealed that there was indeed autorrecation of the errors (p-value = 0.000), so the standard errors were clustered on the individual identifier variable. The clustering allows for arbitrary correlation within individuals, which produces errors that are robust to both autocorrelation and heteroskedasticity.

```
. xtreg amount_cigarettes L1.diagnosis L1.smoker D.amount_cigarettes $x i.wavenum, fe cluster(nomem_encr
> )

Fixed-effects (within) regression               Number of obs      =     28652
Group variable: nomem_encr                      Number of groups   =      8165

R-sq:  within  = 0.6543                         Obs per group: min =         1
       between = 0.1488                                        avg =       3.5
       overall = 0.2833                                        max =         6

                                                F(15,8164)         =   1084.82
corr(u_i, Xb)  = -0.2351                        Prob > F           =    0.0000

                             (Std. Err. adjusted for 8165 clusters in nomem_encr)
```

| amount_cigarettes | Coef. | Robust Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| diagnosis | | | | | | |
| L1. | .0960304 | .0845042 | 1.14 | 0.256 | -.0696193 | .2616801 |
| smoker | | | | | | |
| L1. | 4.125557 | .3457444 | 11.93 | 0.000 | 3.44781 | 4.803304 |
| amount_cigarettes | | | | | | |
| D1. | .419032 | .0157268 | 26.64 | 0.000 | .3882035 | .4498605 |
| gpvisit_last12months | .0031393 | .0057121 | 0.55 | 0.583 | -.0080579 | .0143365 |
| age | .2007219 | .0215431 | 9.32 | 0.000 | .1584919 | .2429519 |
| num_kids | -.0957242 | .0822115 | -1.16 | 0.244 | -.2568796 | .0654312 |
| civilstatus | -.1053332 | .0694467 | -1.52 | 0.129 | -.2414663 | .0307999 |
| gross_income | 8.89e-08 | 3.99e-06 | 0.02 | 0.982 | -7.74e-06 | 7.92e-06 |
| education | .1555526 | .0618926 | 2.51 | 0.012 | .0342272 | .2768779 |
| freq_drink_last12months | -.051578 | .0352428 | -1.46 | 0.143 | -.120663 | .0175069 |
| wavenum | | | | | | |
| 3 | .6678151 | .1496925 | 4.46 | 0.000 | .3743797 | .9612506 |
| 4 | -.5175678 | .0625641 | -8.27 | 0.000 | -.6402094 | -.3949262 |
| 5 | -2.722099 | .1221095 | -22.29 | 0.000 | -2.961465 | -2.482734 |
| 6 | -3.109593 | .1401947 | -22.18 | 0.000 | -3.384411 | -2.834776 |
| 7 | -3.392601 | .1442227 | -23.52 | 0.000 | -3.675314 | -3.109887 |
| _cons | -5.498421 | 1.083009 | -5.08 | 0.000 | -7.621394 | -3.375449 |
| sigma_u | 6.1143905 | | | | | |
| sigma_e | 2.9849484 | | | | | |
| rho | .80754326 | (fraction of variance due to u_i) | | | | |

*Output 4*: OLS model with fixed-effects, with interaction

```
. xtreg amount_cigarettes L1.diagnosis##L1.smoker D.amount_cigarettes $x i.wavenum, fe cluster(nomem_enc
> r)

Fixed-effects (within) regression               Number of obs      =     28652
Group variable: nomem_encr                      Number of groups   =      8165

R-sq:  within  = 0.6544                         Obs per group: min =         1
       between = 0.1495                                        avg =       3.5
       overall = 0.2838                                        max =         6

                                                F(16,8164)         =   1015.50
corr(u_i, Xb)  = -0.2342                        Prob > F           =    0.0000

                            (Std. Err. adjusted for 8165 clusters in nomem_encr)
------------------------------------------------------------------------------------
                       |               Robust
     amount_cigarettes |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------------+------------------------------------------------------------
           L.diagnosis |
                    1  |    .205403   .0927879     2.21   0.027     .0235151    .3872909
                       |
              L.smoker |
                    1  |   4.158587   .3469653    11.99   0.000     3.478446    4.838727
                       |
  L.diagnosis#L.smoker |
                  1 1  |  -.5639092   .2152934    -2.62   0.009     -.985939   -.1418794
                       |
     amount_cigarettes |
                   D1. |   .4190824   .0157321    26.64   0.000     .3882434    .4499213
                       |
  gpvisit_last12months |   .0034166   .0057564     0.59   0.553    -.0078675    .0147007
                   age |   .2003898   .0216835     9.24   0.000     .1578846     .242895
              num_kids |  -.0961984     .08218    -1.17   0.242    -.2572921    .0648954
           civilstatus |  -.1045988   .0694724    -1.51   0.132    -.2407824    .0315848
          gross_income |  -1.44e-08   4.10e-06    -0.00   0.997    -8.05e-06    8.02e-06
             education |   .1546212   .0619432     2.50   0.013     .0331968    .2760456
freq_drink_last12months|  -.0510583   .0352757    -1.45   0.148    -.1202075     .018091
                       |
               wavenum |
                    3  |   .6669774   .1497579     4.45   0.000     .3734138     .960541
                    4  |  -.5156033   .0627392    -8.22   0.000     -.638588   -.3926185
                    5  |  -2.722189    .122326   -22.25   0.000    -2.961979   -2.482399
                    6  |  -3.108113   .1405689   -22.11   0.000    -3.383664   -2.832562
                    7  |  -3.391448   .1447545   -23.43   0.000    -3.675203   -3.107692
                       |
                 _cons |  -5.489643   1.089388    -5.04   0.000    -7.625121   -3.354165
-----------------------+------------------------------------------------------------
               sigma_u |  6.1096344
               sigma_e |  2.9845151
                   rho |  .80734643   (fraction of variance due to u_i)
------------------------------------------------------------------------------------
```

## RESOURCES


- American Heart Association (2011). “Why is it so hard to quit?”. American Heart Association. Retrieved May 7, 2015, from <http://www.heart.org/HEARTORG/GettingHealthy/QuitSmoking/QuittingSmoking/Why-is-it-so-hard-to-quit_UCM_324053_Article.jsp>.
- Baltagi, Badi. (2008). *Econometric analysis of panel data*. John Wiley & Sons, 2008. Ch.1.
- Basset, J.C. et al. (2012). Impact of a bladder cancer diagnosis on smoking behavior. *Journal of Clinical Oncology*. Vol. 30, No. 15, 1871-1878.
- Becker, G. S, & Murphy, K. M. (1988). "A theory of rational addiction." *The journal of political economy* (1988): 675-700.
- Becker, G. S., Grossman, M., Murphy, K. M. Rational addiction and the effect of price on consumption. (1991).
- Becker, G.S., Grossman, M., and Murphy, K.M. (1994). "An empirical analysis of cigarette addiction", American Economic Review 84(3):396-418.
- CBS Statistics Netherlands. (2013). “Population; key figures”. Accessed 31 May, 2015 from <http://www.cbs.nl/en-GB/menu/themas/bevolking/cijfers/>
- Cawley, J., & Ruhm, C. (2011). The economics of risky health behaviors. National Bureau of Economic Research.
- CDC- Center for Disease Control and Prevention. (2009). CDC - Fact Sheet - Health Effects of Cigarette Smoking. Retrieved May 7, 2015, from <http://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/effects_cig_smoking/>.
- Chaloupka, F. J. (1991), "Rational addictive behavior and cigarette smoking", Journal of Political Economy 99(4):722-742.
- Chaloupka, F. J., & Warner, K. E. (2000). The economics of smoking. *Handbook of health economics*, *1*, 1539-1627.
- Chaloupka, F. J., Tauras, J. A., and Grossman, M. "The economics of addiction." *Tobacco control in developing countries* (2000): 107-129.
- Chaloupka, F.J., Emery, S., and Liang, L. (2003). "Evolving models of addictive behaviour: from neoclassical to behavioral economics." *Choice, Behavioral Economics and Addiction* (2003): 71-89.
- Darden, M. (2011). A dynamic stochastic model of lifetime smoking behavior. Ph.D Dissertation, The University of North Carolina at Chapel Hill.
- Dotinga, R. (2015). “Many smokers try to quit after cancer diagnosis”. Health Day. Accessed 21 May, 2015 at <http://consumer.healthday.com/cancer-information-5/mis-cancer-news-102/many-smokers-try-to-quit-after-cancer-diagnosis-698579.html?lexp=true&utm_expid=38353063-4.pIV1hUrQR8K_MJ1_OqjLag.1&utm_referrer=https%3A%2F%2Fwww.google.nl%2F>
- Douglas, S. (1998), "The duration of the smoking habit", Economic Inquiry 36(1):49-64
- Elster, J. (2000). Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints. Section I.7.
- Eriksen, M., Mackay, J.,Schluger, N., Gomeshtapeh, F. I., Drope, J. *Tobacco Atlas*. 2015. The American Cancer Society.
- Evans, W.N., and Farrelly, M.C. (1998). "The compensating behavior of smokers: Taxes, tar and nicotine", RAND Journal of Economics 29(3):578-595.
- Fischhoff, B., Bostrom, A., and Quadrel, M.J. (1993). “Risk Perception and Communication,” Annual Review of Public Health 14, 183–203.
- Grossman, M. (1995). "Chapter 10: The economic approach to addictive behavior." *The New Economics of Human Behaviour* (1995): 157-171.
- Houthakker, H.S., and L.D. Taylor (1966), Consumer Demand in the United States, 1929-1970: Analyses and Projections (Harvard University Press, Cambridge, MA).
- ITC Project (June 2010). ITC Netherlands National Report. University of Waterloo, Ontario, Canada; STIVORO (Dutch Expert Centre on Tobacco Control), The Hague, The Netherlands.

- Jacobs, M. (1997). From the first to the last ash: the history, economics & hazards of tobacco. Health literacy.
- Kassel, J.D., Paronis, C.A., and Stroud, L.R. (2003). Smoking, stress, and negative affect: correlation, causation, and context across stages of smoking. *Psychological Bulletin*. Vol. 129, No. 2, 270-304.
- Lewit, E.M., and Coate, D. (1982). The potential for using excise taxes to reduce smoking. *Journal of Health Economics*, 1(2), 121-145.
- Marshall, A. (1920). Principles of Economics, 8th edition. (Macmillan, London). Appendix H.7.
- Medical News Today. (2013). "Stress From 9/11 Linked To Nationwide Resurgence In Smoking Among Americans Who Had Quit". Retrived on 22 May, 2015 < http://www.medicalnewstoday.com/releases/262299.php >.
- Nakajima, R. (2007). "Measuring peer effects on youth smoking behaviour." *The Review of Economic Studies* 74.3 (2007): 897-935.
- Oster, Emily, Ira Shoulson, and E. Ray Dorsey. (2013). "Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease." *American Economic Review*, 103(2): 804-30.
- Oster, E. (2014). "Diabetes and Diet: Behavioral Response and the Value of Health". <http://faculty.chicagobooth.edu/emily.oster/papers/diabetes.pdf>.
- Schelling, T.C. (1978). "Egonomics, or the art of self-management", American Economic Review 68:290-294.
- Sloan, F.A., Smith, V.K. and Taylor, D.H. (2003). The Smoking Puzzle: Information, Risk Perception, and Choice. Harvard University Press.
- Smith, V.K., Taylor, D.H., Sloan, F.A., Johnson, F.R., Desvousges, W.H. (2001). Do smokers respond to health shocks? The Review of Economics and Statistics, 83 (4), 675–687.
- Spinnewyn, F. (1981). Rational habit formation. *European Economic Review*, 15, 91-109.
- Suranovic, S. M., Goldfarb, R., and Leonard, T.C.. "An economic theory of cigarette addiction." *Journal of Health Economics* 18.1 (1999): 1-29.
- The Tobacco Atlas. (2015). "Secondhand Smoke | The Tobacco Atlas." Retrieved 8 May. 2015 <http://www.tobaccoatlas.org/topic/secondhand-smoke/>.
- US Department of Health and Human Services. (1998). "Tobacco use among U.S. racial/ethnic minority groups", A Report of the Surgeon General (US Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; Government Printing Office, Washington, DC).
- Westmaas, J.L. et al. (2015). Does a Recent Cancer Diagnosis Predict Smoking Cessation? An Analysis From a Large Prospective US Cohort. *Journal of Clinical Oncology*.
- WHO. (2013). WHO Report on the Global Tobacco Epidemic, 2013. Retrieved May 7, 2015, from <www.who.int/entity/tobacco/surveillance/policy/country_profile/nld.pdf>.
- Wills, T. A., & Shiffman, S. (1985). Coping and substance use: A conceptual framework. In S. Shiffman & T. A. Wills (Eds.), Coping and substance use (pp. 3–24). New York: Academic Press.
- Young, T. (1983). "The demand for cigarettes: Alternative specifications of Fujii's model", Applied Economics 15:203-211.