Erasmus University Rotterdam
Erasmus School of Economics

Master Thesis [Economics and Business - Health Economics specialisation]

The Consequent Effects of Acquiring Health Information Online on Health Behaviour

Fanny Tallgren
495548

Supervisor: Pilar Garcia Gomez
Second assessor: Carin A. Uyl-de Groot

30th April 2021

Abstract

What impact can a search for health information online have on health behaviour? And what are some individual characteristics associated with an individual searching for health information? This thesis attempts to answer these two questions using cross-sectional data from an Italian population survey for the years 2013 to 2016. The analysis uses an instrumental variable method to estimate the effect of searching for health information online on health care use, weight and diet. The instrument used is a dummy variable equal to 1 if the percentage of households with access to broadband in an Italian region is higher than the median share. Health information online was only found to have a positive and significant impact on health care use. The results also showed that individuals with health problems and a poor self-rated health state as well as those with a high socio-economic status were more likely to search for health information.

Table of Contents

1. Introduction

Wonderful things can be done with the World Wide Web. Pope Francis reportedly described the internet as "a gift from God" (Gander, 2014). It has revolutionised every field in the world, including health care. People have always tried to acquire health and self-care information. For a long time, this was mainly done using resources such as books and telephone advices (Wagner et al., 2001). Nowadays, the internet is increasingly used as a source of information for any concerns with one's health, leading to the emergence of a new type of patient, the so called "e-patient" (Ferguson & Frydman, 2004) and a new type of doctor, "Dr Google". Individuals who excessively use the internet for health information have been coined "cyberchondriacs" (Loos, 2013, Menon et al., 2020, Starcevic et al., 2019). Already in 1997, a national survey found that 41% of internet users in the United States (US) had searched for health information online (Rice, 2006). More recently, a study published in 2020 by Bujnowska et al. found that 76.8% of the survey respondents has used health resources online. One can only imagine what the rates have increased to now, in the midst of a global pandemic.

Is the increased use of online health information good news or bad news? It appears that health information on the internet is a double-edged sword. One can quickly envisage the advantages of access to online health resources, as it enables information to be acquired fast and improves the average health knowledge of an individual. This is beneficial since the health sector suffers from a large information asymmetry between health professionals and the general population (Folland et al., 2017, p. 273).

The disadvantages from easy access to health resources online must be acknowledged. Firstly, the availability of information online is problematic when people choose to use the internet in cases where they should have consulted a health care professional. In other words, e-health can become a substitute to health care. In a US 2012 survey, 35% of the respondents who admitted to have tried to diagnose themselves online had not sought information from a professional (Pew, 2013). A second issue is to do with the reliability and accuracy of information online. Since there is no regulation, countless websites with misleading and erroneous information are accessible. This is especially the case for the topics of nutrition. Even if the provided information is correct, it may be too difficult for the average citizen to fully comprehend (Mitsutake et al., 2016). The health information literature have developed the concept of eHealth literacy, which is the ability to find information and to evaluate its credibility (Iverson, et al., 2008, Mitsutake et al., 2016). In case an individual does not have an adequate level of eHealth literacy, applying the information found online can have damaging impacts on their health (Iverson, et al., 2008). Therefore, the problem is twofold; firstly, there is a lot of inaccurate information online, and secondly, e-patients may not be able to recognise when the information is unreliable (Eysenbach et al., 2002).

Assuming that the internet's flaws could be taken care of, the internet can be considered a powerful tool to promote healthy behaviour. At least in Europe, the majority of the population and especially the younger generations use the internet daily. Ideally, reliable and understandable health information online provided by health authorities would help reduce unhealthy behaviour. This, in turn can potentially decrease health care costs as well as inequality at a relatively low cost. Currently, obesity, poor diets and low exercise levels are significant public health problems (Fruhbeck et al., 2013). They have been found to impact quality of life and health care costs as well as to cause chronic and noncommunicable diseases and premature deaths (Cawley & Ruhm, 2012, Cawley & Meyerhoefer, 2012, Frazão 1999, Fontaine et al. 2003, Specchia et al., 2014, Migliore et al., 2013). On top of this, smoking, obesity, physical inactivity, unhealthy diets and alcohol can have a significant impact on labour market outcomes and GDP (Devaux & Sassi, 2015, Popkin et al., 2006).

Two counterarguments with regards to considering obesity and unhealthy lifestyles a public health problem must be recognised. Firstly, it is possible that the individuals who currently lead healthy lifestyles may later on suffer from costly old age diseases such as Alzheimer's. Therefore, attempts to minimise unhealthy behaviours may not reduce health care costs. However, this can be considered a separate issue. Secondly, some hold the strong conviction that a person's diet and weight is ultimately a private matter and that the government should not intervene. Also, not everyone who is obese is a problem for society, and not everyone who has a "normal" weight is necessarily healthy (Folland et al., 2017, p. 207). However, there is considerable evidence, especially in behavioural economics, that individuals do not behave optimally and can make errors in their decision-making (Laibson, 1997, Loewenstein, 1996 , O'Donoghue & Rabin, 1999, Sunstein & Thaler, 2003).

In order for e-health resources to be used as a tool to promote healthy behaviour, there should be proof that online information has an impact on people's actions (Mitsutake et al., 2016). Unfortunately, economic research has not paid considerable attention to studying the effects of online health information on people's health behaviour. Notable studies by DiNardi et al. (2019), Farajallah et al. (2015), Schmid (2015), Dwyer and Liu (2013), Suziedelyte (2012) and Wagner &

Jimison (2003) will be reviewed later on. However, they have mainly analysed the impact of health information on health care use, but not on behaviour. Consequently, this thesis will try to answer the question whether health information online can influence people's health behaviour.

The health behaviour analysed will be visits to a health care professional and eating a healthy diet (to be defined later). The impact of e-health on weight will also be analysed. Although weight can be considered more of a health outcome than a health behaviour, it is an outcome that can be modified through behaviour (DiNardi et al., 2019). Because of the availability of erroneous information online, the thesis will also attempt to assess whether online information influences patients' health behaviour for the better or the worse. The analysis will also try to identify determinants of health information seeking online. Conclusions will be reached based on an instrumental variable analysis of cross-sectional data for the years 2013- 2016 from an annual survey, *The Aspects of Daily Life*, conducted in Italy.

The reasons for using a dataset from Italy are manifold. Italy is an interesting country for analyses on diet and weight, especially due to the alleged presence of the Mediterranean diet (Cavaliere et al., 2014). According to Italy's 2019 country health profile by the OECD and the European Observatory on Health Systems and Policies (EOHSP), the leading cause of death in Italy were cardiovascular diseases. There is a considerable amount of medical evidence that cardiovascular diseases can be prevented through health behaviour, such as exercising and eating a healthy diet (Gupta & Wood, 2019). More specifically, the OECD & EOHSP report claims that 16 % of deaths in Italy in 2017 were related to eating a poor diet. Additionally, the percentage of overweight and obesity has been increasing, especially amongst children and teenagers (OECD & EOHSP, 2019), although it remains low relative to other countries. For example, according to a European wide Health Behaviour in School-aged Children (HBSC) survey in 2013-2014, 18% of 15-year-olds in Italy were found to be overweight or obese (OECD & EOHSP, 2019). A very recent study in the field of nutrition by Vitale et al. (2021) called for interventions to promote healthier diets in Italy based on the results of their analysis of Italian's food consumption. Therefore, it is beneficial to find out whether access to health information online in Italy can lead to a change in habits from a perspective of population health.

The main research question is the following: "What are the consequences of searching for health information online on an individual's health behaviour?". The research question can be divided further into the following four sub questions:

1. What are the determinants of searching for health information online?

What are the consequences of searching for health information online on

2. An individual's use of health care services?
3. an individual's diet?
4. An individual's weight?

The following are three hypotheses that will be tested:

H1: Searching for health information online increases health care use.

H2: Searching for health information online has a positive impact on a person's diet

H3: Searching for health information online reduces obesity

This thesis is structured as follows. In the first section, a literature review of similar studies is conducted. This is followed by a coverage of economic theory for health information seeking. The third section explains the data set and methodology used to answer the research questions. In the fourth section, the results of the analysis are presented. Consequently the fifth section is a discussion of the results and the limitations of the study as well as a short conclusion. This is followed by a list of references and the appendix.

2. Literature review

How does searching for health information online affect health behaviour? This section will cover the moderate amount of research carried out in the field of economics with regards to the topic.

A 2019 study by DiNardi et al. in the US found that internet use had an impact on peoples' health behaviour. They looked at six outcomes: being overweight (body mass index (BMI) $\geq$ 25), obese (BMI $\geq$ 30), extremely obese (BMI $\geq$ 40), exercise activity in the last 30 days, binge drinking (five or more drinks in one occasion) in the last 30 days, and smoking. Similarly to this thesis, they use the rollout of broadband providers across the US during the 2000s as a source of exogenous variation in Internet access and use. After performing the analysis separately by race and gender, they find that internet access potentially increases the body weight of white women, although not for non-white women or men, and that it increases alcohol consumption but also exercise for both men and women. They argue that increased access to the Internet can lead to increased information, increased social connections, and increased income. In turn, these changes influence the individual's allocation of time between sedentary and dynamic activities, which has an impact on their weight.

Farajallah et al. (2015) tried to estimate whether internet users were looking for information to complement health care use (in order to understand or confirm their doctor's diagnosis) or to substitute it (to replace or contest their doctor's diagnosis). Their study (N=1344) conducted in France consisted of three logit econometric models. They find that individuals searched for health information online to find out about a disease or symptoms of disease. Women, individuals with poor health and individuals who spend considerable time on the Internet were most likely to search for information online.

Schmid (2015) found that a high level of consumer health information reduced physician visits by 0.55 per year. However, he finds no effect on the likelihood of visiting a physician. He estimates a two-part Poisson hurdle model for physician visits based on the 2007 Swiss Health Survey (N=14,393). The data included information about individuals' health status, insurance, health behaviour, health care use, health expertise, and socioeconomic status. However, he did not look at online health information specifically.

Dwyer and Liu (2013) analysed whether consumers' use of health information from non-physician information sources was as a substitute or complement for health services. They found that health information increased the likelihood of visiting a physician, but also the average frequency of visits. They controlled for patients' trust in physicians, and find some evidence of patients with a low trust in physicians replacing health care visits with health information. With respect to emergency room visits, they found that consumers who sook health information made significantly fewer visits.

Suziedelyte (2012) studied whether searching for health information online affected individuals' demand for health care services. Using the US Health Information National Trend Survey (HINTS) data for 2003 to 2007, she found that people who searched for health information online were significantly more likely to use health care in terms of number of visits of health care professionals. She finds the effect to be larger for individuals who search for health information online more frequently. She uses US state-level telecommunication regulations which allegedly affected the supply of Internet services as an instrument for health information online. Therefore, she concludes that the e-health is a complement rather than a substitute for health care. Her findings were robust even after controlling for third factors that could affect both searching for information and health care use. She found that individuals who had an intrinsic interest in health were more likely to search for health information online and actively use health care in an attempt to maximise their health.

An earlier study conducted in 2003 by Wagner & Jimison did not find a significant effect of using a computer for health information on self-reported health care visits. They studied the effect based on survey (N=5909) results using ordinary least squares, instrumental variables, fixed effects, and fixed-effects instrumental variables models. Exposure to a health information intervention (the Healthwise Communities Project) as well as computer ownership and Internet access were used as instrumental variables. However, since the study dates back from 2003, it is highly likely that the internet has a stronger impact on people's health care use nowadays than at the beginning of the century. Much earlier studies have also been conducted. In their 1982 study, Hay & Leahy found a positive effect of information on the likelihood of visiting a physician. Several years later in 1990, Kenkel also found a positive correlation between health information and health care demand.

A similar analysis outside of the field of economics by McCully et al. (2013) used multiple linear regression to look at the associations between Internet use for diet, weight, and physical activity (DPWA) and three health behaviours: fruit intake, vegetable intake, and physical activity. Similar to Suziedelyte (2012), they analysed US HINTS data on internet users (N=4827) from 2007 to 2011. They found that 43% of survey respondents had searched information online on improving their diet, weight or physical activity. They also found that internet users who were younger, more educated, married and who had a higher BMI were more likely to use the e-Health for DWPA. They did not find a correlation between gender and using the Internet for these health behaviours. Individuals who had used the internet for DWPA had a higher vegetable and fruit intake, and did more moderate exercise.

An Italian study in the field of public health by Siquilini et al. (2011) did not find a correlation between e-health use and a change in one's lifestyle and/or in joining preventive programmes (i.e., vaccination, screening). However, they still concluded that the internet could be used as a useful tool to nudge the population's health behaviour. Their aim was to assess differences in e-health use by socio-demographic and health-related variables. They used questionnaire results (N= 3018) conducted in hospital laboratories by physicians from six representative Italian cities and analysed the data using descriptive statistics and logistic regressions. They decided to exclude respondents older than 65 years from the sample as they claimed Internet usage in Italy among elderly people to be low. 57% of the respondents who used the internet reported using it to search for health-related information. They were young, female and affected by chronic diseases.

Multiple interesting and relevant studies outside of the field of economics have concluded that people consult health information resources with the aims of adjusting their lifestyle behaviours and to take better care of themselves (Bujnowska et al., 2020, Hsieh & Lin, 1997, Lee, 2008). The studies mentioned show that information seekers stated intentions of wanting to change their health behaviour. However, evidently this is only what the respondents claim they would do. It is unknown whether they actually changed their habits.

There is a considerable amount of studies in the field of economics with regards to the impact of other types of information on health behaviour. A few studies have analysed the impact of providing nutritional information to consumers on individuals' diets and calories consumed, for example studies by Variyam, 2007, Craig et al., 2020 and Stranieri et al., 2010. Carrera et al. (2020) studied whether a health risk assessment affected the eating behaviour of hospital employees at their canteen. They find that employees diagnosed with high-risk levels of cholesterol mad small, short-term reductions in their spending at the cafeteria. The changes where only modest and temporary among individuals who were unaware of their high cholesterol. A study by Oster (2018) looked at household food purchases among individuals who had been diagnosed of diabetes using data from the Nielsen HomeScan panel. She was able to find data on health outcomes for a subset of the households from a secondary survey. Using a household fixed-effect framework, she finds that a diagnosis leads to an improvement in diet in terms of a reduction in calories purchased from non-whole grains, soda, red meat and whole milk products.

Carrieri et al. (2019) found that the spread of negative information online on the MMR vaccine in Italy led to a significant decrease in child immunisation rates for *all* vaccines. Carrieri & Principe in 2018 also found that WHO's health warning on carcinogenic effects of consuming red meat led to a short-term decrease in the consumption of red meat. In both cases, the effects were stronger for the less educated than for the higher educated.

All of the above mentioned studies differ from the purpose from this thesis in three ways. Firstly, the survey data used in this thesis did not ask the health information seekers whether the search had an impact on their health behaviour. Asking people directly may potentially lead to biased results. Instead, the survey asked separately about the respondent's health information searching and their health behaviour. After that, statistical analysis is done to see if there is a correlation.

Secondly, this thesis will contribute to the existing literature by not only analysing the impact of health information on health care use, but also on other types of health behaviour, such as diet. This is believed to be of use as for example Suziedelyte (2012) concluded in her study that interesting future research topics would be the impact of health information online on risky behaviour and efforts to improve their health. Thirdly, many of the studies did not differentiate between health information *seeking* and health information *scanning* (Lee, 2008). Some papers have simply looked at the effect of being exposed to information, which is potentially difficult to analyse as health information is currently provided on every type of media. The focus of this thesis is on the impact of health information seeking on health behaviour.

3. The theoretical framework

This section will cover the theoretical arguments explaining why people would search for health information online in order to improve their health, and why this could change their health behaviour.

Health as an investment good

According to classical economic theory, humans are utility maximisers. In other words, they make decisions by weighing the utility (or happiness) costs and benefits of different options and choose the one that maximises their utility. For example, in the context of consumer goods, a person chooses to buy a good because it brings them more utility than it costs them and choose the good that brings them most utility. The choices are subject to their income and time constraints.

Applying this argument in the context of health is more tricky as health in itself cannot be consumed. However, people can make decisions and take actions that improve their health. For example, a person may go for a run if they believe it will increase their health, on top of the utility they might get from running itself (Grossman, 1972). When people take these actions, are they maximising utility or health? How does improved health contribute to utility? Maybe it is the satisfaction of having contributed to their health that provides utility. There are multiple reasons why good health, or the knowledge of being in good health, could increase utility. Good health enables people to live longer, to feel and look good and to be more productive (Folland et al., 2017, p. 196). Days spent in poor health decrease utility, whereas being healthy means days can be spent on activities that increase utility, such as working to earn money and purchasing commodities (Grossman, 1972). Being healthy can also lead to a longer life, which implies more days in total that may provide utility (Grossman, 1972).

Grossman's economic model of health capital provides a convincing argument for explaining people's health behaviour (Grossman, 1972, Wagner et al., 2001, Folland et al., 2017, p. 190, Hsieh & Lin, 1997). Good health can be considered an investment / capital good. In this model, every person is born with a certain stock of health, which will depend on factors such as hazard and genetics. The initial stock of health can last for a long time, but it naturally decreases and even depreciates at an increasing rate as a person becomes older. People can try to increase their stock by investing in their health through their actions. Therefore, a production function of health exists as people consume health inputs to produce health. Health inputs can include health care use and

/ or having a healthy lifestyle by eating a healthy diet and exercising. The production function also depends on environmental factors, such as education (Grossman, 1972). Having a certain level of education implies the ability and curiosity to continuously find out about ways on how to better produce health, which increases the efficiency of the production of health (Grossman, 1972).

Cropper (1977) and Phelps (1978) also argue within the framework of Grossman's model that people purchase preventive care in order to increase health but also to decrease the probability of ending up in a bad health state in the future. Acquiring health information online can be seen as prevention and an input to the health production function. However, it is important to note that simply acquiring information does not directly produce health, but may encourage actions that improve health. Therefore, searching for information could even be seen as an input to inputs in the health production function.

The decision to invest in health

After concluding that individuals can invest in their health through their actions, the question then becomes: how do people decide whether and how much to invest in their health? What determines the level of inputs in the production function? Recall that according to classical economists, the optimal amount of inputs is decided by making an assessment of the marginal costs and marginal health benefits of each action (Folland et al., 2017, p. 207, Kenkel, 1990). A person will stop investing in their health once the marginal costs and benefits of any additional investment are equal to each other. This framework can even help explain why people take actions that have a negative impact on health. For example, a person may compare the added health risk and monetary cost from smoking a cigarette with the improved mental health and social interaction that may follow. Ample amounts of research in economics has been done to explain the rationality behind the consumption of cigarettes, drugs and alcohol, such as the theory of "rational addiction" (Frank, 2004, Wagner & Jimison, 2003).

However, a criticism made against this argument is that people do not carry out cost-benefit analyses in conditions of perfect information (Folland et al., 2017, p. 273). Most people do not exactly know the health costs and benefits from their actions (Kenkel, 1990). Especially in the context of health, there is a large information asymmetry between health professionals and the general population. The relationship between a physician and a patient can be described as principal-agent, where the patient (the agent) hires the physician (the principal) to conduct the health information learning and problem solving for them (Laugesen et al., 2015). This is because the patient does not have the required knowledge or tools to do so (Laugesen et al., 2015). However, this knowledge gap means that in theory, health professionals can take advantage of the information asymmetry and induce demand by recommending services whose benefits may not necessarily outweigh their costs (Kenkel, 1990). This is called the physician induced demand hypothesis. Therefore, physicians do not necessarily act as "perfect agents" (Wagner et al., 2001). For this reason, some may find it important to acquire health information in order to ensure they receive the strictly necessary and good quality health care (Wagner et al., 2001, Hay & Leahy, 1982).

Therefore, acquiring additional health information (through a search online, for example) reduces uncertainty and may change a person's actions if their perception of the marginal benefits and costs of an action change (Hsieh & Lin, 1997). For example, Kenkel (1990) claims that people often tend to underestimate the marginal benefit of health care use. Therefore, because of the known informational asymmetry, people have an incentive to acquire health information which may impact their decision-making (Hsieh & Lin, 1997). Therefore, health information may provide increases in utility (Schmid, 2015).

Determinants of the decision to search for health information online

What factors can lead to the decision to seek health information? Multiple factors have been suggested in the literature. First of all, the decision is likely to depend on an individual's health condition. A person may seek information for preventative purposes, to treat an existing condition or to diagnose an unknown illness or injury (Cauley, 1987, Wagner et al., 2001). As Kenkel (1990) puts it, health status provides an incentive to acquire information. The study in Italy by Siquilini et al (2011) found that respondents with a chronic disease or who used medicines daily were significantly more likely to use online health resources. Arguably this is because their perception of the marginal benefit of the search is much higher than for someone who does not have or is not aware of having health problems (Wagner et al, 2001, Kenkel, 1990).

Secondly, the decision to search for health information will also depend on the price of acquiring information and consequently on an individual's disposable income and budget (Vistnes & Hamilton, 1995). Assuming that health information is a normal good, the demand curve for health information is downward sloping and is inversely related to price (Wagner et al, 2001). A decrease in the price of information should lead to an increased demand for health information (Wagner et al, 2001). Arguably the availability of health information online has greatly decreased the cost of acquiring information. Similarly, consumers who have a lower cost of acquiring information online will be more likely to do so than someone who faces a higher price, leading to differences between individuals in their propensity to search for health information online (Hsieh & Lin, 1997). Acquiring online information does not have a direct monetary cost as most websites are freely accessible. However, owning a device that connects to the internet and internet access does have a monetary cost. Therefore, it is expected that owning a device that can connect to the internet or not will have a strong effect on whether the person will be likely to search for information online (Rice, 2006). It is also likely that higher educated and income households have better access to the internet (Tustin, 2010, Zillien & Hargittai, 2009). This cost can be considered a fixed cost rather than a variable cost directly related to a specific online health search. Additionally, the fixed cost is split between various activities that can be done online, as people do not solely get an internet connection and a device to search for health information online.

The price of acquiring information does not only include the monetary cost of obtaining information. It also depends on the time spent obtaining information and consequently on the value of time (Cauley, 1987, Wagner et al, 2001). The travel time required to access health care is likely to have an impact on the use of online health resources. It is an opportunity cost of acquiring health care (Cauley, 1987), which will depend on the person's wage. Therefore, the impact of income on searching for health information is unclear. As Wagner et al. (2001) suggest, there could be a substitution or an income effect at play. Those who have a large cost of time may decide to substitute in-person consultations to online resources (Wagner et al, 2001). Therefore, people with a high income may be more likely than low income individuals to use online health resources due to their higher opportunity cost. At the same time, ignoring the substitution between online health and health care services, individuals with a higher wage will have a higher opportunity cost of spending time on searching for health information, especially for preventive purposes. Therefore, the substitution effect would make it less worth it to conduct one's own research on health. Alternatively, searching for health information online does not require transport costs, baby-sitting costs or waiting time costs (Cauley, 1987). This could mean that people with a lower income will be more likely to use online e-health resources than high income people. However, an income effect is also as likely. As income increases, the demand for health information may increase, assuming that preferences do not change.

Based on the substitution effect argument, one would expect that employed people will be less likely to search for health information than those not on the labour force due to the opportunity cost of time. The results of the Italian study by Siquilini et al. (2011) seem to confirm this, at least with regards to women. They found that unemployed women and students were significantly more likely to use online health resources than employed women, whereas retired women were the least likely to use online health resources. For men, the picture was different. They found that male students were more likely to use online health resources than employed males, whereas unemployed males were less likely. In the same vein, Kenkel (1990) notes that those with an occupation or education in health are perhaps less or more likely to search for health information online due to their higher level of information compared to people working in other fields.

Additionally, finding appropriate websites, evaluating the information's reliability as well as processing the different arguments requires a considerable amount of time (Wagner et al, 2001). This in turn depends on the education level of the individual. Many papers have argued education to have an impact on the acquirement of health information (Schultz, 1975, Ippolito, 1990). This could be due to higher educated population's larger experience with and understanding of the internet (Siquilini et al., 2011, Tustin, 2010, Zillien & Hargittai, 2009, Wagner et al, 2001, Rice, 2006). This is because less educated individuals may be more likely to work in manual labour and may prefer asking professionals about health, rather than doing their own research (Wangberg et al., 2015). The huge impact of education on the use of health information raises concerns that online health information may even further increase health inequalities between the less and more educated. The role of education on acquiring health information can also be explained in terms of discounting (Folland et al., 2017, p. 157). Higher educated people tend to have a lower discount rate in terms of costs and benefits that occur in the future, compared to less educated people. Therefore, they are more likely to invest in health in the present time, as they value being in good health in the future more highly (Folland et al., 2017, p. 157, Wangberg et al., 2015).

Thirdly, searching for health online information can also depend on whether a person has used health care services. People may acquire information in advance of an appointment in order to prepare themselves (Lee & Lin, 2016). Likewise, people may search for information about their symptoms and then decide whether they should book an appointment or not (Kelly et al, 2013, Bujnowska et al., 2020). Alternatively, they may wish to consult the internet after an appointment in order to gain further information, either because they are curious to learn more or they did not agree / trust the given advice (Lee & Lin, 2016). People who have been dissatisfied with their health care services have been found to be more likely to substitute the internet for health care services in the future (Tustin, 2010). For example, Siquilini et al (2011) found in their study that the rates of e-health use were significantly higher for respondents who reported a bad health care experience compared to those who did not.

There is considerable evidence showing that health care use and health depends on socio-economic status, which in turn affects an individual's probability of searching for health information. Cutler & Lleras-Muney (2006) estimated based on data from the National Longitudinal Mortality Study that an individual's additional year of education increases their life expectancy by 0.18 years (using a 3% discounting rate). Hay and Leahy (1982) also found that people with higher income and higher education levels or in a blue collar job have also been found to use health care more than people with lower income or fewer years of education. Baum & Ruhm (2003) find a negative relationship between a child's body weight and their parental education, with an extra year of the mother's education increasing a child's BMI by 1.2 percentage points. This effect follows to adulthood. However, a German study by Jurges et al. (2011) did not find an impact of additional education on being overweight or obese, but did find an impact of education on smoking, especially for women. The OECD country profile for Italy estimated that women with low education are three times more likely to be overweight than women with more education, and lower educated men are 1.3 times more likely to be overweight than more highly educated men (OECD 2021).

Due to the youths' greater use of the internet, one could expect that they would be more likely than older people to search for health information online. However, the young are also in better health than the older, meaning that they have less of an incentive to seek for health information (Kenkel, 1990). Hsieh & Lin (1997) argue that older people may be less likely to seek health information (not only online) as their payoff period of the investment is much shorter. Siquilini et al (2011) found that the impact of age depends on gender. Using the internet for health information was decreasing with age for males. However, they found that for females, the 30-41 year olds had the highest rate of using the internet for health purposes (78.8%).

Many studies, mainly conducted in the US, report the large negative impact of having health insurance on health information use online (Wagner et al, 2001). This is less likely to have an impact in countries such as Italy which has national health service (Servizio Sanitario Nazionale, SSN) that provides universal coverage largely free of charge at the point of delivery (Ferré et al., 2014). According to the 2014 Health in Transition (HiT) report, private voluntary health insurance only accounted for 0.9% of total health-care expenditure in 2012 (Ferré et al., 2014).

Lastly, whether an individual will search for health information online will depend on their preferences. Someone who cares about their health is very likely to search for health information online and to engage in other types of positive health behaviour (Suziedelyte, 2012). Meanwhile, a person who cares little about their health is likely to have an unhealthy lifestyle, even if they were suffering from health issues.

To conclude, this section has introduced to the framework used in the data analysis and has covered findings in the literature which will be guiding the data analysis. It has highlighted the perceive importance of socio-economic and health status in determining an individual's health behaviour and likelihood of acquiring health information.

4. Data and Methods

This section will cover the methodology of the thesis. Firstly, the data sources will be presented and the variables described. Secondly, the possible endogeneity of variables will be discussed alongside the solution to endogeneity, the use of an instrument. Lastly, a description of the regression methods will take place.

The data source

The data used to analyse the research questions is repeated cross-sectional data from an annual household survey, "*Aspects of daily life*", by the Italian National Institute of Statistics (Istat). Each year Istat interviews around 25,000 randomly chosen households across the whole of Italy. Since the chosen households are obliged to respond by law (Istat, 2019), the sample can be considered nationally representative. Every Italian person thus has an equal probability of being selected to respond to the survey. The survey respondents answer questions related to their family, education, health, work, leisure, and the use of information and communications technologies. The data is therefore observational, as answers were collected via an interview in a retrospective manner.

Data from the *Aspects of daily life* surveys from years 2013 to 2016 was pooled. This means that the sample size is large (N= 153,813) and can be considered nationally representative. However, among the 153,813 observations, only 68,517 respondents answered all the questions of interest, resulting in the regressions varying in sample sizes depending on who provided information for all variables. On top of the survey data, additional information was extracted from other sources. As will be further explained later, additional data was needed for an instrument. Data available on the Eurostat website on the differences in broadband access in NUTS2 regions in Italy was extracted. Similarly, data on regional household disposable income as well as regional population density was obtained from Eurostat.

Methodology

Analysis was conducted using STATA15. A wide range of regression methods were used, including probit, ordinary least squares as well as two stage instrumental variable regressions. This section will provide a description of the regression variables, the regression methods, issues of endogeneity and solutions.

Based on the four research questions, the analysis includes multiple regressions with different dependent variables. They are the following dummy variables *intsearch, healthcareuse, healthydiet,* *vegetables, fruits, overweightorobese* or *obese.* Worthy of note is that *intsearch* is also the variable of interest in the other 4 regressions models. In its simplest form, the regression equation is as shown in formula 1 below, with Y_{iry} being the health outcome or behaviour of interest. i indexes individuals, r indexes NUTS2 regions, and t indexes years. The variable of interest is *intsearch* and β_1 is the coefficient of interest. λ_r and γ_t are region and year fixed effects, respectively. They are included in order to take into account time invariant differences between different regions, and differences across years. The regressions control for different sets of covariates X_{irt}. They are added separately into the regression in order to see whether the coefficient of *intsearch* is sensitive to them. The reasoning behind the inclusion of the control variables is explained in the next paragraph. The equations also include a constant β as well as the error term ε_{irt}. In the case of research question 1, the dependent variable is *intsearch* as seen in equation 2.

(1)$Y_{irt} = \beta + \beta_1 intsearch_{irt} + \beta_2 X_{irt} + \lambda_r + \gamma_t + \varepsilon_{irt}$

(2) $Y_{irt} = \beta + \beta_1 X_{irt} + \lambda_r + \gamma_t + \varepsilon_{irt}$

The covariates

Covariates in the analysis included NUTS2 region, survey year, household internet access, demographics (gender, age, marital status, children in household), the presence of one or more medical conditions, self-reported health status, education level, employment status, self-reported level of economic resources, health behaviour such as smoking, frequent alcohol consumption, keeping track of one's weight and salt consumption, difficulty reaching first aid, and readership of books and / or newspapers. Table 1 below shows summary statistics for the control variables used in the analysis. 68% of respondents in the sample are aged between 26 and 64 years. 18% at least have an undergraduate degree. 18% also suffered from a disease or long-term health problem in last 6 months.

Table 1

Type of Variable	Variable name	Definition	Mean	Standard Deviation	Minimum-Maximum
Fixed effects	**year**	Survey year			2013-2016
	regions	NUTS2 regions			10-999
	popdensity	Persons per square kilometre for each NUTS2 region	213.3	122.9	39.2- 439.2

Demographics	**female**	1= yes , 0=no	0.479	0.5	0-1
	youth	1= if <=25 years old, 0=no	0.26	0.439	0-1
	adult	1= if >= 26 & <= 64 years old, 0=no	0.682	0.466	0-1
	senior	1= if >= 65 years old, 0=no	0.058	0.233	0-1
	unmarried	1= yes , 0=no	0.485	0.5	0-1
	divorcedwidow	1= divorced/separated /widow(er), 0=no	0.095	0.293	0-1
	married	1= Married / cohabitant with spouse / civilly united, 0=no	0.42	0.494	0-1
	children	1= household with children, 0=no	0.752	0.432	0-1
Socio-economic status	**university**	1= undergraduate &/ postgraduate education , 0=no	0.184	0.388	0-1
	highschool	1= high school education , 0=no	0.442	0.497	0-1
	middleschool	1= middle school education, 0=no	0.27	0.444	
	primarynone	1= primary school or no education, 0=no	0.104	0.305	0-1
	Books	1= read a book in the past 12 months or reads the newspaper 3-5 times p/ week , 0= no	0.641	0.48	0-1
	dispincome	Regional (NUTS2) household disposable income	16672.05	3203.333	11,300 - 21,250 PPS
	higheconresources	1= excellent / adequate household economic resources, 0= scarce / absolutely insufficient household economic resources	0.615	0.487	0-1
	Insurance	1= private health or accident insurance, 0= no	0.213	0.41	0-1
	Employed	1= yes, 0= no	0.559	0.497	0-1
	Unemployed	1= yes, 0= no	0.136	0.343	0-1

	Inactive	1= inactive in the labour market, 0= no	0.306	0.461	0-1
Health state	**healthproblems**	1= disease or long-term health problem in last 6 months, 0= no	0.18	0.384	0-1
	goodselfhealth	1= very good / good self-reported health status, 0= not good not bad / very bad / bad	0.81	0.392	0-1
	notgoodnotbadselfhealth	1= not good not bad self-reported health status, 0= very good / good / very bad / bad	0.169	0.374	0-1
	badselfhealth	1= very bad / bad self-reported health status, 0= very good / good / not good not bad	0.021	0.143	0-1
Health behaviour	**frequentalcohol**	1= consumption of wine / beer / aperitifs/ bitters / liquors on weekly basis, 0= less than weekly	0.235	0.424	0-1
	smoker	1= current smoker , 0= no never or in the past	0.217	0.412	0-1
	saltconscious	1= pays attention to their level of salt, 0= no	0.682	0.466	0-1
	weightconscious	1= checks their weight, 0= no	0.87	0.337	0-1
E-patient characteristics	**intaccess**	1: internet access at home, 0: no	0.97	0.169	0-1
	difficultyfirstaid	1: a little bit difficult / very difficult to reach first aid, 0: no difficulty / I don't know	0.514	0.5	0-1

As highlighted in the previous section, controlling for socio-economic status is absolutely crucial. Due to a lack of information on household income, the variable *insurance* is used as a control for a person's disposable income on top of level of education. There are various papers in the literature arguing that voluntary supplementary insurance is positively correlated with income (Jones et al., 2006, Besley, 2001, Propper, 2000). Having additional private health insurance in a public health care system may be a sign of wealth or of having a job with fringe benefits. According to the 2014 HiT report, private insurance in Italy is more likely among higher socio-economic groups, with 16.3% of the families in the highest quintile compared to 1.4% in the lowest quintiles having it, as well as among people with higher education (Ferré et al., 2014). Of course, having insurance or extensive insurance has long been argued to increase health care use Kenkel, 1990, Hay and Leahy, 1982).

Including covariates for someone's health status or a disease is important as health problems naturally increases demand for health care use and will affect an individual's health behaviour (Kenkel, 1990, Cauley, 1987). The healthy group of individuals is used as the reference group in the regressions. Controlling for age is also necessary. The elderly are more likely to have health problems, and hence may have a higher health care use and possibly more incentive to take preventive measures such as searching for health information online and eating a healthy diet (Dittus et al., 1995). However, at the same time, Baum & Ruhm (2009) found that BMI can increase by around 0.6 percentage points every year of age, and hence the elderly may be more likely to be obese.

Controlling for gender is important, since women may use more health care for reasons including being pregnant or outliving their partner and thus requiring more formal care (Kenkel, 1990, Hay and Leahy, 1982). Health behaviour may also vary between genders. It has been suggested that women may be more likely to search for health information (Farajallah et al., 2015, Bujnowska et al, 2020). It is also possible that individuals who are single, divorced or widowed are more likely to have negative health behaviour than those who are married, as they are less likely to engage in activities that put their life and health in danger (Umberson, 1987, Roos et al. 1998, Kamphuis et al., 2006). Siquilini et al (2011) found that unmarried, separated or divorced respondents were more likely to use health information resources online than respondents who were cohabiting. Being married and having children may also increase the probability of using health care (Hay and Leahy, 1982, Tiffin & Arnoult, 2010, Cliff et al., 2019).

The dependent variables

Searching for health information online

The variable *intsearch* is equal to 1 if the respondent had searched health information online regarding "accidents, illnesses, nutrition, health improvements, etc.). Across the whole aggregated sample of the population, 22.78% of respondents had searched for health information online. However, it is important to note that 47.44% of the total respondents did not answer the question. After dropping all the observations with a missing variable for *intsearch*, the variable of interest, the sample size decreased from 153,813 observations to 80,839 observations. Table 2 shows that the percentage of households who searched for health information online decreased over the years, which is against expectations. There appears to be a huge drop of around 10% between 2013 and 2014. This also coincides with a huge increase in missing variables, from 10,432 missing answers in 2013 to 22,203 in 2014. This is possibly due to a change in the questionnaire or survey collection.

Table 2

Year	% respondents who searched for health information online	Missing values for intsearch
2013	51.2	10,432
2014	42.5	22203
2015	42.6	21260
2016	41.6	19079

There are some considerable differences between the total sample, the sample restricted to those who answered the question about searching for health information online, and the sample of respondents who did not answer the question (table 5 in the appendix). For example, the percentage of respondents with internet access of is much lower among the excluded portion of the sample (50%) compared to percentage in the final sample of 97%. The percentage of the respondents aged 65 or more is also much lower in the restricted sample (0,06%) compared to the total or the excluded sample.

Table 3 below shows differences in the characteristics of those who had searched and had not searched for health information online. For each characteristic, the rows show the composition of respondents who had or had not searched for health information online. As expected, the two groups seem to differ in terms of age, gender, education, employment status, health status, economic resources, health behaviour. For example, the group of respondents who had searched for e-health information consisted of much more adults compared to the group who had not searched for information.

Table 3

	Respondents who had searched for health information online	Respondents who had not searched for health information online
Internet access	.976794	.9655707
Difficulty reaching first aid	.5075355	.518744
Youth	.1728892	.3269294
Adult	.7620312	.6213514
Senior	.0650796	.0517192
Female	.5211223	.4459339
University	.2334875	.146687
High school	.4867272	.4077502
Middle school	.2328595	.2986792
Primary school or none	.0469258	.1468835

Employed	.5869629	.534692
Unemployed	.1315525	.1395602
Inactive	.2814845	.3257478
Children	.720757	.7765964
Health problems	.2232164	.1473072
Good self-reported health	.7754753	.8372448
Bad self-reported health	.0272307	.0161118
Self-reported high economic resources	.6354845	.5994072
Self-reported low economic resources	.3645155	.4005928
Health or accident insurance	.2393335	.1930447
Reads books and /or newspapers frequently	.7061044	.5907956
Overweight or obese	.3676714	.3442637
Health care use	.2554292	.1062821
Exercises frequently	.4853589	.4763311
Eats a healthy diet	.2566008	.2158827
Frequently consumes alcohol	.2506886	.2217158
Smoker	.21973	.2145923
Salt conscious	.7338022	.6430396
Weight conscious	.8936281	.8513255

The sub-groups with the highest percentage of respondents who had searched for health information were respondents who had used health care in the last x months (65%), respondents who reported having bad health (56%), respondents who had a university education (55%), respondents who had health problems (54%) and the sub-group of the population aged 65+ years (49%). The percentage of respondents who had searched for health information online was the lowest amongst the following sub-groups: respondents whose highest education completed was primary or who had no education (20%), respondents aged 25 or under (29%), respondents without internet access at home (34%), respondents who did not pay attention to their weight (35%) and respondents who did not read books or newspapers frequently (35%). This information can be found in table 1 of the appendix.

Health care use

The dependent variable *healthcareuse* is equal to 1 if the respondent had been to first aid / emergency room, outpatient medical examination, home medical examination, hospitalization, used a P.A. or public service managers to book medical visits in recent months. Regrettably the "*Aspects of daily life*" survey did not ask directly whether the respondent had visited a general practitioner or a specialist in recent months. Therefore, health care use only includes very specific types of health care use. For this reason, in the final sample only 17% of respondents had used health care according to the above definition. The percentage of respondents who used health care increased in 2014 and 2015 compared to 2013, but fell in 2016.

As expected, the rates were the highest among respondents with a bad self-reported health state (41%), respondents with health problems (26%), respondents with a university degree (22%), respondents age 65 or more (21%) and respondents who had health or accident insurance (21%). The lowest percentages were among respondents who had at the most completed primary school (10%), , respondents aged less than 25 years (10%), respondents who had at the most completed middle school (15%), respondents who reported having a good health state (15%) as well as respondents who did not have any health problems (15%). This information can be found in table 2 of the appendix.

Healthy diet

The dependent variable *healthydiet* is equal to 1 if the respondent ate vegetables and fruits at least once a day, and ate salty snacks (chips, popcorn, pretzels, olives) and sweets / desserts (cakes, sweet snacks, ice cream, etc.) only a few times per week or less. For the rest of the thesis, this behaviour will be referred to as eating a healthy diet. The criteria for the variable is based on what is generally considered a healthy diet and what dietary guidelines usually recommend. For example, the OECD considers a poor diet to be low fruit and vegetable consumption as well as high sugar and salt consumption (OECD & EOHSP, 2019). It is also how healthy diet is defined in a 2013 paper by Anekwe & Rahkovsky. Following the approach by Roos et al., the respondents were classified into two categories; those whose food behaviour is in line with dietary guidelines and those whose diet goes against the guidelines (Roos et al., 1998). Additionally, separate regressions are run using a daily consumption of vegetables or fruits as the dependent variables respectively.

Across the final sample, only 23% of respondents ate a healthy diet. It appears that the majority of the respondents did not eat a healthy diet. The frequency of healthy eaters increased over the study years. The highest percentages of respondents eating a healthy diet were among respondents aged 65 or more (41%), respondents who had completed university education (30%), respondents with a self-reported bad health state (30%), respondents with health problems (29%) and respondents who were salt conscious (27%). The lowest percentage was among the respondents who had completed primary school at the most (12%), respondents aged 25 or younger (12%), respondents that did not pay attention to their salt consumption (16%), respondents that did not pay attention to their weight (18%) and respondents who had at the most completed middle school (20%). This information can be found in table 3 of the appendix.

Weight

The dependent variable *overweightorobese* is equal to 1 if the respondent's body mass index (BMI) is equal to or higher than 25. It must be noted that BMI has been criticised for not being an accurate indicator of obesity, as it does not distinguish between fat and muscle (Folland et al., 2017, p. 204). However, it was the best measure of weight available. Data on each respondent's weight and height was provided. Based on this, the respondents aged 18 or above were categorised into the BMI categories underweight, standard weight, overweight, obese. The BMI value was calculated according to the following formula: weight (kg) / (height (m))^2 (Folland et al., 2017, p. 204). People with a BMI higher than 25 are considered overweight, and those with a BMI higher than 30 are considered obese (Folland et al., 2017, p. 204). Respondents below the age of 18 were categorised based on the extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. The reason for combining respondents considered overweight and obese instead of defining the variable based on only being obese is due to the data set for the first survey year having already combined being overweight and obese for respondents younger than 18. Additionally, separate regressions are run using a binary variable for being obese, *obese*, as the dependent variable. This regression excludes respondents younger than 18 years.

In the final sample, 35% of respondents were considered overweight or obese according to their BMI. The percentage of respondents overweight or obese increased sharply between 2014 (33%) and 2016 (37%). The percentage of overweight or obese respondents was highest in the following sub-groups: respondents aged 65 and over (59%), respondents with a bad self-reported health state (51%), respondents with health problems (48%), respondents who were frequent consumers of alcohol (47%) and males (46%). The lowest percentage were among respondents aged 25 or younger (16%), females (24%), respondents who had at the most completed primary school (26%), respondents who exercised frequently (30%) and respondents out of the labour force (31%). This information can be found in table 4 of the appendix.

Endogeneity

As commonly found with observational data, correlation between the dependent and the explanatory variable of interest does not necessarily mean causation (Cunningham, 2021). The binary explanatory variable of interest *intsearch* is likely to suffer from bias when used in the regressions. This is because in the absence of randomised experiment, whether an individual searches for information online is not random, but chosen by the individual. Therefore, there is self-selection into the group of e-health users.

In this analysis, there are at least two potential causes of biases. Firstly, the model may suffer from omitted variable bias. It could be that a third factor impacts both the explanatory and dependent variables. For example, an intrinsic interest in health can cause a person to search for information online and use health care more often. Therefore, there is unobserved individual heterogeneity in searching for information online. Since part of the variation in the independent variable is caused by the omitted variables, the independent variable is correlated with the error term, which means that the conditional independence assumption does not hold (Cunningham, 2021). To ensure exogeneity, the variation of the explanatory variable has to be independent of the error term (Cunningham, 2021). This means that a simple ordinary least squares regression will not provide the causal estimate of the effect of searching for information online on health behaviour.

Secondly, there may be reverse causality. The direction of the causality between *intsearch* and the dependent variables is not always certain. For example, a health care visit can cause a person to search for further information online, especially if they were not satisfied with their visit. Alternatively, an individual may search for nutritional information online because of their weight, instead of information affecting a person's weight. Similarly, there is potentially reverse causality between BMI and some control variables. While a person's exercise levels and diet have an impact on their weight, a person's weight may in turn also impact their exercise level, occupation or even relationship status.

Instrumental variable regression

Using an exogenous instrumental variable (IV) is a common approach to resolving endogeneity issues (Terza et al., 2008). It is especially appropriate in this case for two reasons. Firstly, a randomised experiment was not conducted and secondly, it is not possible to control for unobservable variables such as intrinsic interest in health. Using an instrument helps to get unbiased and consistent estimates despite the endogeneity, as if a simple ordinary least squares regression was used (Woolridge, 2016).

An instrument commonly used in the literature when analysing health information online has been broadband access to the internet (Carrieri et al., 2019, Suziedelyte, 2012). While there are structural differences in broadband coverage across Italy (Carrieri et al., 2019), finding an appropriate instrumental variable for *intsearch* was not an easy task. The survey itself asks whether the respondent has access to the internet in their home. However, this can be correlated with the household's income, for example, which cannot be controlled for in the regressions. Therefore, the variable intaccess cannot be used as an instrument and an exogenous source of variation in broadband access was needed.

Data was available on the Eurostat website on the differences in percentages in household broadband access in NUTS2 regions in Italy. A variable was then constructed equal to the percentage of households with access to broadband for each NUTS2 region and for every survey year 2013-2016. Average household broadband access increased across all regions, from 67.9% in 2013 to 76.7% in 2016. Since NUTS2 regions and survey year can be controlled for in the regressions, the resulting variation in broadband access can be considered as exogenous. In order to further ensure exogeneity, the instrument used is a dummy variable *highbroadband*, which is equal to 1 if the percentage of households with access to broadband in a region is higher than or equal to the median percentage of 75%, and equal to 0 if the percentage is below the median.

The regression methods

Due to the dependent variables and the endogenous explanatory variable *intsearch* being binary, a non-linear regression model was needed. The commonly used probit or logit models would be appropriate to use in this case. However, the analysis requires an instrumental variable. The basic two stage least squared (2SLS) regression model does not apply since the model is non-linear. There are various alternatives in the literature on two-stage regression methods for non-linear models.

The most straightforward method would be using the *ivprobit* regression developed by Newey (1987) since it is appropriate for regressions with binary dependent variables and endogenous explanatory variables (equation 4). However, *ivprobit* assumes that the errors are normally distributed and that the endogenous explanatory variable is continuous. *Intsearch* is a binary variable. This means that the errors in the first stage regression are not normally distributed and that the estimates may not be valid.

Terza et al. (2008) have developed another IV regression model, the two-stage residual inclusion (2SRI) approach (equation 6). It was specifically recommended for health economic research. In this two stage regression, the first stage is the same as in *ivprobit*. However, in the second stage, the endogenous variables are not replaced, but the first-stage residuals are added as regressors (Terza et al., 2008). Having a binary endogenous variable is thus not a problem. Terza et al. (2008) recommend correcting the standard errors of the results. However, due to the large sample size, correcting for the standard errors using the bootstrap method was too time-consuming. Therefore, the standard errors could not be corrected, meaning that the analysis of the statistical significance of the coefficients in the output were not necessarily accurate.

Lewbel and Dong's special regressor method was also considered. It estimates a binary choice model with endogenous regressors (Baum, 2012). The assumption is that the model includes a special regressor that is exogenous, appears additively in the model and is continuously distributed. However, deciding which variable to use as a special regressor was not obvious. When using age as the special regressor, following the example given by Lewbel and Dong, the results were not consistent with the other two-stage regression results (2SRI and *ivprobit*). Therefore, the use of this method was dropped.

Due to a lack of a perfect method, ivprobit and 2sri, were both used and compared since both methods have some crucial flaws. Additionally, a one stage probit (equation 3), ordinary least squares (OLS) (equation 5) and 2SLS regressions (equation 6) were ran alongside to check the robustness of the results. Therefore, the following five different regression methods are used. They can be written in these five equations:

$$(3)\ E(Y|X) = \Phi(\beta + \beta_1 intsearch_{irt} + \beta_2 X_{irt} + \lambda_r + \gamma_t)$$

$$(4)\ E(Y|X) = \Phi(z + z_1 highbroadband_{rt} + z_2 X_{irt} + \lambda_r + \gamma_t)$$

$$E(Y|X) = \Phi(\beta + \beta_1 \widehat{intsearch}_{irt} + \beta_2 X_{irt} + \lambda_r + \gamma_t)$$

$$(5)\ Y_{irt} = \beta + \beta_1 intsearch_{irt} + \beta_2 X_{irt} + \lambda_r + \gamma_t + \varepsilon_{irt}$$

$$(6)\ intsearch_{irt} = z + z_1 highbroadband_{rt} + z_2 X_{irt} + \lambda_r + \gamma_t + \epsilon_{irt}$$

$$Y_{irt} = \beta + \beta_1 \widehat{intsearch}_{irt} + \beta_2 X_{irt} + \lambda_r + \gamma_t + \varepsilon_{irt}$$

$$(7)\ )\ intsearch_{irt} = z + z_1 highbroadband_{rt} + z_2 X_{irt} + \lambda_r + \gamma_t + \epsilon_{irt}$$

$$Y_{irt} = \beta + \beta_1 \widehat{intsearch}_{irt} + \beta_2 X_{irt} + \epsilon_{irt} + \lambda_r + \gamma_t + \varepsilon_{irt}$$

Additionally, the reduced forms of the two stage regressions are also estimated using probit and OLS.

$$(8)\ E(Y|X) = \Phi(\beta + \beta_1 highbroadband_{rt} + \beta_2 X_{irt} + \lambda_r + \gamma_t)$$

$$(9)\ Y_{irt} = \beta + \beta_1 highbroadband_{rt} + \beta_2 X_{irt} + \lambda_r + \gamma_t + \varepsilon_{irt}$$

Instrumental validity

For an instrumental variable to be valid, it needs to satisfy four conditions. Firstly, the instrument z should be correlated with the endogenous explanatory variable (instrumental relevance) (Woolridge, 2016). It should explain variation in the endogenous variable. Instrumental relevance can be tested by running a simple regression between the instrument and the explanatory variable (Woolridge, 2016). The null hypothesis that the instrument has no impact on the endogenous variable needs to be rejected at a low significance level (1% for example) in order for the instrument to be considered relevant (Woolridge, 2016). Broadband access is arguably a relevant instrument since naturally a person is more likely to search for information online if they have access to internet.

Similarly related to relevance, the correlation between the instrument and the endogenous variable should not be weak. A weak instrument means that the instrument only explains very little variation in the independent variable (Terza et al., 2008), leading to estimates being inconsistent / biased (Woolridge, 2016). Also the sign as well as the magnitude of the coefficient of the instrument on the endogenous variable should be paid attention to (Woolridge, 2016).

Finding out whether the instrument is weak can be done by using an F-test of joint significance of the instrument in the first stage of the regression. The null hypothesis is that the coefficient of the instrument is equal to 0. Following Staiger & Stock's (1997) rule of thumb, the F-statistic should be larger than 10 in order to conclude that the instrument is not weak. From the output in table 4, it can be seen that the instrument *highbroadband* is relevant. It is correlated with the the endogenonus variable *intsearch* and a significant determinant of its variation. However, the magnitude of the effect is very small. When including all the relevant control variables, the F-statistic of the test of joint significance falls below 10, although it remains very close to 10.

Table 4

	Coefficient OLS	**F-statistic**	**Coefficient Probit**	**F-statistic**
Baseline model(i) N = 80716	0.0239*** (0.00634)	14.23***	0.0620*** (0.0161)	14.73***
Incl. intaccess (ii) N = 80715	0.0235*** (0.00634)	13.75***	0.0611*** (0.0162)	14.30***

Incl. demographics N = 80715 (iii)	0.0243*** (0.00626)	15.06***	0.0663*** (0.0164)	16.38***
Incl. health care use, health status & insurance (iv) N = 77906	0.0205*** (0.00628)	10.66***	0.0587*** (0.0169)	12.08***
Incl. health behaviour (v) N = 72834	0.0202*** (0.00651)	9.64***	0.0573*** (0.0174)	10.84***
Incl. education, employment & econ resources (vi) N = 72873	0.0230*** (0.00662)	12.05***	0.0626*** (0.0172)	13.20***
Incl. all variables N = 69310	0.0184*** (0.00667)	7.61***	0.0526*** (0.0179)	8.66***

i: Model includes variables *year* (separate dummies*), regions (*separate dummies), *dispincome, popdensity*
ii: Add *intaccess* to baseline variables
iii: Add *intaccess, youth, senior, female, unmarried, divorcedwidow, children* to baseline variables
iv: Add *healthcareuse, difficultyfirstaid, notgoodnotbadselfhealth, badselfhealth, healthproblems, overweightobese, insurance* to model iii
v: Add *frequentalcohol, smoker, saltconscious, weightconscious, healthydiet, frequentexercise* to model iv
vi: Add *university, middleschool, primarynone, books, employed, unemployed, higheconresources, insurance* to model iii

The second required condition is that the instrument should be uncorrelated with the error term (called exogeneity) (Woolridge, 2016). In other words, it should not be correlated with any of the other determinants of the dependent variable. This assumption is hard to test or prove (Woolridge, 2016). The risk of endogeneity can be minimised by controlling variables in the regression that are believed to be correlated with the instrument which also affect the dependent variable (Woolridge, 2016). The regressions in the analysis control for various factors, including regional income and population density.

Thirdly, the instrument should also have no direct impact on the dependent variable (the exclusion restriction) (Woolridge, 2016). This means that the only impact it should have on the dependent variable is through the endogenous variable (Woolridge, 2016). This assumption also cannot be tested. Broadband access can be considered a generally exogenous instrument, as differences in broadband coverage are unlikely to affect people's health behaviour. Lastly, the monotonicity assumption must hold. Essentially, it is assumed that there are no defiers, i.e. the simple presence of an instrument does not lead to a behavioural reaction.

5. The results

This section will present the results of the regression analyses. The output will be used to reach conclusions regarding the four research questions presented in the introduction. Whilst the output cannot be used to accept the hypotheses, they can lead to their rejection. According to common customs, the coefficients are considered significant if their p-value is smaller than 0.1.

Internet search

The following table 5 shows the regression results of a probit regression of covariates on the dependent variable *intsearch*. The model includes 69,310 respondents. The pseudo r-squared was 0.063.

Table 5

Variable	Coefficient
dispincome	0.000223*** (0.0000476)
popdensity	0.00499** (0.00245)
highbroadband	0.0526*** (0.0179)
intaccess	0.214*** (0.0300)
year=2013	0 (.)
year=2014	-0.118*** (0.0237)
year=2015	-0.283*** (0.0367)
year=2016	-0.336*** (0.0450)
youth	-0.175*** (0.0180)
senior	-0.0886*** (0.0244)
female	0.164*** (0.0110)
unmarried	-0.0348*** (0.0129)
divorcedwidow	-0.0526*** (0.0175)
children	-0.0190 (0.0124)
healthcareuse	0.581*** (0.0132)
difficultyfirstaid	0.0182* (0.00998)
notgoodnotbadselfhealth	0.0814*** (0.0142)
badselfhealth	0.0821** (0.0361)
healthproblems	0.128*** (0.0143)
overweightobese	0.00758 (0.0111)
insurance	0.0342*** (0.0123)
university	0.0869*** (0.0134)
middleschool	-0.194*** (0.0121)
primarynone	-0.456*** (0.0337)
booksnewspapers	0.182*** (0.0111)
employed	-0.0128 (0.0144)
unemployed	0.0264 (0.0179)
higheconresources	0.00665 (0.0107)
frequentalcohol	0.0654*** (0.0121)
smoker	0.0263** (0.0121)
saltconscious	0.137*** (0.0111)
weightconscious	0.177*** (0.0150)
healthyeating	-0.0169 (0.0117)
frequentexercise	0.0674*** (0.0102)

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Firstly, it is important to look at whether any of the other dependent variables have a significant impact on whether the respondents searched for health information online. Worthy of note is the significant (at 1%) positive correlation between *healthcareuse* and *intsearch*. This is an indication of potential reverse causality, as health care use seems to increase the likelihood of searching for health information online by 0.581 percentage points on top of *intsearch* potentially leading to increased health care use. This finding gives support to the choice of using an instrument for *intsearch*. Being overweight or obese as well as eating a healthy balanced diet were not found to have a significant impact on *intsearch*.

As seen in the data description, there appears to be a significant decreasing trend in searching for e-health information in Italy over the survey years (2013-2016). There are also significant differences in the number of individuals searching online between regions. The coefficients for the regions variables are not included in the tables due to their large number (20). Respondents coming from regions with a higher population density and average household disposable income were significantly more likely to have searched for e-health than respondents from regions with lower levels. However, the coefficients are very small in magnitude. All of these interpretations are made under the assumptions that other factors are kept constant. As expected, having access to internet in the house as well as having a higher than the median share of households with access to broadband in a region significantly increases searching for health information online.

Similarly, having difficulties to reach first aid, health problems and a perception of being in poor health increased the likelihood of e-health information seeking compared to who do not experience difficulties. Being salt or weight conscious was also found to be significantly positively correlated with *intsearch*. Surprisingly, frequent alcohol drinkers and smokers were significantly more likely to search for health information than respondents who do not drink or smoke regularly. However, their effects are very small in magnitude (<0.1). Respondents who exercised frequently were significantly more likely to search for health information online than those who do not.

Females were significantly more likely to have searched for information online than men, ceteris paribus. The output also shows that age has an impact on internet searching. Respondents aged 25 or less and those aged 65 or more were found significantly less likely to search for information than 26 to 64 year old respondents. Unmarried, divorced or widow(er)s respondents were also significantly less likely to have searched for information than respondents who were married. Meanwhile, no significant correlation between having children and *intsearch* was found.

Socio-economic status also seems to play a role. Respondents who had completed university were significantly more likely to have searched for e-health information than respondents who competed high school, while those who had at the most completed middle or primary school were significantly less likely to have searched for information. Additionally, respondents who read books or newspapers frequently were significantly more likely to have searched for health information online. Employment and economic resources were not found to have a significant impact on health information seeking. However, respondents with private health or accident insurance were significantly more likely to search for health information online.

Health care use

Table 6 shows the coefficients of *intsearch* on health care use using *highbroadband* as an instrument using seven different regression methods. Searching for information online is positively and significantly correlated with health care use when using the probit, OLS and 2SRI methods. *Highbroadband* and health care use are also positively correlated in the OLS reduced form. This is not necessarily an endogeneity issue, as long as the correlation between *highbroadband* and health care use is due to searching for health information online. However, the correlation between *intsearch* and health care use is not significant in the IV Probit and 2SLS regressions. The *intsearch* coefficient varies considerably from 0.131 to 1.214 depending on the regression used. The results are therefore sensitive to the type of regression used. The standard errors also vary greatly, and they are higher when using the instrument. Including the different sets of control variables only slightly varied the coefficient (see table 6 in the appendix). For the IV probit and 2SLS, the coefficients were statistically significant until including all of the relevant control variables.

Table 6

	Probit	**Reduced form Probit**	**IV Probit**	**OLS**	**Reduced form OLS**	**2SLS**	**2SRI**
Intsearch	0.515***	0.0217	1.214	0.131***	0.00885*	0.435	0.432*
N = 69310	(0.0119)	(0.0210)	(0.798)	(0.00295)	(0.00514)	(0.271)	(0.247)

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

When looking at the impact of control variables in table 7 below, all methods show that the respondents were significantly more likely to have used health care in 2014, 2015 and 2016 compared to 2013. Being divorced or widowed was positively and significantly correlated with health care use. Respondents with a "not good, not bad" or "bad or very bad" self-rated health state were significantly more likely to have used health care than those whose health state was "good or very good". Meanwhile, having health problems was not significantly correlated with health care use in all regression methods. Having insurance was significantly positively correlated with health care use. Employed and unemployed respondents were more likely to have used health care compared to those inactive on the labour market.

Table 7

	Probit	**IV Probit**	**OLS**	**2SLS**	**2SRI**
Intaccess	0.118*** (0.0371)	0.0474 (0.101)	0.0385*** (0.00765)	0.00124 (0.0247)	0.00142 (0.0226)
2014	0.408*** (0.0303)	0.394*** (0.0591)	0.0906*** (0.00566)	0.100*** (0.00829)	0.100*** (0.00721)
2015	0.419*** (0.0436)	0.443*** (0.0430)	0.0861*** (0.00961)	0.122*** (0.0225)	0.122*** (0.0205)
2016	0.279*** (0.0536)	0.331*** (0.0647)	0.0474*** (0.0119)	0.0956*** (0.0299)	0.0952*** (0.0274)
Population density	0.00615* (0.00323)	0.00389 (0.00450)	0.00171*** (0.000563)	0.000665 (0.000974)	0.000646 (0.000873)
Youth	-0.0807*** (0.0222)	-0.0261 (0.0756)	-0.0218*** (0.00488)	0.00756 (0.0190)	0.00731 (0.0173)
Senior	-0.0188 (0.0277)	0.00891 (0.0441)	-0.0124* (0.00740)	0.00342 (0.0125)	0.00323 (0.0115)
Female	0.0546*** (0.0130)	0.00299 (0.0678)	0.0220*** (0.00319)	-0.00659 (0.0181)	-0.00634 (0.0165)
Married	-	-	-	-	-
Unmarried	-0.0363** (0.0151)	-0.0226 (0.0244)	-0.0110*** (0.00385)	-0.00418 (0.00586)	-0.00421 (0.00543)
Divorced or widowed	0.0571*** (0.0198)	0.0655*** (0.0198)	0.0141** (0.00548)	0.0213*** (0.00724)	0.0212*** (0.00675)
Children	0.00278 (0.0143)	0.00795 (0.0151)	-0.000455 (0.00377)	0.00273 (0.00443)	0.00275 (0.00414)
Difficultyfirstaid	-0.0159 (0.0118)	-0.0192* (0.0116)	-0.00351 (0.00291)	-0.00608* (0.00346)	-0.00610* (0.00322)
Overweight or obese	0.0172 (0.0129)	0.0138 (0.0137)	0.00346 (0.00325)	0.00208 (0.00355)	0.00210 (0.00330)
Notgoodnotbad healthstate	0.187*** (0.0160)	0.144** (0.0690)	0.0537*** (0.00447)	0.0355*** (0.0123)	0.0357*** (0.0112)
Badhealthstate	0.653*** (0.0368)	0.553*** (0.178)	0.219*** (0.0132)	0.185*** (0.0252)	0.185*** (0.0233)
Healthproblems	0.143*** (0.0161)	0.0903 (0.0767)	0.0478*** (0.00457)	0.0221 (0.0167)	0.0222 (0.0153)
Insurance	0.0805*** (0.0141)	0.0620* (0.0321)	0.0238*** (0.00383)	0.0159** (0.00636)	0.0159*** (0.00591)
University	0.0756*** (0.0153)	0.0419 (0.0493)	0.0264*** (0.00417)	0.00928 (0.0115)	0.00944 (0.0105)
Middle school	-0.0405*** (0.0147)	0.0181 (0.0751)	-0.0201*** (0.00339)	0.0131 (0.0210)	0.0128 (0.0191)
Primaryornone	0.00692 (0.0393)	0.129 (0.150)	-0.0234*** (0.00904)	0.0491 (0.0461)	0.0486 (0.0421)
booksnewspaper	0.0318** (0.0133)	-0.0223 (0.0690)	0.0174*** (0.00314)	-0.0136 (0.0195)	-0.0134 (0.0178)
Employed	0.108*** (0.0171)	0.100*** (0.0245)	0.0263*** (0.00414)	0.0260*** (0.00445)	0.0260*** (0.00410)
Unemployed	0.0945*** (0.0216)	0.0776** (0.0351)	0.0227*** (0.00499)	0.0163** (0.00669)	0.0164*** (0.00612)
Higheconresources	-0.0140 (0.0128))	-0.0143 (0.0123)	-0.00316 (0.00309)	-0.00384 (0.00332)	-0.00388 (0.00307)

Frequentalcohol	0.0618***	0.0374	) 0.0178***	0.00569	0.00585 (0.00766)
	(0.0141)	(0.0370)	(0.00364)	(0.00842)	
Smoker	0.00759	0.0000573	0.00123	-0.00292	-0.00289 (0.00420)
	(0.0143)	(0.0168)	(0.00353)	(0.00456)	
Saltconscious	0.0189	-0.0210	0.0108***	-0.0121	-0.0120 (0.0133)
	(0.0134)	(0.0511)	(0.00315)	(0.0146)	
Weightconscious	0.0271	-0.0235	0.0150***	-0.0140	-0.0137 (0.0169)
	(0.0181)	(0.0655)	(0.00414)	(0.0186)	
Healthydiet	0.0121	0.0155	0.00179	0.00424	0.00421 (0.00372)
	(0.0136)	(0.0135)	(0.00350)	(0.00399)	
Frequentexercise	0.0497***	0.0252	0.0166***	0.00394	0.00404 (0.00772)
	(0.0120)	(0.0355)	(0.00298)	(0.00844)	
dispincome	0.0000327	-0.0000302	0.0000153	-0.0000217	-0.0000215
	(0.0000573)	(0.0000960)	(0.0000136)	(0.0000277)	(0.0000255)

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Diet

Table 8 shows the results of the regressions of using *healthydiet, vegetables* and *fruits* as dependent variables. Based on the output, it cannot be concluded that searching for health information online has a significant impact on a person's diet. Even the sign of the correlation remains unclear. *Intsearch* and *healthydiet* appear to be negatively correlated, meanwhile *fruits* and *intsearch* seem to have a positive correlation, while between *vegetables* and *intsearch* it is mainly negative. When not including the instrument (probit and OLS regressions), searching for health information seems to significantly increase the likelihood of consuming vegetables daily. The probit reduced form shows that the instrument *highbroadband* has a significant (at 1%) negative impact on eating a healthy diet. Meanwhile *highbroadband* does not have a significant impact on fruit or vegetable consumption, although the correlation is positive for fruits and negative for vegetables.

Model specifications including demographic, regional and health state control variables found a positive correlation between *healthydiet* and *intsearch* (see table 7 in the appendix). However, after including other health behaviour as well as socio-economic controls, the coefficient was no longer significant. When including all relevant control variables, the coefficient turned negative. The reduced form shows that *highbroadband* does not have a significant impact on vegetable consumption, and the correlation becomes negative when taking into account health behaviour and socio-economic variables in the probit regressions, and already turns negative when adding health state control variables in the OLS and 2SRI regressions (table 8 in the appendix). In the probit regression, the size of the coefficient greatly decreases when adding more control variables. For the regressions without instruments, the coefficient for *fruits* was positive and significant when health behaviour variables were not controlled for (table 10 in appendix). When including both health and socio-economic status variables, the coefficient became negative.

Table 8

	Probit	Reduced form probit	IV Probit	OLS	Reduced form OLS	2SLS	2SRI
Intsearch on healthy diet N = 69310	-0.0169 (0.0111)	-0.0167 (0.0194)	-0.863 (0.839)	-0.00468 (0.00335)	-1.17e-17*** (1.15e-18)	-0.270 (0.327)	-0.269 (0.312)
Intsearch on vegetables N = 69307	0.0292*** (0.0103)	-0.00345 (0.0179)	-0.180 (0.964)	0.0105*** (0.00375)	-0.00163 (0.00651)	-0.0886 (0.355)	-0.0880 (0.352)
Intsearch on fruits N = 68517	-0.00220 (0.0111)	0.000207 (0.0193)	0.0111 (1.008)	-0.00120 (0.00340)	0.000785 (0.00589)	0.0411 (0.309)	0.0425 (0.318)

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

According to the output of all five regression methods (table 9 below), respondents younger than 25 years old are significantly less likely to eat a healthy diet than those aged 25 to 64 years. Meanwhile, respondents older than 65 are more likely to eat healthy. Female respondents are significantly more likely to eat a healthy than men. Unmarried respondents are significantly less likely to eat a healthy diet and divorced or widowed to eat fruits compared to married ones. Households with children were also significantly less likely to eat healthy compared to households without.

Respondents who have at the most completed middle school are significantly less likely to be eat a healthy diet than those who have only completed high school. Respondents that frequently read books or newspapers and reported having high economic resources are significantly more likely to eat a healthy diet than those who do not. Employed and unemployed are significantly less likely to eat a healthy diet compared to those inactive on the labour market.

Respondents who experience difficulty reaching first aid as well as obese or overweight respondents are significantly more likely to consume fruits and vegetables daily (tables 9 and 11 in the appendix). Frequent alcohol consumers were significantly less likely to eat a healthy diet, but more likely to eat vegetables (tables 9 appendix), than respondents who are not frequent drinkers. Smokers were significantly less likely to eat a healthy diet and vegetables than non-smokers. Salt and weight conscious respondents were significantly more likely to eat a healthy diet than respondents who are not. Respondents who exercise frequently were also significantly more likely to eat a healthy diet and vegetables than those who do not.

Table 9

	Probit	IV Probit	OLS	2SLS	2SRI
Intaccess	-0.0580*	0.0125	-0.0161*	0.00439	0.00432 (0.0260)
	(0.0317)	(0.0812)	(0.00974)	(0.0272)	
2014	0.00364	-0.0292	0.000996	-0.00923	-0.00922 (0.0141)
	(0.0246)	(0.0407)	(0.00718)	(0.0148)	
2015	0.0659*	-0.0170	0.0179	-0.00645	-0.00635 (0.0309)
	(0.0378)	(0.0957)	(0.0111)	(0.0324)	
2016	0.118**	0.0158	0.0330**	0.00392	0.00406 (0.0370)
	(0.0469)	(0.121)	(0.0138)	(0.0388)	
Population density	-0.00210	-0.0000733	-0.000655	-0.0000740	-0.0000823
	(0.00252)	(0.00329)	(0.000763)	(0.00108)	(0.00102)
Youth	-0.301***	-0.328***	-0.0754***	-0.0919***	-0.0918***
	(0.0207)	(0.0198)	(0.00541)	(0.0211)	(0.0202)
Senior	0.291***	0.237***	0.107***	0.0984***	0.0984*** (0.0138)
	(0.0249)	(0.0840)	(0.00885)	(0.0145)	
Female	0.173***	0.209***	0.0514***	0.0674***	0.0674*** (0.0192)
	(0.0119)	(0.0233)	(0.00357)	(0.0201)	
Married	-	-	-	-	-
Unmarried	-0.114***	-0.115***	-0.0351***	-0.0386***	-0.0385***
	(0.0139)	(0.0164)	(0.00425)	(0.00612)	(0.00584)
Divorced or widowed	0.00935	-0.00810	0.00383	-0.00138	-0.00143 (0.00871)
	(0.0183)	(0.0254)	(0.00614)	(0.00905)	
Households with children	-0.0762***	-0.0757***	-0.0242***	-0.0261***	-0.0261***
	(0.0132)	(0.0148)	(0.00423)	(0.00497)	(0.00478)
healthcareuse	0.0109	0.195	0.00323	0.0613	0.0610 (0.0683)
	(0.0142)	(0.183)	(0.00442)	(0.0717)	
Difficulty first aid	0.0125	0.0171	0.00352	0.00528	0.00527 (0.00384)
	(0.0108)	(0.0109)	(0.00325)	(0.00402)	
Overweight or obese	0.0204*	0.0207*	0.00475	0.00538	0.00538 (0.00371)
	(0.0119)	(0.0116)	(0.00363)	(0.00388)	
Not good not bad health state	-0.0201	0.00721	-0.00620	0.00179	0.00174 (0.0105)
	(0.0152)	(0.0324)	(0.00469)	(0.0110)	
Bad health state	0.0177	0.0421	0.00536	0.0135	0.0134 (0.0156)
	(0.0372))	(0.0427)	(0.0124)	(0.0164)	
Health problems	0.0340**	0.0723*	0.0115**	0.0244	0.0244 (0.0159)
	(0.0152)	(0.0378)	(0.00485)	(0.0167)	
Insurance	0.0180	0.0275*	0.00594	0.00939	0.00938 (0.00580)
	(0.0132)	(0.0149)	(0.00414)	(0.00605)	
University	0.0197	0.0464	0.00756	0.0165	0.0165 (0.0115)
	(0.0141)	(0.0283)	(0.00461)	(0.0120)	
Middle school	-0.0745***	-0.129***	-0.0214***	-0.0404*	-0.0403* (0.0226)
	(0.0135)	(0.0489)	(0.00382)	(0.0238)	
Primary or none	-0.0939***	-0.223*	-0.0292***	-0.0720	-0.0717 (0.0512)
	(0.0357)	(0.124)	(0.0104)	(0.0538)	
Booksnewspaper s	0.142***	0.187***	0.0404***	0.0582***	0.0581*** (0.0212)
	(0.0122)	(0.0334)	(0.00347)	(0.0222)	
Employed	-0.0336**	-0.0348**	-0.00986**	-0.0112**	-0.0111**
	(0.0157	(0.0152)	(0.00481)	(0.00527)	(0.00505)
Unemployed	-0.0657***	-0.0518*	-0.0181***	-0.0155**	-0.0156**
	(0.0200)	(0.0276)	(0.00560)	(0.00669)	(0.00636)
High economic resources	0.0340***	0.0330***	0.0105***	0.0111***	0.0110***
	(0.0117)	(0.0119)	(0.00343)	(0.00367)	(0.00350)
Frequent alcohol consumption	-0.105***	-0.0760*	-0.0317***	-0.0253***	-0.0253***
	(0.0133)	(0.0411)	(0.00385)	(0.00885)	(0.00847)
Smoker	-0.114***	-0.0958***	-0.0313***	-0.0287***	-0.0288***
	(0.0135)	(0.0310)	(0.00378)	(0.00504)	(0.00477)
Salt conscious	0.267***	0.287***	0.0740***	0.0872***	0.0872*** (0.0159)
	(0.0125)	(0.0131)	(0.00337)	(0.0167)	
Weightconscious	0.112***	0.156***	0.0303***	0.0471**	0.0471** (0.0203)
	(0.0169)	(0.0371)	(0.00453)	(0.0212)	

Frequentexercise	0.107***	0.119***	0.0329***	0.0396***	0.0396***
	(0.0110))	(0.0108)	(0.00333)	(0.00892)	(0.00852)
dispincome	-0.0000875*	-0.0000110	-0.0000238	-0.00000212	-0.00000224
	(0.0000519)	(0.0000976)	(0.0000154)	(0.0000312)	(0.0000298)

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Weight

The following table (table 10) shows the *intsearch* coefficients from regressions on being overweight or obese as well as only on being obese. The output shows a positive correlation between searching for health information online and being overweight or obese. However, none of the coefficients are significant at conventional levels. The coefficients are very sensitive to the regression method used as they vary from 0.000808 to 0.198. When using *obese* as the dependent variable, the coefficient is negative when not using the instrument. However, it changes sign from negative to positive when using the instrument. This is likely due to the positive correlation between *highbroadband* and *obese*, which can be seen from the coefficient in the reduced forms. Only one coefficient is significant at conventional levels. The coefficients also greatly vary depending on the control variables included (see tables 12 and 13 in appendix).

Table 10

	Probit	**Reduced form probit**	**IV Probit**	**OLS**	**Reduced form OLS**	**2SLS**	**2SRI**
Intsearch for overweight or obese	0.00760	0.00366	0.198	0.00209	0.000808	0.0439	0.0439
N = 69310	(0.0108)	(0.0188)	(1.009)	(0.00355)	(0.00611)	(0.332)	(0.332)
Intsearch for obese	-0.00864	0.0424	1.586***	-0.00152	0.00586	0.341	0.318
N = 64603	(0.0157)	(0.0275)	(0.507)	(0.00216)	(0.00373)	(0.257)	(0.202)

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

When looking at the impact of control variables (table 11 below and table 14 in the appendix for the dependent variable *obese*), all methods show that the respondents are significantly more likely to be obese in 2016 compared to 2013. Respondents are significantly less likely to be overweight or obese as the population density in a region increases. Young respondents are significantly less likely to be overweight or obese than adults, whereas older respondents are significantly more likely to be overweight or obese. Female respondents are significantly less likely to be overweight or obese than men. Unmarried and divorced or widowed respondents are significantly less likely to be overweight or obese compared to married ones. The same applies to households with children compared to households without.

Respondents with a self-rated health state that is “not good, not bad” or “bad or very bad” are significantly more likely to be overweight or obese than those whose health state is “good or very good”. The same applies to respondents who claim to have health problems vs those who do not. Respondents with insurance are also significantly more likely to be overweight or obese than those without insurance. Frequent alcohol consumers were significantly less likely to be obese than respondents who do not and smokers were significantly less likely to be overweight or obese than non-smokers. Respondents who exercise frequently were also significantly less likely to be overweight or obese than those who do not.

Respondents who have completed university education are significantly less likely to be overweight or obese than those who have only completed high school, whereas those who at the most completed middle or primary school are significantly more likely to be overweight or obese. Respondents who reported having high economic resources are also significantly less likely to be overweight or obese than those who reported having low economic resources. Respondents who are employed or unemployed are significantly more likely to be overweight or obese compared to those inactive on the labour market.

Table 11

	Probit	**IV Probit**	**OLS**	**2SLS**	**2SRI**
Intsearch	0.00760 (0.0108)	0.198 (1.009)	0.00209 (0.00355)	0.0439 (0.332)	0.0439 (0.332)
Intaccess	0.0596* (0.0305)	0.0446 (0.0861)	0.0198* (0.0101)	0.0165 (0.0276)	0.0165 (0.0276)
2014	0.0210 (0.0237)	0.0282 (0.0449)	0.00810 (0.00768)	0.00971 (0.0149)	0.00971 (0.0149)
2015	0.0303 (0.0364)	0.0477 (0.0984)	0.0104 (0.0119)	0.0142 (0.0326)	0.0142 (0.0326)
2016	0.0583 (0.0450)	0.0789 (0.117)	0.0193 (0.0148)	0.0239 (0.0391)	0.0239 (0.0391)
Population density	-0.00651*** (0.00243)	-0.00690** (0.00310)	-0.00217*** (0.000798)	-0.00226** (0.00108)	-0.00226** (0.00108)
Youth	-0.599*** (0.0207)	-0.585*** (0.0921)	-0.171*** (0.00584)	-0.168*** (0.0217)	-0.168*** (0.0217)
Senior	0.200*** (0.0252)	0.205*** (0.0354)	0.0765*** (0.00893)	0.0779*** (0.0142)	0.0779*** (0.0142)
Female	-0.688*** (0.0114)	-0.697*** (0.0326)	-0.231*** (0.00363)	-0.234*** (0.0203)	-0.234*** (0.0203)
Unmarried	-0.367*** (0.0133)	-0.363*** (0.0326	-0.128*** (0.00454)	-0.127*** (0.00638)	-0.127*** (0.00637)
Divorced or widowed	-0.104*** (0.0178)	-0.0996*** (0.0302)	-0.0393*** (0.00631)	-0.0385*** (0.00911)	-0.0385*** (0.00910)
Households with children	-0.0699*** (0.0128)	-0.0682*** (0.0165)	-0.0235*** (0.00443)	-0.0232*** (0.00505)	-0.0232*** (0.00504)
Health care use	0.0157 (0.0139)	-0.0260 (0.222)	0.00437 (0.00461)	-0.00479 (0.0729)	-0.00479 (0.0728)

Difficulty first aid	0.0195* (0.0105)	0.0182 (0.0130)	0.00626* (0.00343)	0.00598 (0.00410)	0.00598 (0.00409
Not good not bad health state	0.173*** (0.0145)	0.166*** (0.0409)	0.0609*** (0.00507)	0.0596*** (0.0113)	0.0596*** (0.0113)
Bad health state	0.135*** (0.0372)	0.128** (0.0525)	0.0491*** (0.0133)	0.0478*** (0.0168)	0.0478*** (0.0168)
Health problems	0.178*** (0.0148)	0.168*** (0.0589)	0.0619*** (0.00515)	0.0599*** (0.0170)	0.0599*** (0.0170)
Insurance	0.0463*** (0.0128)	0.0436** (0.0199)	0.0165*** (0.00433)	0.0160*** (0.00614)	0.0160*** (0.00614)
University education	-0.210*** (0.0141)	-0.216*** (0.0283)	-0.0707*** (0.00461)	-0.0721*** (0.0120)	-0.0721*** (0.0120)
Middle school education	0.114*** (0.0128)	0.127* (0.0687)	0.0420*** (0.00420)	0.0450* (0.0242)	0.0450* (0.0242)
Primary school education or none	0.297*** (0.0336)	0.326** (0.153)	0.106*** (0.0114)	0.113** (0.0548)	0.113** (0.0547)
Books or newspapers	-0.0401*** (0.0116)	-0.0528 (0.0672)	-0.0113*** (0.00383)	-0.0142 (0.0227)	-0.0142 (0.0227)
Employed	0.0509*** (0.0154)	0.0517*** (0.0156)	0.0171*** (0.00496)	0.0173*** (0.00522)	0.0173*** (0.00521)
Unemployed	0.0781*** (0.0194)	0.0759*** (0.0234)	0.0228*** (0.00609)	0.0224*** (0.00688)	0.0224*** (0.00687)
High economics resources	-0.0421*** (0.0113)	-0.0423*** (0.0113)	-0.0130*** (0.00368)	-0.0131*** (0.00376)	-0.0131*** (0.00376)
Frequent alcohol consumption	0.0236* (0.0124)	0.0190 (0.0281)	0.0133*** (0.00428)	0.0123 (0.00901)	0.0123 (0.00900)
Smoker	-0.0753*** (0.0127)	-0.0767*** (0.0141)	-0.0260*** (0.00423)	-0.0264*** (0.00524)	-0.0264*** (0.00523)
Salt conscious	0.0688*** (0.0118)	0.0590 (0.0551)	0.0222*** (0.00376)	0.0201 (0.0171)	0.0201 (0.0171)
Weight conscious	0.0546*** (0.0157)	0.0423 (0.0682)	0.0192*** (0.00514)	0.0165 (0.0217)	0.0165 (0.0217)
Healthy diet	0.0169 (0.0122)	0.0180 (0.0133)	0.00530 (0.00406)	0.00556 (0.00452)	0.00556 (0.00452)
Frequent exercise	-0.171*** (0.0108)	-0.175*** (0.0209)	-0.0576*** (0.00350)	-0.0587*** (0.00908)	-0.0587*** (0.00907)
Disposable income	-0.000000527 (0.0000501)	-0.0000161 (0.0000964)	9.01e-08 (0.0000164)	- 0.00000332 (0.0000316)	-0.00000332 (0.0000316)

6. Discussion & Conclusion

This section will draw conclusions from the results presented in the previous section and provide some potential explanations for the findings. The research questions will be answered and some of the limitations of the analysis are discussed. Based on the regression results, searching for health online does not seem to be significantly impacting peoples' health behaviour.

For the dependent variable *healthcareuse,* some of the regression methods showed a significant positive correlation. The answer to the second research question "What are the consequences of searching for health information online on an individual's use of health care services?" is therefore that searching for health information online is positively correlated with a person's use of health care services. The first hypothesis, "searching for health information online increases health care use" is not rejected. Based on this, it can be suggested in line with previous literature (or example, Suziedelyte, 2012 and Farajallah et al., 2015) that health information is not a substitute, but a complement to health care use. However, this finding goes against the conclusions taken by Dwyer and Liu (2013) and Schmid (2015).

The answer to the third research question, "What are the consequences of searching for health information online on an individual's diet?" is that searching for health information online was not found to have any consequences on a person's diet. The coefficients were not significant and changed sign between the different regression methods. Therefore, the second hypothesis, "searching for health information online has a positive impact on a person's diet" is rejected. As expected, having a socio-economic status was positively correlated with eating a diet high in fruits and vegetables. Higher educated people may be more likely to have a diet in line with dietary guidelines than those with a lower level of education (Roos et al., 1998, Ryden & Hagfors, 2011). Additionally, a perception of a high cost of eating a diet high in fruits and vegetables can deter low income households. This is in line with many other studies from various fields (Dittus et al., 1995, Tiffin & Arnoult, 2010, Rehm et al. 2011, Kamphuis et al., 2006, Giskes et al., 2010).

Another important finding was that younger respondents (aged 25 years or less) were significantly less likely to eat a healthy diet than the adult population, meanwhile the elderly were more likely. This is in line with findings from the 2019 WHO European Childhood Obesity Surveillance Initiative (COSI) survey for Italy which found that only 1 in 4 children consumed fruits and vegetables every day (WHO, 2019). Already in 2006, Larson et al. found that the young were severely lacking cooking skills, which has been argued to be crucial for eating a healthy diet (Hartmann et al., 2013). The elderly also have more leisure time, which may imply more time to cook and to think about their food choices.

The answer to the fourth research question is that e-health information searches also do not appear to have a significant impact on an individual's weight. However, the output showed a positive correlation between searching for health information online and being overweight or obese. This may come from the sedentary lifestyle of internet users (DiNardi et al., 2019). The sign of the correlation was not clear for only obese respondents. Therefore, the third hypothesis, "searching for health information online reduces obesity" is rejected. The output showed that females were significantly less likely to be obese compared to men. This is in line with three earlier Italian studies in the field of medicine by Micciolo et al. (2010), Cavaliere et al. (2014) and Osella et al. (2014).

The analysis does however shed light over who searches for health information online. A surprising finding was that over the study years, less individuals searched for health information online. The answer to the first research question "What are the determinants of searching for health information online?" is as follows. Having health problems and a perception of being in poor health was positively and significantly correlated with e-health information seeking. Individuals who take other actions towards their health, such as watching their intake of salt and looking at their weight as well as exercising frequently were more likely to search for health information online. However, unexpectedly, the same was found for frequent alcohol drinkers and smokers.

Females were significantly more likely to have searched for information online than men. This was also found to be the case previously in Italy, as Siquilini et al (2011) found that 61.6% of female users and 50.2% of male users had used online health information resources. In spite of their higher internet use and skills, young individuals (aged 25 or less) were significantly less likely to search for information than 26 to 64 year old respondents. Additionally, no significant correlation between having children and *intsearch* was found. As expected, education played an important role. Respondents who had completed university were significantly more likely to have searched for e-health information than respondents who competed high school, while the opposite held for those who had a lower than high school education. Respondents who read books or newspapers frequently were significantly more likely to have searched for health information online. However, individual employment and economic resources were not found to have a significant impact on health information seeking. At the regional level, higher disposable income and population density were positively correlated with internet searching.

The limitations of the analysis will now be discussed. The analysis suffers from some internal validity concerns. The instrument used, a dummy variable for a higher than the median percentage of households with broadband access, is not flawless. It is possible that it is not a very strong instrument, as it did not pass the Staiger & Stock rule of thumb and the magnitude of the coefficient was very small. However, it proved very hard to find an instrument which could be considered both relevant and exogenous. Additionally, it may be reason for concern that health care use was found to have a significant on searching for health information online. However, this reverse causality should be taken care of by using the instrument.

The dependent variables also suffer from some drawbacks. Firstly, health care use does not include primary or specialised health care visits. Secondly, the dependent variables for being overweight and obese are based on self-reported height and weight. There may be reporting errors and they are usually biased, with height overestimated and weight underestimated. As pointed out by Baum & Ruhm (2009), these reporting inaccuracies could differ with socio-economic status.

The data set did not include information on frequency of the online search, or about the cause or content of the search. It would have been interesting to know which specific health issue(s) the individual searched the internet for. Frequency of searching has been found to have an effect in previous studies (Suziedelyte, 2012). Also, additional important control variables could have been included, such as whether the individual was pregnant, for example.

Another limitation is that the survey data used is cross-sectional, and not panel data, meaning that it does not follow the same sample of respondents over time. Perhaps had the data been longitudinal, some significant long-term effects changes in health behaviour could have been found which did not show in this short-term analysis.

There also some external validity concerns. No survey weights were used after excluding observations who had missing variables for *intsearch*. Therefore, while the original sample can be considered representative of the Italian population, this may not hold for the final sample size used in the analysis. The latest survey year included in the analysis was 2016. Searching for health information online may have changed since, especially since the COVID-19 pandemic.

To conclude, the analysis did not find searching for health information online to have a considerable impact on individuals' health behaviour, although it was positively correlated with emergency health care use. However, it is important to keep in mind that a considerable proportion of the sample (43% after excluding missing variabes) did nevertheless search for health information online, and that those in poor health were more likely to use e-health.

References

Anekwe, T. D., Rahkovsky, I. (2013). Economic Costs and Benefits of Healthy Eating. *Current Obesity Reports*, 2, 225–234

Baum, C., L., Ruhm, C. J. (2009). Age, socioeconomic status and obesity growth. *Journal of Health Economics*, 28, 3, 635-648

Baum, C. F. (2012). SSPECIALREG: Stata module to estimate binary choice model with discrete endogenous regressor via special regressor method. Statistical Software Components S457546, Boston College Department of Economics

Besley, T. (2001). The demand for health care and health insurance. *Oxford Review of Economic Policy*, 5, 21–33

Bujnowska-Fedak, M.M., Wegierek, P. (2020). The Impact of Online Health Information on Patient Health Behaviours and Making Decisions Concerning Health. *International Journal of Environmental Research and Public Health*, 17, 880

Carrera, M., Hasan, S. A., Prina, S. (2020). Do health risk assessments change eating habits at the workplace? *Journal of Economic Behaviour & Organization*, 172, 236-246

Carrieri, V., Madio, L., Principe, F. (2019).Vaccine hesitancy and (fake) news: Quasi-experimental evidence from Italy. *Health Economics*, 28, 11, 1377-1382

Carrieri, V., Principe, F. (2020).Who and for How Long? An Empirical Analysis of the Consumers' Response to Red Meat Warning. IZA Discussion Paper No. 13882, Available at SSRN: https://ssrn-com.eur.idm.oclc.org/abstract=3734750

Cauley, S. D. (1987) The Time Price of Medical Care. *The Review of Economics and Statistics,* 69, 1, 59-66

Cavaliere, A., De Marchi, E., Banterle, A. (2014). Healthy–unhealthy weight and time preference. Is there an association? An analysis through a consumer survey. *Appetite,* 83, 135-143

Cawley, J., Meyerhoefer, C. (2012). The medical care costs of obesity: an instrumental variables approach. J Health Econ, 31, 219–230

Cawley, J., Ruhm, C.J. (2012) The Economics of Risky Behaviours. *Handbook of Health Economics,* 2

Cliff, B. Q., Townsend, T., Wolfson, J. A. (2019). Examining Household Changes in Produce Purchases Among New Parents. *Journal of Nutrition Education and Behaviour*, 51, 7, 798-805

Craig, J., Moul, C.C., Niemesh, G.T. (2020). Do menu-labelling laws translate into results? The disparate impacts on population obesity and diabetes, *Applied Economics*, 52, 14, 1592-1605

Cropper, M., L. (1977). Health investment in health, and occupational choice. *Journal of Political Economy*, 85,6, 1273-1294

Cunningham, S. (2021). *Causal Inference: The Mixtape.* Yale University Press

Cutler, D.M., Adriana, L.-M., (2006). Education and Health: Evaluating Theories and Evidence. National Bureau of Economic Research Working Paper No. 12352

Devaux M, Sassi F. (2015). The Labour Market Impacts of Obesity, Smoking, Alcohol Use and Related Chronic Diseases. OECD Health Working Paper No. 86. Doi: https://doi.org/10.1787/5jrqcn5fpv0v-en

DiNardi, M., Guldi, M. & Simon, D. (2019). Body weight and Internet access: evidence from the rollout of broadband providers. *Journal of Population Economics,* 32, 877–913 https://link-springer-com.eur.idm.oclc.org/article/10.1007/s00148-018-0709-9#citeas

Dittus, K.L., Hillers, V.N., Beerman, K.A. (1995). Benefits and Barriers to Fruit and Vegetable Intake: Relationship between Attitudes and Consumption. *Journal of Nutrition Education*, 27, 3

Dwyer, D., S., Liu, H. (2013). The impact of consumer health information on the demand for health services. *The Quarterly Review of Economics and Finance,* 53, 1, 1-11,

Eysenbach, G., Powell, J., Kuss, O., Sa, E. (2002) Empirical Studies Assessing the Quality of Health Information for Consumers on the World Wide Web. *Journal of the American Medical Association*, 287, 20

Farajallah, M., Le Goff-Pronost, M., Pénard, T. & Suire, R. (2015). Quoi de neuf docteur ? Une étude économétrique sur la recherche en ligne d'informations médicales par les patients, *Journal de gestion et d'économie médicales*, 4-5, 33, 231-251

Ferré, F., de Belvis, A. G., Valerio, L., Longhi, S., Lazzari, A., Fattore, G., Ricciardi, W., Maresso, A. (2014). Italy: Health System Review. Health Systems in Transition, 16, 4, 1–168.

Ferguson, T., Frydman, G. (2004). The first generation of e-patients: These new medical colleagues could provide sustainable healthcare solutions, *British Medical Journal*, 328, 7449, 1148-1149

Folland, S., Goodman, A.C., & Stano, M. (2017). *The Economics of Health and Health Care*. Oxon & New York: Routledge

Fontaine, K.R., Redden, D.T., Wang C., Westfall, A.O., Allison, D.B. (2003). Years of Life Lost Due to Obesity. JAMA, 289, 2, 187–193

Frank, R.G (2004) Behavioral Economics and Health Economics. *Working Paper 10881*

Frazão, E. (1999). High costs of poor eating patterns in the United States. Chapter 1 in America's Eating Habits: Changes and Consequences. *USDA Economic Research Service*

Fruhbeck, G., Toplak, H., Woodward E., Yumuk, V., Maislos, M., Oppert, J. (2013). Obesity: the gateway to ill health—an EASO position statement on a rising public health, clinical and scientific challenge in Europe. *Obesity Facts*, 6, 117–20

Gander, K. (2014, January 23). Pope Francis: The internet is a 'gift from God'. The Independent. Retrieved from: https://www.independent.co.uk/news/world/politics/pope-francis-calls-the-internet-a-gift-from-god-9080117.html

Giskes, K et al. (2010). A systematic review of studies on socioeconomic inequalities in dietary intakes associated with weight gain and overweight/obesity conducted among European adults. *Obesityß Reviews.*, 11, 6, 413–29

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *The Journal of Political Economy*, 80, 2, 223-255.

Gupta, R. & Wood, D.A. (2019) Primary prevention of Ischaemic heart disease: populations, individuals, and health professionals. *The Lancet*, 394, 10199.

Hay, J. & Leahy, M.J. (1982). Physician Induced Demand – An Empirical Analysis of the Consumer Information Gap. *Journal of Health Economics* 1, 231-244

Hartmann C, Dohle S, Siegrist M. (2013). Importance of cooking skills for balanced food choices. *Appetite*, 65, 125-131

Hsieh, C. & Lin, S. (1997). Health Information and the Demand for Preventive Care among the Elderly in Taiwan. *The Journal of Human Resources*, 32, 2, 308-333

Iverson, S. A., Howard, K. B., Penney, B. K. (2008) Impact of Internet Use on Health-Related Behaviors and the Patient-Physician Relationship: A Survey-Based Study and Review. *Journal of the American Ostheopathic Association*, 108, 12

Instituto Nazionale di Statistica (2019). Informazioni sulla Rilevazione: Indagine Multiscopo sulle Famiglie: Aspetti Della Vita Quotidiana. Retrieved from: https://www.istat.it/it/archivio/91926

Ippolito, P.M & Mathios, A.D. (1990). Information, Advertising and Health Choice: a Study on the Cereal Market. *Rand Journal of Economics*, 21,3, 459-480

Jones, A. M., Koolman, X., van Doorslaer, E. (2006). The Impact of Having Supplementary Private Health Insurance on the Use of Specialists. *Annales d'Économie et de Statistique*, 83/84, 251-275

Jurges, H., Reinhold, S., Salm, M. (2011). Does schooling affect health behavior? Evidence from the educational expansion in Western Germany. *Economics of Education Review*, 30,5, 862-872

Kamphuis, C et al. (2006). Environmental determinants of fruit and vegetable consumption among adults: a systematic review. *British Journal of Nutrition*, 96, 4, 620–635

Kelly, L., Jenkinson, C., Ziebland, S. (2013). Measuring the effects of online health information for patients: Item generation for an e-health impact questionnaire. *Patient Education and Counseling*, 93, 433–438

Kenkel, D. (1990) Consumer Health Information and the Demand for Medical Care. *The Review of Economics and Statistics,* 72, 4, 587-595

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112, 2, 443-477.

Larson NI et al. (2006). Food preparation by young adults is associated with better diet quality. *Journal of the Academy of Nutrition and Dietetics*, 106, 12, 2001–2007

Laugesen, J., Hassanein, K., Yuan, Y. (2015). The Impact of Internet Health Information on Patient Compliance: A Research Model and an Empirical Study. *Journal of Medical Internet Research*, 17, 6

Lee, C-J. (2008). Does the Internet Displace Health Professionals? *Journal of Health Communication, International Perspectives*, 13, 5, 450-464

Lee, S.T., Lin, J. (2016) A Self-Determination Perspective on Online Health Information Seeking: The Internet vs. Face-to-Face Office Visits With Physicians, *Journal of Health Communication*, 21, 6

Loewenstein, G. (1996). Out of Control: Visceral Influences on Behavior. *Organizational Behavior and Human Decision Processes*, 65, 3, 272-292

Loos, A. (2013). Cyberchondria: too much information for the health anxious patient? *Journal of Consumer Health on the Internet*, 17, 4, 439-445

McCully, S. N., Don, B., P., Updegraff, J., A. (2013). Using the Internet to Help With Diet, Weight, and Physical Activity: Results From the Health Information National Trends Survey (HINTS). *Journal of Medical Internet Research*, 15,8 doi: 10.2196/jmir.2612

Menon et al. (2020). Cyberchondria: conceptual relation with health anxiety, assessment, management and prevention, *Asian Journal of Psychiatry*, 53

Micciolo, R., Di Francesco, V., Fantin, F., Canal, L., Harris, T. B., Bosello, O., Zamboni, M. (2010). Prevalence of Overweight and Obesity in Italy (2001–2008): Is There a Rising Obesity Epidemic? *Annals of Epidemiology*, 20, 4, 258-264

Migliore, E., Pagano, E., Mirabelli, D., et al. (2013). Hospitalization rates and cost in severe or complicated obesity: an Italian cohort study. *BMC Public Health*, 13, 544

Mitsutake, S., Shibata, A., Ishii, K., Oka, K. (2016). Associations of eHealth Literacy With Health Behavior Among Adult Internet Users. *Journal of Medical Internet Research*, 18, 7

Newey, W. K. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics*, 36, 231-250.

O'Donoghue, T., Rabin, M. (1999). Doing It Now or Later. *American Economic Review*, 89, 1,103-124.

OECD & **EOHSP** (2019). Italy: Country Health Profile 2019, State of Health in the EU, OECD Publishing, Paris & European Observatory on Health Systems and Policies, Brussels. Retrieved from: https://www.oecd-ilibrary.org/docserver/cef1e5cb-en.pdf?expires=1591618671&id=id&accname=guest&checksum=1825112EC3A94B69C6956C230A95DBCE

OECD (2021). Obesity and the Economics of Prevention: Fit not Fat - Italy Key Facts. OECD. Accessed 30/04/2021 at: https://www.oecd.org/els/health-systems/obesityandtheeconomicsofpreventionfitnotfat-italykeyfacts.htm#Further_Reading

Osella, A. R., Díaz, M del P., Cozzolongo, R., Bonfiglio, C., Franco, I., Abrescia, D.I., Bianco, A., Giampiero, E.S., Petruzzi, J., Elsa, L., Mario, C., Mastrosimni, A.M., Giocchino, L. (2014). Overweight and Obesity in Southern Italy: their association with social and life-style characteristics and their effect on levels of biologic markers. *Revista Facultad de Ciencias Médicas Universidad Nacional de Córdoba*, 71, 3, 113-124

Oster, G., Edelsberg, J., O'Sullivan, A.K., Thompson, D. (2000). The clinical and economic burden of obesity in a managed care setting. *American Journal of Managing Care*, 6, 681–9

Oster, E. (2018). Diabetes and diet: Purchasing behavior change in response to health information. *American Economic Journal: Applied Economics,* 10, 4, 308-348

Pew (2013). Health online 2013. Retrieved from: https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf

Phelps, C. (1978). Illness Prevention and Medical Insurance. *Journal of Human Resources*, 13,2, 183-207

Popkin, B.M., Kim, S., Rusev, E.R., Du, S., Zizza, C. (2006). Measuring the full economic costs of diet, physical activity and obesity-related chronic diseases. Obesity Reviews, 7,3, 271-293

Propper C. (2000). The demand for private health care in the UK. *Journal of Health Economics,* 19, 855–876.

Rehm, C. D., Monsivais, P., Drewnowski, A. (2011). The quality and monetary value of diets consumed by adults in the United States. *American Journal of Clinical Nutrition*, 94, 5, 1333–1339

Rice, R. (2006) Influences, usage, and outcomes of Internet health information searching: Multivariate results from the Pew surveys. *International Journal of Medical Informatics*, 75, 8-28

Roos, E., Lahelma, E., Virtanen, M., Prättälä, R. & Pietinen, P. (1998). Gender, Socioeconomic Status and Family Status as Determinants of Food Behaviour. *Social Science & Medicine,* 46, 12, 1519-1529

Ryden, P. J., & Hagfors, L. (2011). Diet cost, diet quality and socio-economic position: how are they related and what contributes to differences in diet costs? *Public Health Nutrition*, 14, 9, 1680 - 1692

Schmid, C. (2015). Consumer Health Information and the Demand for Physician Visits. *Health Economics*, 24, 12, 1619-1631

Schultz, T.W. (1975). The Value of the Ability to Deal with Disequilibria. *Journal of Economic Literature*, 13,3, 827-846

Sindelar, J. (1982). Differential Use of Medical Care by Sex, *Journal of Political Economy* , 90, 5, 1003-1019.

Siquilini, R., Ceruti M., Lovato, E. et al. (2011) Surfing the internet for health information: an Italian survey on use and population choices. *BMC Medical Informatics and Decision Making*, 11, 21. Doi: https://doi.org/10.1186/1472-6947-11-21

Specchia, M.L., Veneziano, M.A., Cadeddu, C., Ferriero, A.M., Mancuso, A., Ianuale, C., Parente, P., Capri, S., Ricciardi, W. (2015). Economic impact of adult obesity on health systems: a systematic review. European Journal of Public Health, 25, 2, 255-62

Staiger, D. & Stock, J.H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65, 3, 557-586.

Starcevic, V., Baggio, S., Berle, D., Khazaal, Y., K. Viswasam, K. (2019). Cyberchondria and its relationships with related constructs: a network analysis, *Psychiatr. Q.,* 90, 3, 491-505

Stranieri, S., Baldi, L., Banterle, A. (2010). Do Nutrition Claims Matter to Consumers? An Empirical Analysis Considering European Requirements. *Journal of Agricultural Economics*, 61, 1, 15-33

Sunstein, C., and Thaler, R. (2003). Libertarian Paternalism Is Not an Oxymoron. *University of Chicago Law Review*, 70, 4, 1159-202.

Suziedelyte, A. (2012). How does searching for health information on the Internet affect individuals' demand for health care services? *Social Science & Medicine*, 75, 1828-1835

Terza, J. V., Basu, A., Rathouz, P.J. (2008). Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics*, 27, 3, 531-543

Tiffin, R., Arnoult, M. (2010). The demand for a healthy diet: estimating the almost ideal demand system with infrequency of purchase. *European Review of Agricultural Economics*, 37, 4, 501–521.

Tustin, R. (2010). The Role of Patient Satisfaction in Online Health Information Seeking. *Journal of Health Communication*, 15, 1, 3-17

Umberson, D. (1987). Family Status and Health Behaviors: Social Control as a Dimension of Social Integration. *Journal of Health and Social Behavior*, 28, 3, 306-319

Variyam, J. N. (2007). Do nutrition labels improve dietary outcomes? *Health Economics*, 17, 6, 695-708

Vistnes, J.P., Hamilton, V. (1995). The Time and Monetary Costs of Outpatient Care for Children. *The American Economic Review*, 85, 2

Vitale, M., Giosuè, A., Vaccaro, O., & Riccardi, G. (2021). Recent Trends in Dietary Habits of the Italian Population: Potential Impact on Health and the Environment. *Nutrients,* 13, 2, 476

Wagner, T.H., Hu, T., Hibbard, J.H. (2001). The demand for consumer health information. *Journal of Health Economics, 20,* 1059–1075

Wagner, T.H., Jimison, H. B. (2003). Computerized Health Information and the Demand for Medical Care. *Value in Health*, 6,1, 29-39

Wangberg, S., Sorensen, T., Andreassen, H.K. (2015). Using the Internet to Support Exercise and Diet: A Stratified Norwegian Survey. *Medicine 2.0, 4,2*

WHO (9th December 2019). Italy: over 20% of children are overweight, says new report. World Health Organisation, Regional Office for Europe. Accessed 30/04/2021 at: https://www.euro.who.int/en/countries/italy/news/news/2020/12/italy-over-20-of-children-are-overweight,-says-new-report

Woolridge, J. M. (2016). *Introductory Econometrics*, 6e. Cengage Learning, Boston

Zillien, H., Hargittai, E. (2009). Digital Distinction: Status-Specific Types of Internet Usage. *Social Science Quarterly*, 90, 2, 274-291

Appendix

Table 1 – *intsearch* descriptive statistics of characteristics

	% of respondents who searched for health information online
Internet access	.4362253
No internet access	.3401674
Difficulty reaching first aid	.428222
No difficulty reaching first aid	.4392411
Youth	.2879897
Adult	.4840096
Senior	.490428
Female	.4719644
Male	.3979742
University	.5490301
High school	.4772594
Middle school	.3735519
Primary school or none	.1963688
Employed	.4791471
Unemployed	.4413164
Inactive	.4199991
Children	.4151555
No children	.488759
Health problems	.5375306
No health problems	.411334
Good self-reported health	.4146647
Bad self-reported health	.5638298
High self-reported economic resources	.4478554
Low self-reported economic resources	.4104362
Health or accident insurance	.4867495
No insurance	.4189651
Reads books and /or newspapers frequently	.477984
Does not read books or newspapers frequently	.3549378
Overweight or obese	.4495986
Not overweight or obese	.4244764
Health care use	.6476128
No health care use	.3891544
Exercises frequently	.4380355
Does not exercise frequently	.4291539
Eats a healthy diet	.4764431
Does not eat a healthy diet	.4205762
Frequently consumes alcohol	.4755776
Does not consume alcohol frequently	.4357275
Smoker	.4508446
Not a smoker	.4433736
Salt conscious	.4668347
Does not pay attention to their salt consumption	.3639468
Weight conscious	.446117

Does not pay attention to their weight	.3544157

Table 2 - *healthcareuse* descriptive statistics of characteristics

	% of respondents that used health care
Survey year 2013	.0908998
Survey year 2014	.1869922
Survey year 2015	.1953053
Survey year 2016	.1639689
Female	.1805595
Male	.1620608
Difficulty reaching first aid	.1660849
No difficulty reaching first aid	.1762625
Youth	.1023701
Adult	.193405
Senior	.2141166
University	.2193613
High school	.1816985
Middle school	.1466612
Primary school or none	.1019725
High self-reported economic resources	.1760123
Low self-reported economic resources	.1626274
Health or accident insurance	.2131205
No insurance	.1599447
Health problems	.2577688
No health problems	.1525283
Good self-reported health	.1515087
Bad self-reported health	.4097099
Overweight or obese	.1853819
Not overweight or obese	.1629682

Table 3 – *healthydiet* descriptive statistics of characteristics

	% of respondents who had a healthy diet
Survey year 2013	.2234552
Survey year 2014	.2273047
Survey year 2015	.234273
Survey year 2016	.2427493
Youth	.1236715
Adult	.2606754
Senior	.4081413
Female	.2670791
Male	.2027473
University	.2982468
High school	.2540083
Middle school	.1995687
Primary school or none	.1193386
Health problems	.287395
No health problems	.22176
Good self-reported health	.2248368
Bad self-reported health	.2988166

High self-reported economic resources	.2490635
Low self-reported economic resources	.2082186
Overweight or obese	.2505157
Not overweight or obese	.2242203
Exercises frequently	.2427062
Does not exercise frequently	.2251529
Salt conscious	.2693439
Does not pay attention to their salt consumption	.155033
Weight conscious	.240662
Does not pay attention to their weight	.1835217

Table 4 - *overweightorobese* descriptive statistics of characteristics

	% of respondents who were overweight or obese
Total	.3544081
Survey year 2013	.3533476
Survey year 2014	.3313425
Survey year 2015	.3599232
Survey year 2016	.3710308
Youth	.1584728
Adult	.4096414
Senior	.5855023
Female	.2395109
Male	.4598396
University	.3147862
High school	.3759131
Middle school	.3813819
Primary school or none	.2627807
Employed	.4073618
Unemployed	.3580492
Inactive	.3050571
Children	.3269651
No children	.4377998
Health problems	.4812872
No health problems	.3266367
Good self-reported health	.3238652
Bad self-reported health	.5124113
High self-reported economic resources	.3509474
Low self-reported economic resources	.3599677
Health care use	.3844092
No health care use	.3482183
Exercises frequently	.3032415
Does not exercise frequently	.401922
Eats a healthy diet	.3801464
Does not eat a healthy diet	.3465356
Frequently consumes alcohol	.4653465
Does not consume alcohol frequently	.3282752
Smoker	.3794544
Not a smoker	.3553634
Salt conscious	.3721784

Does not pay attention to their salt consumption	.316025
Weight conscious	.3553247
Does not pay attention to their weight	.349113

Table 5- descriptive statistics comparison between sample sizes

	Total sample N = 150,000 something	**Restricted sample answering the question about searching for health information online N = 80,839**	**Sample characteristics for those whose answer was missing N = 72,974**
Internet access	.7498566	.9704347	.5042832
Difficulty reaching first aid	.5532656	.5138842	.5969953
Youth	.2330882	.2601715	.203086
Adult	.5357089	.6823192	.3732973
Senior	.2312028	.0575094	.4236166
Female	.5170174	.478519	.5596651
University	.1180589	.1843046	.0360065
High school	.3260053	.4419773	.1823614
Middle school	.2921255	.2701543	.3193393
Primary school or none	.2638103	.1035639	.4622928
Employed	.4012026	.558524	.2070657
Unemployed	.1144624	.1359093	.0879967
Inactive	.484335	.3055668	.7049377
Children	.6457907	.7523967	.5276948
Health problems	.2581218	.180258	.3465853
Good self-reported health	.6883618	.8104751	.5530874
Bad self-reported health	.0617243	.0209305	.1069148
Self-reported high economic resources	.5636598	.615045	.5065769
Self-reported low economic resources	.4363402	.384955	.4934231
Insurance	.1538209	.2131068	.0880486
Reads books and /or newspapers frequently	.5216452	.6408151	.370865
Overweight or obese	.4299511	.3544081	.5235191
Health care use	.1617808	.1709117	.1515714
Exercises frequently	.3913383	.4802439	.2874278
Eats a healthy diet	.270109	.233539	.3135308
Frequently consumes alcohol	.2552129	.2346109	.2819675
Smoker	.1895497	.2168785	.1540788
Salt conscious	.7088493	.6824453	.7403329
Weight conscious	.8364882	.8696919	.7989664

Table 6 - Health care use – coefficients for *intsearch* using different model specifications

	Probit	**Reduced form Probit**	**IV Probit**	**OLS**	**Reduced form OLS**	**2SLS**	**2SRI**
Baseline model (i)	0.595***	0.0322*	1.397**	0.151***	0.0115**	0.483**	0.483**
N = 80638	(0.0109)	(0.0193)	(0.554)	(0.00275)	(0.00475)	(0.216)	(0.195)

Incl. intaccess (ii)	0.594***	0.0318*	1.401**	0.151***	0.0114**	0.485**	0.485**
N = 80637	(0.0109)	(0.0193)	(0.561)	(0.00275)	(0.00475)	(0.220)	(0.199)
Incl.demographics	0.550***	0.0347*	1.432***	0.141***	0.0118**	0.484**	0.485**
(iii)	(0.0111)	(0.0194)	(0.522)	(0.00277)	(0.00472)	(0.212)	(0.192)
N = 80637							
Incl. health status	0.532***	0.0247	1.224*	0.135***	0.00918*	0.406*	0.403*
& insurance &	(0.0113)	(0.0199)	(0.683)	(0.00281)	(0.00479)	(0.223)	(0.207)
difficulty first aid							
(iv)							
N = 77906							
Incl. health	0.522***	0.0250	1.249*	0.133***	0.00960*	0.429*	0.427**
behaviour (v)	(0.0117)	(0.0205)	(0.692)	(0.00288)	(0.00497	(0.237)	(0.218)
N = 72834							
Incl. all SES (vi)	0.542***	0.0328	1.448**	0.140***	0.0114**	0.500**	0.497**
N = 72811	(0.0115)	(0.0203)	(0.565	0.00288)	(0.00506)	(0.244)	(0.217)
Incl. all variables	0.515***	0.0217	1.214	0.131***	0.00885*	0.435	0.432*
N = 69310	(0.0119)	(0.0210)	(0.798)	(0.00295)	(0.00514)	(0.271)	(0.247)

i: Model includes variables *year* (separate dummies*), regions (*separate dummies), *dispincome, popdensity*
ii: Add *intaccess* to baseline variables
iii: Add *intaccess, youth, senior, female, unmarried, divorcedwidow, children* to baseline variables
iv: Add *difficultyfirstaid, notgoodnotbadselfhealth, badselfhealth, healthproblems, overweightobese, insurance* to model iii
v: Add *frequentalcohol, smoker, saltconscious, weightconscious, healthydiet, frequentexercise* to model iv
vi: Add *university, middleschool, primarynone, books, employed, unemployed, higheconresources, insurance* to model iii

Table 7 - Healthy diet model specifications

	Probit	**Reduced form Probit**	**IV Probit**	**OLS**	**Reduced form OLS**	**2SLS**	**2SRI**
Baseline model (i)	0.135***	-0.0126	-0.488	0.0413***	-0.00300	-0.126	-0.125
N = 80584	(0.00988)	(0.0177)	(0.672)	(0.00304)	(0.00538)	(0.229)	(0.225)
Incl. intaccess (ii)	0.136***	-0.0122	-0.481	0.0416***	-0.00289	-0.123	-0.123
N = 80583	(0.00988)	(0.0177)	(0.686)	(0.00305)	(0.00538)	(0.233)	(0.229)
Incl.demographics	0.0458***	-0.00980	-0.390	0.0135***	-0.00247	-0.102	-0.102
(iii)	(0.0102)	(0.0180)	(0.708)	(0.00307)	(0.00528)	(0.220)	(0.217)
N = 80583							
Incl. health status,	0.0408***	-0.0182	-0.840	0.0120***	-0.00495	-0.255	-0.254
insurance (iv)	(0.0106)	(0.0184)	(0.716)	(0.00317)	(0.00537)	(0.275)	(0.262)
N = 77788							
Incl. health	0.00950	-0.0194	-0.898	0.00330	2.69e-17***	-0.279	-0.278
behaviour (v)	(0.0109)	(0.0190)	(0.727)	(0.00325)	(1.77e-18)	(0.290)	(0.274)
N = 72834							
Incl. all SES (vi)	0.00666	-0.00904	-0.386	0.00218	-0.00244	-0.107	-0.106
N = 72784	(0.0106)	(0.0188)	(0.782)	(0.00323)	(0.00568)	(0.250)	(0.247)
Incl. all variables	-0.0169	-0.0167	-0.863	-0.00468	-1.17e-17***	-0.270	-0.269
N = 69310	(0.0111)	(0.0194)	(0.839)	(0.00335)	(1.15e-18)	(0.327)	(0.312)

i: Model includes variables *year* (separate dummies*), regions (*separate dummies), *dispincome, popdensity*
ii: Add *intaccess* to baseline variables
iii: Add *intaccess, youth, senior, female, unmarried, divorcedwidow, children* to baseline variables
iv: Add *healthcareuse, difficultyfirstaid, notgoodnotbadselfhealth, badselfhealth, healthproblems, overweightobese, insurance* to model iii
v: Add *frequentalcohol, smoker, saltconscious, weightconscious, frequentexercise* to model iv
vi: Add *university, middleschool, primarynone, books, employed, unemployed, higheconresources, insurance* to model iii

Table 8 – Vegetables model specifications

	Probit	**Reduced form Probit**	**IV Probit**	**OLS**	**Reduced form OLS**	**2SLS**	**2SRI**

Baseline model (i)	0.170***	0.00367	0.164	0.0644***	0.00105	0.0441	0.0442
N = 80573	(0.00913)	(0.0163)	(0.688)	(0.00344)	(0.00622)	(0.261)	(0.259)
Incl. intaccess (ii)	0.170***	0.00363	0.166	0.0644***	0.00104	0.0444	0.0446
N = 80572	(0.00914)	(0.0163)	(0.700)	(0.00344)	(0.00622)	(0.266)	(0.264)
Incl.demographics	0.0992***	0.00466	0.209	0.0366***	0.00131	0.0545	0.0544
(iii)	(0.00938)	(0.0165)	(0.681)	(0.00346)	(0.00612)	(0.254)	(0.252)
N = 80572							
Incl. health status,	0.0869***	0.000246	0.0286	0.0320***	-	-0.0159	-0.0155
insurance (iv)	(0.00973)	(0.0168)	(0.825)	(0.00359)	0.000324	(0.306)	(0.304)
N = 77778					(0.00623)		
Incl. health	0.0558***	-0.00136	-0.0562	0.0202***	-	-0.0453	-0.0448
behaviour (v)	(0.0101)	(0.0174)	(0.862)	(0.00366)	0.000915	(0.316)	(0.313)
N = 72831					(0.00637)		
Incl. all SES (vi)	0.0573***	-0.000391	-0.0106	0.0209***	-	-0.0251	-0.0250
N = 72769	(0.00986)	(0.0174)	(0.761)	(0.00360)	0.000576	(0.280)	(0.278)
					(0.00640)		
Incl. all variables	0.0292***	-0.00345	-0.180	0.0105***	-0.00163	-0.0886	-0.0880
N = 69307	(0.0103)	(0.0179)	(0.964)	(0.00375)	(0.00651)	(0.355)	(0.352)

i: Model includes variables *year* (separate dummies*), regions (*separate dummies), *dispincome, popdensity*
ii: Add *intaccess* to baseline variables
iii: Add *intaccess, youth, senior, female, unmarried, divorcedwidow, children* to baseline variables
iv: Add *healthcareuse, difficultyfirstaid, notgoodnotbadselfhealth, badselfhealth, healthproblems, overweightobese, insurance* to model iii
v: Add *frequentalcohol, smoker, saltconscious, weightconscious, frequentexercise* to model iv
vi: Add *university, middleschool, primarynone, books, employed, unemployed, higheconresources, insurance* to model iii

Table 9– Vegetables determinants

	Probit	**IV Probit**	**OLS**	**2SLS**	**2SRI**
Intsearch	0.0292***	-0.180	0.0105***	-0.0886	-0.0880 (0.352)
	(0.0103)	(0.964)	(0.00375)	(0.355)	
Intaccess	0.00496	0.0211	0.00165	0.00931	0.00927 (0.0292)
	(0.0294)	(0.0797)	(0.0108)	(0.0295)	
2013	-	-	-	-	-
2014	0.0209	0.0127	0.00756	0.00374	0.00375 (0.0158)
	(0.0224)	(0.0440)	(0.00815)	(0.0159)	
2015	0.0202	0.000711	0.00888	-0.000228	-0.000160 (0.0344)
	(0.0346)	(0.0960)	(0.0126)	(0.0347)	
2016	0.0331	0.00980	0.0138	0.00297	0.00305 (0.0413)
	(0.0429)	(0.116)	(0.0158)	(0.0418)	
Population	-0.00227	-0.00180	-0.000810	-0.000593	-0.000596
density	(0.00233)	(0.00321)	(0.000845)	(0.00115)	(0.00114)
Adult	-	-	-	-	-
Youth	-0.172***	-0.185***	-0.0643***	-0.0705***	-0.0704***
	(0.0180)	(0.0549)	(0.00664)	(0.0231)	(0.0229)
Senior	0.117***	0.109**	0.0385***	0.0351**	0.0352** (0.0147)
	(0.0252)	(0.0458)	(0.00846)	(0.0148)	
Female	0.289***	0.300***	0.105***	0.111***	0.111*** (0.0216)
	(0.0110)	(0.0462)	(0.00401)	(0.0219)	
Married	-	-	-	-	-
Unmarried	-0.0488***	-0.0513***	-0.0174***	-0.0187***	-0.0187***
	(0.0131)	(0.0166)	(0.00474)	(0.00663)	(0.00655)
Divorced or	0.0284	0.0242	0.00977	0.00782	0.00781 (0.00933)
widowed	(0.0179)	(0.0269)	(0.00623)	(0.00933)	
Children	-0.0476***	-0.0488***	-0.0166***	-0.0173***	-0.0173***
	(0.0126)	(0.0134)	(0.00449)	(0.00516)	(0.00513)

healthcareuse	0.0525*** (0.0135)	0.0981 (0.209)	0.0186*** (0.00477)	0.0403 (0.0780)	0.0401 (0.0771)
Difficulty first aid	0.0303*** (0.0101)	0.0316*** (0.0112)	0.0111*** (0.00364)	0.0117*** (0.00434)	0.0117*** (0.00431)
Overweight or obese	0.0187* (0.0112)	0.0191* (0.0112)	0.00701* (0.00404)	0.00724* (0.00414)	0.00724* (0.00413)
Good health state	-	-	-	-	-
Not good not bad health state	-0.0534*** (0.0143)	-0.0468 (0.0345)	-0.0190*** (0.00517)	-0.0160 (0.0119)	-0.0160 (0.0117)
Bad health state	-0.0446 (0.0361)	-0.0379 (0.0481)	-0.0155 (0.0129)	-0.0125 (0.0170)	-0.0125 (0.0167)
Health problems	0.0112 (0.0146)	0.0213 (0.0485)	0.00419 (0.00517)	0.00900 (0.0180)	0.00898 (0.0179)
Insurance	0.000441 (0.0126)	0.00317 (0.0177)	0.0000939 (0.00448)	0.00138 (0.00643)	0.00138 (0.00638)
University education	0.0343** (0.0137)	0.0411 (0.0336)	0.0125*** (0.00484)	0.0158 (0.0129)	0.0158 (0.0128)
High school education	-	-	-	-	-
Middle school education	-0.0291** (0.0122)	-0.0440 (0.0688)	-0.0109** (0.00446)	-0.0180 (0.0259)	-0.0179 (0.0256)
Primary school education or none	-0.00396 (0.0326)	-0.0377 (0.159)	-0.00130 (0.0118)	-0.0173 (0.0586)	-0.0171 (0.0578)
Reads books or newspapers frequently	0.173*** (0.0111)	0.187*** (0.0576)	0.0647*** (0.00415)	0.0713*** (0.0242)	0.0713*** (0.0239)
Inactive on the labour market	-	-	-	-	-
Employed	-0.0514*** (0.0146)	-0.0521*** (0.0147)	-0.0186*** (0.00519)	-0.0191*** (0.00551)	-0.0190*** (0.00547)
Unemployed	-0.0900*** (0.0179)	-0.0875*** (0.0224)	-0.0336*** (0.00659)	-0.0326*** (0.00747)	-0.0326*** (0.00741)
High economics resources	0.0471*** (0.0108)	0.0473*** (0.0107)	0.0174*** (0.00394)	0.0176*** (0.00405)	0.0176*** (0.00401)
Frequent alcohol consumption	0.0403*** (0.0122)	0.0451* (0.0246)	0.0150*** (0.00443)	0.0173* (0.00965)	0.0173* (0.00960)
Smoker	-0.126*** (0.0121)	-0.123*** (0.0192)	-0.0466*** (0.00449)	-0.0457*** (0.00562)	-0.0457*** (0.00556)
Salt conscious	0.218*** (0.0111)	0.227*** (0.0395)	0.0812*** (0.00415)	0.0861*** (0.0182)	0.0861*** (0.0180)
Weight conscious	0.0942*** (0.0149)	0.107* (0.0584)	0.0354*** (0.00552)	0.0416* (0.0231)	0.0416* (0.0229)
Frequent exercise	0.151*** (0.0103)	0.156*** (0.0200)	0.0541*** (0.00372)	0.0566*** (0.00967)	0.0566*** (0.00959)
dispincome	0.0000438 (0.0000479)	0.0000608 (0.0000907)	0.0000134 (0.0000173)	0.0000215 (0.0000337)	0.0000215 (0.0000333)

Table 10 - Fruits different model specifications

	Probit	**Reduced form probit**	**IV Probit**	**OLS**	**Reduced form OLS**	**2SLS**	**2SRI**
Baseline model (i) N = 79607	0.110*** (0.00970)	-0.00378 (0.0174)	-0.155 (0.693)	0.0357*** (0.00314)	-0.000817 (0.00567)	-0.0330 (0.229)	-0.0355 (0.237)

Incl. intaccess (ii)	0.110***	-0.00377	-0.157	0.0357***	-0.000812	-0.0333	-0.0359
N = 79606	(0.00970)	(0.0174)	(0.705)	(0.00315)	(0.00567)	(0.233)	(0.241)
Incl.demographics	0.0451***	-0.00236	-0.0986	0.0140***	-0.000593	-0.0237	-0.0248
(iii)	(0.0100)	(0.0176)	(0.701)	(0.00316)	(0.00559)	(0.224)	(0.230)
N = 79606							
Incl. health status,	0.0395***	-0.00360	-0.172	0.0121***	-0.000935	-0.0440	-0.0460
insurance (iv)	(0.0104)	(0.0179)	(0.835)	(0.00328)	(0.00570)	(0.269)	(0.278)
N = 76872							
Incl. health	0.0174	-0.00747	-0.357	0.00471	-0.00175	-0.0846	-0.0863
behaviour (v)	(0.0108)	(0.0187)	(0.869)	(0.00333)	(0.00579)	(0.281)	(0.284)
N = 72003							
Incl. all SES (vi)	0.0240**	-0.000988	-0.0443	0.00720**	0.0000651	0.00272	0.00253
N = 71910	(0.0106)	(0.0187)	(0.780)	(0.00329)	(0.00582)	(0.243)	(0.253)
Incl. all variables	-0.00220	0.000207	0.0111	-0.00120	0.000785	0.0411	0.0425
N = 68517	(0.0111)	(0.0193)	(1.008)	(0.00340)	(0.00589)	(0.309)	(0.318)

i: Model includes variables *year* (separate dummies*), regions (*separate dummies), *dispincome, popdensity*
ii: Add *intaccess* to baseline variables
iii: Add *intaccess, youth, senior, female, unmarried, divorcedwidow, children* to baseline variables
iv: Add *healthcareuse, difficultyfirstaid, notgoodnotbadselfhealth, badselfhealth, healthproblems, overweightobese, insurance* to model iii
v: Add *frequentalcohol, smoker, saltconscious, weightconscious, frequentexercise* to model iv
vi: Add *university, middleschool, primarynone, books, employed, unemployed, higheconresources, insurance* to model iii

Table 11 - Fruits determinants

	Probit	**IV Probit**	**OLS**	**2SLS**	**2SRI**
Intsearch	-0.00220	0.0111	-0.00120	0.0411	0.0425 (0.318)
	(0.0111)	(1.008)	(0.00340)	(0.309)	
Intaccess	-0.0211	-0.0222	-0.00804	-0.0112	-0.0114 (0.0265)
	(0.0315)	(0.0826)	(0.00984)	(0.0254)	
2013	-	-	-	-	-
2014	-0.0124	-0.0119	-0.00378	-0.00212	-0.00210 (0.0145)
	(0.0238)	(0.0461)	(0.00748)	(0.0144)	
2015	0.0207	0.0219	0.00450	0.00840	0.00851 (0.0316)
	(0.0372)	(0.0995)	(0.0115)	(0.0309)	
2016	0.00354	0.00499	-0.00183	0.00280	0.00295 (0.0378)
	(0.0464)	(0.119)	(0.0143)	(0.0368)	
Popdensity	-0.00511**	-0.00514	-0.00153**	-0.00162	-0.00162 (0.00104)
	(0.00248)	(0.00330)	(0.000767)	(0.00103)	
Adult	-	-	-	-	-
Youth	-0.195***	-0.194***	-0.0661***	-0.0636***	-0.0634***
	(0.0188)	(0.0647)	(0.00629)	(0.0198)	(0.0208)
Senior	0.367***	0.367***	0.0787***	0.0801***	0.0801*** (0.0127)
	(0.0302)	(0.0448)	(0.00663)	(0.0124)	
Female	0.129***	0.129**	0.0403***	0.0377**	0.0377* (0.0196)
	(0.0118)	(0.0625)	(0.00365)	(0.0190)	
Married	-	-	-	-	-
Unmarried	-0.202***	-0.202***	-0.0619***	-0.0613***	-0.0613***
	(0.0140)	(0.0197)	(0.00436)	(0.00597)	(0.00598)
Divorced or	-0.0789***	-0.0787***	-0.0214***	-0.0206**	-0.0206**
widowed	(0.0194)	(0.0282)	(0.00556)	(0.00832)	(0.00842)
Children	-0.0735***	-0.0734***	-0.0210***	-0.0207***	-0.0206***
	(0.0136)	(0.0155)	(0.00404)	(0.00459)	(0.00463)
healthcareuse	0.0249*	0.0220	0.00801*	-0.00127	-0.00155 (0.0698)
	(0.0145)	(0.222)	(0.00434)	(0.0679)	
Difficulty first	0.0391***	0.0390***	0.0119***	0.0117***	0.0116***
aid	(0.0108)	(0.0126)	(0.00332)	(0.00385)	(0.00394)
Overweight or	-0.0179	-0.0179	-0.00543	-0.00552	-0.00554 (0.00373)
obese	(0.0120)	(0.0122)	(0.00366)	(0.00372)	
Good health	-	-	-	-	-
state					

Not good not bad health state	-0.0453*** (0.0155)	-0.0457 (0.0337)	-0.0146*** (0.00465)	-0.0158 (0.0103)	-0.0159 (0.0106)
Bad health state	-0.0476 (0.0397)	-0.0480 (0.0494)	-0.0139 (0.0112)	-0.0151 (0.0144)	-0.0152 (0.0148)
Health problems	0.0200 (0.0158)	0.0194 (0.0511)	0.00558 (0.00463)	0.00354 (0.0156)	0.00345 (0.0161)
Insurance	-0.0220 (0.0134)	-0.0221 (0.0190)	-0.00672* (0.00407)	-0.00728 (0.00579)	-0.00729 (0.00581)
University education	0.0396*** (0.0149)	0.0392 (0.0372)	0.0107** (0.00428)	0.00924 (0.0112)	0.00919 (0.0115)
High school education	-	-	-	-	-
Middle school education	-0.0243* (0.0129)	-0.0233 (0.0735)	-0.00767* (0.00411)	-0.00464 (0.0225)	-0.00453 (0.0232)
Primary school education or none	0.0290 (0.0356)	0.0312 (0.166)	0.00948 (0.0104)	0.0163 (0.0507)	0.0165 (0.0524)
Reads books or newspapers frequently	0.140*** (0.0118)	0.139** (0.0693)	0.0448*** (0.00379)	0.0420** (0.0211)	0.0419* (0.0217)
Inactive on the labour market	-	-	-	-	-
Employed	-0.0752*** (0.0157)	-0.0751*** (0.0163)	-0.0231*** (0.00465)	-0.0229*** (0.00483)	-0.0229*** (0.00490)
Unemployed	-0.123*** (0.0191)	-0.123*** (0.0219)	-0.0394*** (0.00611)	-0.0398*** (0.00699)	-0.0398*** (0.00683)
High economics resources	0.0850*** (0.0115)	0.0849*** (0.0118)	0.0258*** (0.00362)	0.0257*** (0.00369)	0.0257*** (0.00369)
Frequent alcohol consumption	-0.0580*** (0.0130)	-0.0583** (0.0273)	-0.0187*** (0.00407)	-0.0197** (0.00846)	-0.0198** (0.00873)
Smoker	-0.301*** (0.0126)	-0.301*** (0.0152)	-0.100*** (0.00429)	-0.101*** (0.00519)	-0.101*** (0.00521)
Salt conscious	0.240*** (0.0116)	0.239*** (0.0518)	0.0787*** (0.00389)	0.0766*** (0.0157)	0.0765*** (0.0164)
Weight conscious	0.184*** (0.0155)	0.183*** (0.0651)	0.0616*** (0.00531)	0.0590*** (0.0199)	0.0589*** (0.0208)
Frequent exercise	0.137*** (0.0111)	0.136*** (0.0278)	0.0422*** (0.00338)	0.0412*** (0.00839)	0.0411*** (0.00869)
dispincome	0.0000283 (0.0000510)	0.0000273 (0.0000968)	0.0000126 (0.0000160)	0.00000911 (0.0000299)	0.00000901 (0.0000305)

Table 12 - Overweight or obese model specifications

	Probit	**Reduced form probit**	**IV Probit**	**OLS**	**Reduced form OLS**	**2SLS**	**2SRI**
Baseline model (i) N = 80716	0.0705*** (0.00918)	-0.549 (0.623)	-0.551 (0.619)	0.0261*** (0.00341)	-0.00527 (0.00608)	-0.220 (0.262)	-0.220 (0.254)
Incl. intaccess (ii) N = 80715	0.0707*** (0.00919)	-0.558 (0.632)	-0.560 (0.628)	0.0262*** (0.00342)	-0.00528 (0.00608)	-0.225 (0.267)	-0.225 (0.259)
Incl.demographics (iii) N = 80715	0.00351 (0.00981)	-0.383 (0.684)	-0.384 (0.680)	-0.000266 (0.00327)	-0.00368 (0.00565)	-0.152 (0.236)	-0.152 (0.232)
Incl. health status, insurance (iv) N = 77906	-0.0139 (0.0102)	-0.258 (0.848)	-0.257 (0.844)	-0.00597* (0.00337)	-0.00250 (0.00573)	-0.122 (0.282)	-0.122 (0.279)

Incl. health behaviour (v) N = 72834	-0.0123 (0.0105)	-0.185 (0.898)	-0.184 (0.896)	-0.00541 (0.00345)	-0.00187 (0.00593)	-0.0922 (0.294)	-0.0922 (0.293)
Incl. all SES (vi) N = 72873	0.0286*** (0.0103)	0.00242 (0.0182)	0.106 (0.790)	0.00910*** (0.00342)	0.000574 (0.00600)	0.0250 (0.261)	0.0250 (0.261)
Incl. all variables N = 69310	0.00760 (0.0108)	0.00366 (0.0188)	0.198 (1.009)	0.00209 (0.00355)	0.000808 (0.00611)	0.0439 (0.332)	0.0439 (0.332)

i: Model includes variables *year* (separate dummies*), regions (*separate dummies), *dispincome, popdensity*
ii: Add *intaccess* to baseline variables
iii: Add *intaccess, youth, senior, female, unmarried, divorcedwidow, children* to baseline variables
iv: Add *healthcareuse, difficultyfirstaid, notgoodnotbadselfhealth, badselfhealth, healthproblems, insurance* to model iii
v: Add *frequentalcohol, smoker, saltconscious, weightconscious, frequentexercise, healthydiet* to model iv
vi: Add *university, middleschool, primarynone, books, employed, unemployed, higheconresources, insurance* to model iii

Table 13 –O*bese* model specifications

	Probit	**Reduced form Probit**	**IV Probit**	**OLS**	**Reduced form OLS**	**2SLS**	**2SRI**
Baseline model (i) N = 69026	-0.0214 (0.0141)	0.0352 (0.0256)	1.175* (0.603)	-0.00309 (0.00205)	0.00488 (0.00368)	0.203 (0.164)	0.204 (0.154)
Incl. intaccess (ii) N = 69025	-0.0218 (0.0141)	0.0351 (0.0256)	1.188* (0.607)	-0.00314 (0.00205)	0.00486 (0.00368)	0.206 (0.168)	0.207 (0.156)
Incl. demographics (iii) N = 69025	-0.0223 (0.0144)	0.0388 (0.0260)	1.254** (0.562)	-0.00335 (0.00205)	0.00543 (0.00365)	0.227 (0.166)	0.223 (0.150)
Incl. health status, insurance (iv) N = 66728	-0.0447*** (0.0152)	0.0414 (0.0268)	1.447*** (0.534)	-0.00658*** (0.00212)	0.00553 (0.00369)	0.280 (0.212)	0.269 (0.180)
Incl. health behaviour (v) N = 64955	-0.0338** (0.0155)	0.0368 (0.0273)	1.420** (0.621)	-0.00491** (0.00215)	0.00498 (0.00373)	0.270 (0.228)	0.245 (0.184)
Incl. all SES (vi) N = 67870	0.00874 (0.0148)	0.0448* (0.0265)	1.464*** (0.479)	0.000933 (0.00207)	0.00629* (0.00367)	0.294 (0.196)	0.274* (0.160)
Incl. all variables N = 64603	-0.00864 (0.0157)	0.0424 (0.0275)	1.586*** (0.507)	-0.00152 (0.00216)	0.00586 (0.00373)	0.341 (0.257)	0.318 (0.202)

i: Model includes variables *year* (separate dummies*), regions (*separate dummies), *dispincome, popdensity*
ii: Add *intaccess* to baseline variables
iii: Add *intaccess, youth, senior, female, unmarried, divorcedwidow, children* to baseline variables
iv: Add *healthcareuse, difficultyfirstaid, notgoodnotbadselfhealth, badselfhealth, healthproblems, insurance* to model iii
v: Add *frequentalcohol, smoker, saltconscious, weightconscious, frequentexercise, healthydiet* to model iv
vi: Add *university, middleschool, primarynone, books, employed, unemployed, higheconresources, insurance* to model iii

Table 14 *Obese* output

	Probit	**IV Probit**	**OLS**	**2SLS**	**2SRI**
Intsearch	-0.00864 (0.0157)	1.586*** (0.507)	-0.00152 (0.00216)	0.341 (0.257)	0.318 (0.202)

Intaccess	0.0597	-0.0826	0.00849	-0.0175	-0.0162
	(0.0448)	(0.0648)	(0.00595)	(0.0208)	(0.0167)
2013	-	-	-	-	-
2014	0.0538	0.0314	0.00728	0.00682	0.0196**
	(0.0348)	(0.0311)	(0.00465)	(0.00563)	(0.00904)
2015	0.0456	0.129***	0.0109	0.0281*	0.0402**
	(0.0342)	(0.0408)	(0.00718)	(0.0155)	(0.0198)
2016	0.0972***	0.198***	0.0198**	0.0432**	0.0547**
	(0.0332)	(0.0502)	(0.00902)	(0.0206)	(0.0238)
Population density	-0.00867**	-0.00851***	0.00113**	-0.00182**	-0.00182***
	(0.00343)	(0.00304)	(0.000510)	(0.000789)	(0.000664)
Adult	-	-	-	-	-
Youth	-0.429***	-0.219	-0.0371***	-0.0249**	-0.0169
	(0.0364)	(0.144)	(0.00293)	(0.00996)	(0.0131)
Senior	-0.00237	0.0678*	0.000628	0.0155	0.0113
	(0.0332)	(0.0346)	(0.00585)	(0.0131)	(0.00890)
Female	-0.274***	-0.272***	-0.0367***	-0.0573***	-0.0560***
	(0.0168)	(0.0501)	(0.00223)	(0.0158)	(0.0124)
Married	-	-	-	-	-
Unmarried	-0.249***	-0.144*	-0.0332***	-0.0297***	-0.0289***
	(0.0200)	(0.0785)	(0.00256)	(0.00402)	(0.00371)
Divorced or widowed	-0.0763***	-0.0196	-0.0129***	-0.00653	-0.00657
	(0.0247)	(0.0368)	(0.00383)	(0.00654)	(0.00557)
Households with children	-0.0382**	-0.0168	-0.00506*	-0.00342	-0.00274
	(0.0181)	(0.0193)	(0.00259)	(0.00329)	(0.00297)
Health care use	0.0147	-0.339***	0.00195	-0.0731	-0.0681
	(0.0199)	(0.116)	(0.00285)	(0.0566)	(0.0445)
Difficulty first aid	-0.00693	-0.0113	-0.000994	-0.00250	-0.00312
	(0.0154)	(0.0117)	(0.00211)	(0.00275)	(0.00250)
Good health state	-	-	-	-	-
Not good not bad health state	0.211***	0.0901	0.0358***	0.0260***	0.0261***
	(0.0197)	(0.0768)	(0.00344)	(0.00835)	(0.00705)
Bad health state	0.222***	0.0960	0.0450***	0.0350***	0.0352***
	(0.0443)	(0.0864)	(0.00976)	(0.0131)	(0.0115)
Health problems	0.208***	0.0574	0.0339***	0.0175	0.0183*
	(0.0200)	(0.0855)	(0.00346)	(0.0130)	(0.0105)
Insurance	0.0351*	-0.00175	0.00507*	-0.000128	0.000896
	(0.0185)	(0.0226)	(0.00264)	(0.00496)	(0.00371)
University education	-0.192***	-0.180***	-0.0221***	-0.0342***	-0.0328***
	(0.0219)	(0.0408)	(0.00245)	(0.00964)	(0.00725)
High school education	-	-	-	-	-
Middle school education	0.129***	0.172***	0.0206***	0.0397***	0.0435***
	(0.0180)	(0.0166)	(0.00283)	(0.0148)	(0.0148)
Primary school education or none	0.202***	0.339***	0.0420***	0.0869**	0.0936***
	(0.0438)	(0.0356)	(0.00957)	(0.0354)	(0.0341)
Reads books or newspapers frequently	-0.0130	-0.119***	-0.00209	-0.0259	-0.0236*
	(0.0169)	(0.0339)	(0.00242)	(0.0181)	(0.0138)
Inactive					
Employed	-0.0537**	-0.0104	-0.00841***	-0.00321	-0.00685**
	(0.0219)	(0.0284)	(0.00295)	(0.00524)	(0.00310)
Unemployed	0.0155	0.0200	0.000574	0.00274	-0.00252
	(0.0282)	(0.0211)	(0.00364)	(0.00461)	(0.00414
High economics resources	-0.123***	-0.0864**	-0.0176***	-0.0192***	-0.0183***
	(0.0165)	(0.0352)	(0.00235)	(0.00301)	(0.00240)
Frequent alcohol consumption	-0.0832***	-0.0845***	-0.0104***	-0.0171***	-0.0181***
	(0.0179)	(0.0197)	(0.00255)	(0.00580)	(0.00544)

Smoker	-0.0537***	-0.0412**	-0.00844***	-0.00988***	-0.0114***
	(0.0186)	(0.0192)	(0.00251)	(0.00317)	(0.00318)
Salt conscious	0.0539***	-0.0441	0.00725***	-0.00967	-0.00886
	(0.0177)	(0.0425)	(0.00229)	(0.0130)	(0.0104)
Weight conscious	0.00488	-0.101***	0.00125	-0.0212	-0.0190
	(0.0230)	(0.0387)	(0.00318)	(0.0173)	(0.0132)
Healthy diet	0.00602	0.0150	0.00116	0.00357	0.00310
	(0.0175)	(0.0135)	(0.00245)	(0.00341)	(0.00275)
Frequent exercise	-0.227***	-0.194***	0.0300***	-0.0404***	-0.0381***
	(0.0162)	(0.0515)	(0.00209)	(0.00821)	(0.00554)
Disposable income		-0.000163***	-0.00000792	-0.0000357	-0.0000340*
		(0.0000609)	(0.0000102)	(0.0000242)	(0.0000194)