Exploring the connection: Overweight, dietary knowledge, and urban-rural disparities in The Netherlands.

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Bachelor Thesis: Economics and Business Economics

Name student: Dennis Stofbergen

Student ID number: 575686

Supervisor: Carel-Peter van Erpecum

Second assessor: xxx

Date final version: 10-07-2023

Table of Contents

Abstract

This study aimed to explore the mediation effect of dietary knowledge in the relationship between urban-rural differences and overweight rates among individuals in the Netherlands. Overweight rates are a major concern for the Dutch population as being overweight has severe health-, environmental-, economic-, and cultural effects. Adequate dietary knowledge is important for individuals in making proper dietary choices. Data from multiple datasets from the LISS panel are utilized and logistic regression and a mediation technique, with an extensive list of control variables, were used to investigate the mediating role of dietary knowledge on the relationship between urban-rural differences and overweight rates. The study did not find statistically significant evidence for the mediating role of dietary knowledge. The effect of urban-rural differences on overweight rates, the effect of urban-rural differences on dietary knowledge and the effect of dietary knowledge on overweight rates were all examined, however, did not provide statistically significant results. Nonetheless, this study still contributes to the existing knowledge by illuminating the complex interplay of urban-rural differences, dietary knowledge and overweight and obesity. In the future, studies with larger, more representative samples to deepen the understanding of these relationships are needed.

Keywords: Dietary knowledge; Urban-Rural Differences; Overweight; Netherlands; Multiple imputation; Mediation; Obesity; BMI; LISS panel

Acknowledgements

I Would like to express my profound gratitude towards Carel-Peter van Erpecum, whose expertise, understanding, and patience, added considerably to my experience in writing this thesis. I appreciate his vast knowledge and skill in many areas, but, more importantly, his ability to share that wisdom and his constant readiness to help has been invaluable to me.

Introduction

An increasing prevalence of overweight is a problematic public health concern (S. Kim & Popkin, 2006). It has major health, economic and environmental consequences. Obesity has not only a heavy negative impact on the health at the individual level, also on society as a whole. This is mainly due to overweight being the cause of various chronic conditions (Must, 1999), such as diabetes, cardiovascular diseases, and cancer. What makes this issue even more complicated is the fact that there are big differences in overweight between rural and urban populations. It was believed that one of the most important causes of the rise in obesity rates is the rising proportion of the population living in urban areas (Rosenberg et al., 2016). However, lately, there is evidence that obesity rates are rising at a faster rate in rural areas compared to urban areas. So, the urban-rural differences gap of obesity is closing or even reversing (Visvikis-Siest, 2019). Hence, it is crucial to unravel the contributors to these differences in obesity and overweight rates between urban and rural populations even more to provide the possibility to target overweight prevention techniques more efficiently.

Dietary knowledge refers to the understanding and awareness of nutritional concepts, food choices and their impact on health and well-being. Dietary knowledge plays an integral part in understanding the divergent patterns of obesity in urban and rural areas. A balanced and nutrient-dense diet is important for maintaining a healthy weight, yet the role of an individual's dietary knowledge, and its impact on their eating habits, is often overlooked. Recent studies suggest that dietary knowledge varies significantly across population groups, and is often influenced by socioeconomic factors, education levels, and the environment (Osler & Hansen, 1993). In addition, a study on young South African black female students has already been conducted regarding the impact of dietary knowledge on urban-rural disparities in overweight (Steyn et al., 2000). However, the study did not find any significant urban-rural differences in the level of dietary knowledge.

In recent years, urbanization and modernization was the cause of lifestyle changes (Popkin, 1999). In urban areas, processed food became readily available in combination with numerous fast-food restaurants. This led to an increase in overweight and obesity (Dunn, 2010). Conversely, rural areas have generally more physically demanding lifestyles. Their work often involves manual labor. Additionally, individuals in rural areas are older on average and have lower levels of health education (Marcellini et al., 2007), as dietary knowledge and health education are strongly correlated (Abu-Baker et al., 2021), dietary knowledge may have some explanatory power to overweight rates. Leading to, even in rural areas, rising rates of obesity. Understanding the factors contributing to these disparities in overweight between urban and rural areas is crucial for designing effective public health interventions to combat overweight.

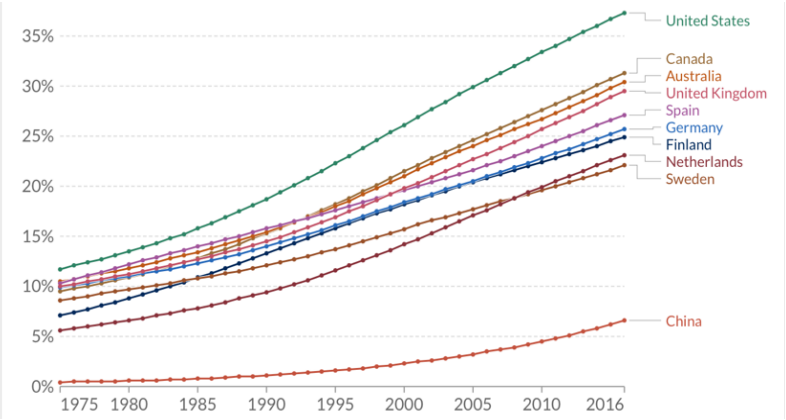
This thesis presents a comprehensive examination of the mediating role of dietary knowledge in the relationship between urban-rural differences and the rates of overweight individuals in the Netherlands. The study starts with a theoretical background, offering an in-depth exploration of the existing literature on dietary knowledge, overweight, and urban-rural differences. Next, the methodology is presented. Here, the techniques used to merge, process, and analyze the data are thoroughly explained. The data is collected from the LISS panel, which is a survey research program that collects data from a representative Dutch sample over time. Following this, the results will be presented. These include both the main findings and auxiliary discoveries that were uncovered during the study. The discussion segment follows, where the results will be interpreted in the context of the literature. Furthermore, the limitations and implications of the study will be considered in the discussion. Finally, in the conclusion the major takeaways from the study are summarized. Moreover, this section also brings forward potential avenues for future research.

Theoretical background

The goal of this theoretical framework is to offer a comprehensive summary of the current understanding concerning urban-rural disparities in overweight, particularly emphasizing dietary knowledge among the Dutch population. First, the global and regional trends of obesity and overweight between urban and rural populations will be examined. Next, the causes of the disparities between overweight in urban and rural areas will be highlighted, enclosing cultural, socioeconomic, and environmental dimensions. Then, the aspect of dietary knowledge and awareness will be thoroughly examined. Furthermore, the theoretical approaches of this study will be discussed, and the research question and the hypotheses will be proposed. This theoretical background will lay the foundation for understanding the complex interplay of overweight and urban and rural populations.

Global trends of overweight and obesity

As there is not a single country in the world where obesity rates were not rising over the last couple of years, obesity is now seen as a worldwide threat. First, a definition of obesity and overweight is important for the understanding of the research. The body mass index, an individual's weight in kilograms divided by the square of height in meters, is commonly used for this purpose. If the BMI is 25 or higher, the individual is defined as overweight (CBS, n.d.-a). If the body mass index is higher than 30, the individual is considered obese. As shown in



Adapted source: (Ritchie & Roser, 2017)

Figure 4.1: Percentage of the overweight proportion of the population.

Figure 4.1, the percentage of the overweight population is rising in all the displayed countries. Further, the percentage obese of the population is even increasing in every country in the world (Kluger, 2023). The reason for displaying these ten countries in the graph is that these countries are considered the most influential countries in the world in terms of market orientation (Valenzuela-Fernández et al., 2018). The causes of the increase are not yet understood entirely, however, there is already a lot of research done to provide possible explanations of factors causing this upwards trend.

Obesity is a very complex subject because, besides external causes, obesity is also dependent to some degree on genetics (Swinburn et al., 2011). Besides genetics, the first, most common cause of obesity is a higher energy intake than energy expenditure (Flegal, 2010). Secondly, eating high-calorie and fat-rich food is nowadays more easily accessible and affordable. Also, the portion size increased over the years and higher proportions of food are pre-packaged and processed, making it less healthy, but more attractive due to its convenience (Rolls, 2003).

On the other hand, physical activity has decreased a lot (Centers for Disease Control and Prevention (CDC), 2007). More time is spent now on behaviors without activity, like watching Netflix and surfing the internet. In addition, simple tasks are getting more convenient and take less time, which is mostly a good thing when thinking about time consumption, but it leads to a decrease in burnt calories during the day. In contrast to the decrease in physical activity in everyday life, when examining general trends of the Netherlands, between 2017 and 2021 there is no statistically significant change in the amount of physical activity in leisure time (CBS, 2022b). Besides the trend in physical activity, the calorie intake of the Dutch population was rising between 2012 and 2016 (Vellinga et al., 2022).

There are more medicines consumed now compared to the past (Kebriaeezadeh et al., 2013). Some of these medicines come with the side effect of weight gain (Keith et al., 2006). Although it is difficult to estimate the causal effect of medicine intake on weight gain, it is known that some of the most prescribed drugs, like antidepressants and antiepileptic drugs, could cause weight gain (Ness-Abramof & Apovian, 2005). Besides increased medicine intake, individuals also seem to act upon the increased health opportunities by quitting smoking. Even though quitting smoking increases health in a lot of aspects, there seems to be a negative effect on BMI when an individual quits smoking (Gruber & Frakes, 2006). Sleep has severe consequences on the body mass index as well. A lack of hours of sleep negatively affects BMI (Gangwisch et al., 2005), and an individual would have an increased desire for food intake with fewer hours of sleep (Spiegel et al., 2004). Besides these aspects, also particular policies, such as the agricultural corn policy, increasing the consumption of nutrient-poor food by subsidizing the production of corn in the United States, have a negative effect on obesity rates (Franck et al., 2013). Furthermore, certain infections, such as Adenoviros-36, and social networks could have effects on the increase in obesity rates (Cohen-Cole & Fletcher, 2008; Ponterio & Gnessi, 2015). A

study has been done with the conclusion that an individual would have an increased chance of 57% of becoming obese when they have a friend who is obese as well (Christakis & Fowler, 2007).

Obesity and overweight trends in The Netherlands

Because this research focuses on the Dutch population, it is important to identify the obesity and overweight trends in the Netherlands. More than half of the Dutch population is overweight or obese in 2022 (CBS, 2022a). Overweight differs for certain subgroups. Men are more likely (17%) than women (12%) to be obese. Also, when focusing on the age of a Dutch individual, there generally is a positive relationship between BMI and age, when the individual gets older up to a certain point, the BMI is rising, on average, as well (Vinke et al., 2020). But when the individual gets older than a certain point, the BMI will decrease again. Besides age, education plays some part in the differences in the prevalence of obesity and overweight. The percentage of overweight decreases from 41.6% with a low education to 30.9% with a high education (CBS, 2022a).

Urban-rural differences

There are distinct differences in obesity and overweight rates between urban and rural areas. While some studies show conflicting results, the prevalence of obesity and overweight is generally higher in rural areas than in urban areas (Johnson & Johnson, 2015). These urban-rural disparities can be attributed to various factors, including disparities in access to healthy food. In urban areas, it is often more challenging for individuals to find fresh products leading to a higher reliance on processed food products (Dekker et al., 2017). Urban areas have more access to physical activity (Martin et al., 2005). In addition, residents of rural areas are more dependent on cars as a way of transport, which negatively impacts the amount of physical activity (Carroll et al., 2021). But, in contrast, urban residents are more inclined to consume more processed food (Sauer et al., 2021). Also, cultural attributes play a significant role. The rural population has a different body image than the urban population, which could potentially contribute to higher obesity as well. In some rural areas, larger body size may be more socially accepted compared to urban areas (Jackson et al., 2003). Lastly, the rural population, on average, has lower incomes and less access to healthcare and health maintenance resources (Sicular et al., 2007). Fresh and healthy food is generally more expensive than nutrient-poor processed food, low income could contribute to individuals not being able to afford healthy food options (Daniel, 2020).

Health effects of overweight and obesity

High BMI is causing a lot of diseases and deaths every year. Cardiovascular diseases are the main cause of death associated with a high BMI (Afshin et al., 2017). 41% of the deaths due to high BMI are because of cardiovascular diseases (Afshin et al., 2017). Furthermore, diabetes, chronic kidney disease, cancer and musculoskeletal disorders are possible diseases due to a high BMI. Besides death, the summed-up diseases could also cause disability-adjusted life years. One disability-adjusted life year

represents a year of living in perfect health being lost. So, obesity and overweight are dangerous for the health of an individual. Also, it is important to note that these health effects are not limited to BMI alone, but also apply to the distribution of body fat. An individual with a lot of excess body fat around the waist may be at a greater health risk than an individual with excess fat in other parts of the body (Frank et al., 2019). Furthermore, obesity comes with several other health effects, not directly resulting in death.

I. Respiratory diseases: obese and overweight individuals are more likely to develop respiratory problems, such as sleep apnea and shortness of breath. (Makker, 2010)
II. Gastrointestinal diseases: Individuals with high BMI have been associated with an increased risk of developing gastrointestinal diseases. (Camilleri et al., 2017)
III. Reproductive health problems: Overweight women have a higher risk of fertility problems, gestational diabetes, and pregnancy complications, while men have a lower sperm count and decreases testosterone levels. (Best & Bhattacharya, 2015)
IV. Joint problems: Excess weight puts more pressure on the joints, which could lead to joint pain and even osteoarthritis. (King et al., 2013)
V. Mental health issues: there is a link between high BMI and mental health issues, such as depression, anxiety, and low self-esteem. (Avila et al., 2015)

Further to the direct health complications of obesity, overweight and obesity could exacerbate the progression of age-related conditions, such as Alzheimer's and other forms of cognitive decline (Tam et al., 2020). The toll that obesity and overweight could have on the immune system of an individual is another critical area of concern (de Heredia et al., 2012). This would suggest a higher vulnerability to infectious diseases.

Environmental effects of overweight and obesity

Besides the, heavily researched, health effects of obesity and overweight, the environmental effects are important as well. Individuals who are overweight generally consume more food and are less active. This is not only a direct cause of obesity. For someone to maintain obese or overweight, they consume, on average, a lot more than an individual who is not obese or overweight. Food production, transportation, and disposal of this added production of food contribute to greenhouse gas emissions. (Swinburn et al., 2019) Furthermore, because of the added health risks, obese individuals require more medical treatment, leading to an increase in medical waste, such as packaging and disposable equipment. (Rutala & Mayhall, 1992) Besides overconsumption, when the weight of individuals is increasing, there could be an increase in vehicle emissions because there is a need for bigger vehicles. Additionally, the added weight of overweight passengers or drivers will cause the vehicle to use more

fuel (Cheah & Lynette W., 2010). And lastly, urban planning needs to be modified. Obese people require, for example, larger sidewalks and more public spaces to encourage physical activity (Day et al., 2013).

Another important aspect to note is that obesity and overweight have the negative external effect of increased air pollution due to various factors, such as increased vehicle usage, higher energy consumption, and increased healthcare. The increase in air pollution has negative health effects on all individuals. Obesity is not only a negative health factor for the overweight or obese individual itself, however, it could also contribute to negative health outcomes for individuals who are not obese or overweight by increased air pollution (Kampa & Castanas, 2008).

Cultural effects of overweight and obesity

Lastly, we need to consider the cultural effects of overweight and obesity as well. Although cultural effects can be broad and variable in different cultures, some common impacts are:

I. Body image: In a lot of societies, being thin is idealized. Individuals who are obese or overweight could face a negative body image which could negatively impact mental health and lower self-esteem. (Schwartz & Brownell, 2004)
II. Discrimination: Overweight and obese people often face discrimination, and the media often portrays obese or overweight individuals negatively (Klaczynski et al., 2004), which could lead to social isolation or lower quality of life. (R. Puhl & Brownell, 2001)
III. Public spaces: Most public spaces have not been designed for overweight and obese individuals, such as airplane seats and public transport (Evans et al., 2012).
IV. Health culture: In cultures where physical fitness is valued highly, individuals may get criticism. Individuals could then feel stigmatized, leading to the creation of a burden to participate in physical activities.

However, in some cultures, there is more and more acceptance towards obesity and overweight. This movement aims to reduce stigma and promote acceptance (R. M. Puhl & Brownell, 2003). The body positivity movement is a social movement to empower and improve the confidence of those who traditionally have been marginalized due to their appearance or body size (Lazuka et al., 2020). It encourages people to adopt more forgiving attitudes towards their bodies. Besides the body positivity movement, the fat acceptance movement is a social movement seeking to change societal attitudes towards fat people (Sturmer et al., 2003). In essence, there is a delicate balance between promoting positivity and acceptance while also acknowledging and addressing potential health risks.

Economic effects of overweight and obesity

Obesity has significant economic impacts at both the individual and the societal level. At the individual level, obesity and being overweight lead to increased healthcare costs due to associated health complications which require ongoing medical care (Hammond & Levine, 2010). Furthermore, obesity and overweight can impact a person's employment status and earning potential (Morris, 2007). This generally is due to discrimination or limited physical abilities. On the societal level, there are several economic consequences of being overweight or obese:

I. An increased prevalence of overweight and obesity can lead to increased healthcare expenditures. This leads to an overconsumption of medical resources and insurance systems (Bertakis & Azari, 2005).
II. Lost productivity due to being overweight and obese can lead to substantial economic losses (Goettler et al., 2017). Obesity is not only a public health concern, but a significant economic issue as well. Overweight and obesity cost the Dutch government approximately 79 billion every year in terms of health costs and lost productivity (Hecker et al., 2022).
III. Obesity is generally more common among low-income groups. The related health issues could lead to bigger economic and health disparities (McLaren, 2007).
IV. For an individual to maintain overweight or obese, their food consumption generally is higher. This could indirectly affect food prices for all customers (Gilbert, 2010).

Dietary knowledge

Dietary knowledge refers to an individual's understanding of how different foods and nutrients affect health and well-being. For example, in terms of portion size, knowing that the recommended portion size for cooked pasta is about 100 grams, rather than serving themselves a larger portion that exceeds the recommended amount. In terms of nutrient balance, it could be, for example, beneficial to understand the importance of adding a source of protein, such as fish, along with whole grains, and fruits and vegetables in a meal to ensure a balanced nutrient intake. Lastly, understanding how dietary choices influence health is important for sufficient dietary knowledge, for example, consuming excessive amounts of sugar could contribute to weight gain and increased risk of type 2 diabetes (Taylor, 1999). When an individual has strong dietary knowledge, this individual is more capable of making healthy food choices (Shimokawa, 2013). Also, these individuals are generally more capable of preventing certain health conditions. So, a strong foundation in dietary knowledge could have positive effects on health outcomes and it contributes to a higher quality of life.

There are lots of factors contributing to dietary knowledge such as education, socioeconomic status, cultural background, access to dietary information, health status, personal interests, media influence,

government policies and access to healthcare professionals (Xu et al., 2020). These factors interact in complex ways to form an individual his or her dietary knowledge. An individual could try to improve dietary knowledge by improving on one or multiple of these factors.

Adequate dietary knowledge is crucial to make healthy food choices. Healthy food and a healthy lifestyle contribute to the prevention and management of obesity. When an individual better understands proper portion sizes, nutritional values of certain food choices and calorie intake, the individual is better informed about overeating and the possible dangers of calorie-dense food. It is important to note that dietary knowledge alone is not always sufficient to prevent or solve obesity; there are numerous other factors involved as well.

Additionally, the food industry is making the relationship between dietary knowledge and obesity and overweight even more complex due to their advertisements containing misconceptions about dietary information (Harris et al., 2009). For example, a cereal may be promoted as 'made with whole grains', which sounds healthy. But the second ingredient listed might be sugar, which is unhealthy.

Theoretical approaches:

I. Social cognitive theory

The social cognitive theory is a psychological framework which supposes behavioral, cognitive, and environmental factors all influence each other (Conner & Norman, 2015). Key components are observational learning, the belief in our own abilities and reinforcement.

I. Cognitive factors: Urban and rural populations might have different dietary knowledge. Possible explanations could be due to differences in education or access to information. To give an example, the urban population might have more individuals, on average, that are highly educated ("Erratum to: The Urban–Rural Education Gap: Do Cities Indeed Make Us Smarter?," 2021), thus, urban residents would have more dietary knowledge.
II. Behavioral factors: Individuals living in urban areas might observe different social norms regarding eating behaviors than rural individuals. For example, less frequent fast-food consumption. The habits of other individuals could influence their own eating habits.
III. Environmental factors: The food environment in urban and rural areas could be significantly different. Urban areas have access to a bigger variety of food options, both fresh and healthy food options and unhealthy processed food options, while rural areas might have less access to multiple food options and are restricted to non-fresh produced food options. An explanation could be that a supermarket in a rural area might not get freshly produced food every day suitable for sale, due to the lack of customers compared to urban areas or due to transportation

complications. Furthermore, supermarkets and other food suppliers in urban areas have generally more generous opening and closing times than in rural areas.

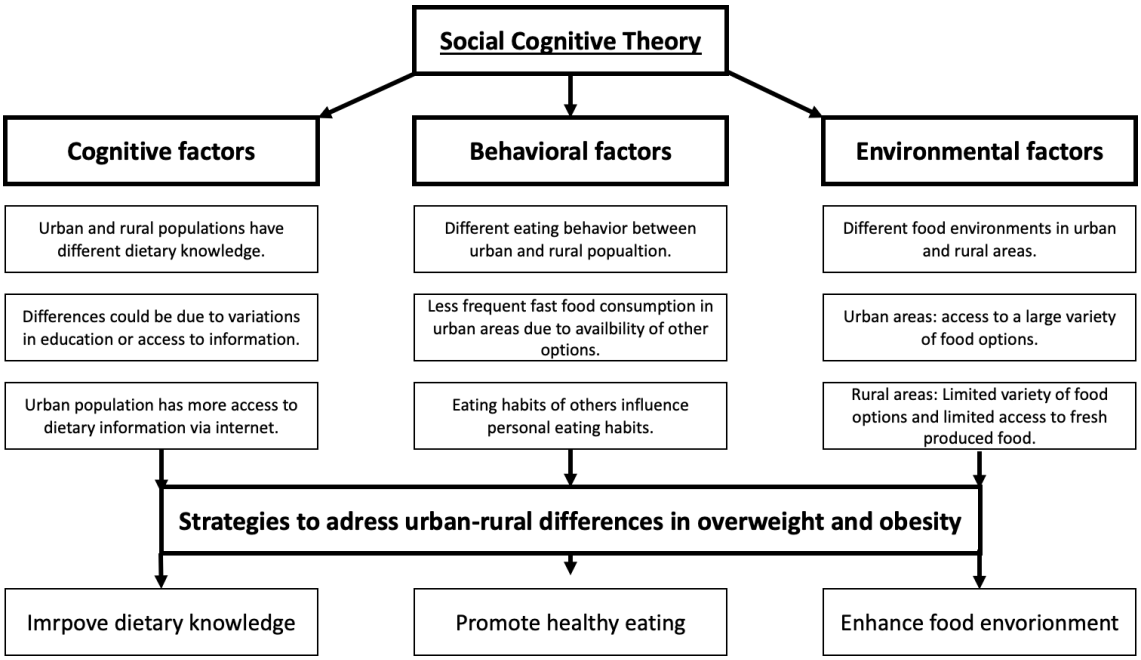


Figure 4.2: Social Flowchart Cognitive Theory modified to this study.

Employing the framework of the social cognitive theory (outlined in figure 4.2), tackling the urban-rural differences in overweight and obesity might be achievable through the enhancement of dietary knowledge, encouragement of healthier eating habits, or betterment of the food environment. The Social Cognitive Theory focuses on the role of cognitive processes. It emphasizes the impact of social and environmental factors on individual behavior and emphasized the role of self-regulation and self-control in changes in behavior.

II. Health belief model

The health model (HBM) is a psychological framework which aims to predict health behaviors based on beliefs and attitudes (Janz & Becker, 1984). The HBM could be applied to explain urban-rural differences in overweight and obesity while focusing on dietary knowledge in the following ways:

I. Perceived severity: The urban and rural populations may have different understandings of the severity of obesity. If the population lacks dietary knowledge about the possible complications of obesity, they may underestimate the severity.
II. Perceived susceptibility: Different populations may have different perceptions of susceptibility to obesity. In urban areas, there may be more awareness of the prevalence of obesity and the

associated health risks. This could be due to less exposure to health education institutions in rural areas.

III. Perceived benefits: Different populations may have different perceptions of the benefits of changing to a healthier diet. Individuals in urban areas may have a greater variety of food options and better access to health-conscious communities and health-educational institutions. They may perceive the benefits of changing to a healthier diet as including weight management, increased energy levels and reduced risk of chronic diseases.

IV. Perceived barriers: Rural populations might have more barriers to healthy eating because they have limited access to healthy food.

V. Self-efficacy: Dietary knowledge can boost self-efficacy. They feel more capable of making healthy dietary changes. If any of the groups (urban or rural) lacks this information, they might feel less capable of making changes.

The Health Belief model uniquely emphasizes the role of an individual in his or her perception and belief in health behavior. It emphasizes the importance of perceived threats, benefits, and barriers in determining if an individual will engage in health-promoting behavior.

The ecological model of health behavior

The ecological model of health behavior is a health model suggesting that our behaviors are based on multiple levels of influence (Diclemente et al., 2002). The model usually consists of five hierarchical levels:

I. Individual level: It could be possible that there are individual differences in dietary knowledge between urban and rural individuals. This could subsequently lead to differences in overweight and obesity rates between the urban and rural populations.

II. Interpersonal level: The norms about diet and body size could differ between the urban and rural populations. This can influence BMI.

III. Organizational level: Urban or rural schools and workplaces might offer different dietary education and different food options. For example, rural schools might have trouble finding resources for nutritional education. This limited access to nutritional education recourses could be due to numerous factors, like limited funding of rural schools compared to urban schools, and limited access to specialized staff (Guo, 2007). The lack of proper recourses could ultimately lead to worse dietary knowledge.

IV. Community level: The food environment can differ a lot between urban and rural areas. Urban areas might have more supermarkets with access to fresh products.

V. Public policy level: Policies can impact rural or urban areas differently. This could affect dietary behaviors differently between urban and rural areas.

The ecological Model of Health Behavior is unique in its approach to behavior change considering multiple levels of influence simultaneously. It acknowledges that individual behavior is influenced by broader contextual factors.

Research question

The proposed research aims to address the following research question: 'To what extent are urban-rural differences in overweight and obesity mediated by variations in dietary knowledge and awareness among individuals living in the Netherlands?' This question aims to address the complex influences between geographical location, dietary knowledge, and obesity trends since it has been clear that these factors are interconnected.

This study aims to shed light on disparities in dietary knowledge between urban and rural populations. This could help policymakers to focus on region-specific policies, making the policies more effective. Also, when dietary knowledge is contributing significantly to the urban-rural differences in overweight and obesity, dietary knowledge could be an important way to tackle the problem of increased rates of overweight and obesity. Furthermore, this research could be a foundation for future research. Underlying factors for disparities in dietary knowledge, such as socioeconomic, cultural, and environmental factors could be studied intensively. Also, the effectiveness of interventions related to improving dietary awareness could be researched. Lastly, while this study is focused on the Dutch population, the findings may also be relevant for other countries coping with similar disparities between urban and rural obesity rates.

The novelty of this study lies in its specific focus. There already is a significant number of studies on factors contributing to obesity and overweight. But the geographical focus on The Netherlands, with its unique sociocultural context and health system, underlines the novelty of this research. The dietary knowledge and obesity prevalence could differ significantly in other countries (Kroneman et al., 2016). To further explain these disparities; The Netherlands has a different education system than other countries. This, and other instances like policies and advertising, influences dietary knowledge. Furthermore, the urban-rural disparities in obesity and overweight are not researched as intensively yet and the focus on dietary knowledge is relatively new.

The study employs a well-suited methodological approach to examine the association between urban-rural differences in obesity and variations in dietary knowledge and awareness among individuals

residing in the Netherlands. Specifically, a logistic regression technique with imputation techniques to account for missing variables is employed to accurately predict their values. This approach mitigates potential biases and enhances the reliability and validity of the findings.

Hypotheses

Based on existing literature and the research question, the study proposes the following hypotheses:

A. It is predicted that rural areas, in comparison to urban areas, will exhibit a higher prevalence of obesity and overweight among individuals living in the Netherlands. This could be due to a range of factors. For example, limited access to freshly produced food and fewer physical activity resources.
B. It is predicted that there will be a negative correlation between dietary knowledge and rates of obesity and overweight. So, individuals with more dietary knowledge are expected to have lower rates of obesity and overweight. This could be because individuals with adequate dietary knowledge are more capable of making healthier food choices.
C. The third hypothesis suggests a mediation effect between urban-rural disparities and dietary knowledge. It is hypothesized that the negative relationship between dietary knowledge and obesity and overweight will be mediated by dietary knowledge.

This study aims to investigate the extent to which urban-rural differences in obesity and overweight are correlated with the level of dietary knowledge. The theoretical framework addressed the global and regional trends of obesity and overweight, with urban and rural differences highlighted. It examines the possible causes and effects of the disparities and provides information on the theoretical approaches of this research. Furthermore, the importance of these insights into dietary knowledge is emphasized.

Methodology

Data collection

'In this paper, I make use of data from the Longitudinal Internet Studies for the Social Sciences (LISS) panel administered by Centerdata' (Tilburg University, The Netherlands) (Centerdata, 2022)[1]. The LISS panel, designed for research, comprises 5,000 households. This grants a total of 7,500 individuals participating. The panel is randomly drawn from population registries. The panel uses probability sampling methods to ensure the representativeness of the general Dutch population. All these

[1] *This sentence is directly copied from the Centerdata website, this is one of their obligatory conditions for using their datasets.*

individuals were initially contacted through a letter in 2007. If there was no response to the letter, a follow-up with a telephone call or in some cases even a home visit is implemented. Even if individuals within the sample do not have access to a computer or the internet, they can still participate with the use of loan equipment.

In later periods, there were refreshment samples. These refreshments samples employ strategic oversampling techniques, where the researchers set a quota for each of the subgroups to ensure adequate representation. The first refreshment sample was in 2009. Now, with the cooperation of Statistics Netherlands, they had the ability to draw a non-random strategic sample. It was intended to improve representativeness by oversampling the groups which were harder to reach. Under-sampled groups were defined with the use of three variables: age, household type, and ethnicity. From October 2011 till May 2012, a new random sample was recruited again in the same way as the first sample. Lastly, from November 2013 till June 2014 there was a non-random sample selected again, to improve the representativeness the same way as in 2009.

Although there will always be over- or underrepresentation of certain groups within the LISS panel. By improving the representativeness with the non-random strategic samples, this sample will almost perfectly represent the Dutch population in terms of the variables used for sampling (age, household type, and ethnicity).

The datasets being used from the LISS panel, all collected through online questionnaires, for this research are:

I. Food choice: This dataset developed by van Kippersluis, H and Koc, H. was used to understand that the socio-economic differences between individuals provide a lot of explanatory power to understand differences in health behavior. Part one of this research was used as a choice experiment, while part two, the one being used in this research, was used to determine health knowledge and awareness. The date of the collection of the data was in May 2014.
II. Background variables: it is important for such a big data panel to keep track of the most important variables of their panel members. Thus, every month, the contact person of the household gets sent a questionnaire to keep track of changes in the household. Some of the variables tracked in this questionnaire are on the household level, such as the number of household members, the type of house they live in, and the income of the whole household. Some of the variables tracked in this questionnaire are on the individual level, such as gender, age, and civil state.
III. Health: In July 2015, a questionnaire on health was held to get more knowledge of the health of individuals within the panel. This questionnaire is part of the 8th wave of the LISS Core Study.
IV. Social integration and leisure: In October 2015, this questionnaire was conducted as part of the 8th wave of the LISS Core Study. The dataset contains information about the social integration

and leisure time of the individuals participating in the panel. This dataset is used in this study to obtain control variables.

V. Personality: This dataset, as part of the 7th wave of the LISS core study, was held in December 2014. This dataset contains mainly information about personal characteristics. This dataset is used to obtain control variables for this study.

These five datasets were all merged into one big dataset. The methods for this merger are discussed later in this paper. The sample sizes of the datasets are displayed in Table 5.1. Additionally, important characteristics of the individuals within the sample are displayed in Appendix A.

Table 5.1: Sample size of the datasets used in the study.

Name of dataset	*Food choice (1)*	*Background variables (2)*	*Health (3)*	*Social integration and leisure (4)*	*Personality (5)*
Number of household members (individuals)	3,527	12,518	7,126	7,037	7,739
Non-respons	370	0	1,117	995	1,178
Respons:	3,157	12,518	6,009	6,042	6,561
Complete	3,147	7,262	5,975	5,930	6,442
Incomplete	10	5,256	34	112	119

Information on the responses of the different used datasets in this research.

Statistical methods

I. Logistic regression

In this research, logistic regression is used. It is commonly used to predict the probability of a binary outcome (Kleinbaum, 1994). In this case, the binary outcome is if the individual is overweight or not. The linear regression equation is transformed into a logistic regression with a range of values between zero and one.

II. Assumptions

Binary logistic regression requires the dependent variable to be binary: The variable 'overweight' is binary and takes the value '1' if the individual is overweight and the value '0' if the individual is not overweight. Secondly, logistic regression requires the independence of observations. Every individual self-reported the values of the variables in the questionnaires used in this study. Although there could still be dependence, because of similar social or cultural factors affecting their responses, for the sake of simplicity and due to the relatively small, expected biases associated with the self-reported variables, the assumption of independence among observations is considered as true. Moreover, it is critical to check for the assumption of linearity between the logit of the outcome variable and the independent variable (DeMaris, 1995). For each continuous independent variable, a new variable was created by squaring the independent variable. Then a logistic regression with both the original and squared

independent variables was conducted. After performing the analysis, none of the newly created, squared variables was significant as is displayed in Table 5.2 (significant if $p < 0.05$). This suggests that there is not enough evidence to reject the linearity assumption. This technique was chosen due to its simplicity and ease of implementation in STATA.

Table 5.2: Results of the method to check the linearity assumption of this study.

Variable	Coefficient* (1)	P-value (2)	95% confidence interval (3)	
Age squared	0.000	0.000	-0.00000503	0.00000701
Number of household members squared	0.001	0.302	-0.0017042	0.0001291
Grade dietary knowledge squared	0.000	0.096	-0.0004974	0.0008316
Log income squared	0.002	0.763	-0.0037722	0.0003032

*The coefficients are rounded to three decimal points.

For checking the multicollinearity of the logistic regression, the variance inflation factor (VIF) of the variables is checked. The average VIF is 2.83 and is well below the frequently used threshold of 5 (Craney & Surles, 2002). Although there are some variables in the regression with high VIF (above 5), the variables of main interest ('Overweight', 'Grade' and 'Urban') are well below 2 and are not concerning in terms of multicollinearity. Thus, multicollinearity among these key variables is unlikely to be a significant issue in interpreting the results of this study. Lastly, the necessary sample size for the logistic regression depends on several factors. But for simplicity, in this study, a rule of thumb is used. The rule of thumb is a minimum of ten events per variable in the regression (Wilson Van Voorhis & Morgan, 2007). The logit regression of this study consists of 43 variables in total, so to have a large enough sample size, a minimum of 430 events are needed. After examining the 'overweight' outcome variable, the number of events, in this case, the number of people being overweight, equals 2,478 (displayed in Table 5.3). Additionally, the dataset includes 2,399 non-events, meaning individuals who are not classified as overweight. This provides a balanced sample for analysis. Therefore, the sample size assumption would not be violated in this case.

Table 5.3: Number of observations for the event and non-event:

Number of observations for individuals who are not overweight.	2,399
Number of observations for individuals who are overweight.	2,478

III. Variables

To approach a causal interpretation of the probability of being overweight, due to urban-rural differences in dietary knowledge, it is crucial to account for potential confounding variables. Although including all variables is not feasible, it is important to incorporate the most influential control variables into the regression. Furthermore, the time for tracking the variables is important. To approach for a

causal interpretation, the independent variable and the mediator variable are tracked earlier than the dependent variable, this is the case in this study. What follows now is a list of the reasonings for adding the variables to the regression.

I. Dependent variable: Probability of being overweight: An individual is overweight with a BMI bigger than 25 and is thus a binary variable.

II. Independent variable: The independent variable, in this case, is the urbanity of the area where the individual lives. If an individual is living in an urban or rural area is determined using the cutoff point of 1,000 addresses per square kilometer, as frequently used by the CBS (Central Bureau of Statistics) (CBS, n.d.-b).

III. Mediator variable: The mediator variable of this study is the dietary knowledge of the individual. The grade is obtained from a multiple-choice questionnaire on dietary knowledge. The answer is considered wrong if the individual gave the wrong answer or when the individual filled in: 'I do not know'.

IV. Confounding variables (control variables)

1) Age: Adding the age of the individual to the regression is important as it could be influential in differences in dietary knowledge (For example, someone who is older has more experience in making healthy dietary choices) and, on the other hand, age could be influential to BMI levels as well.
2) Gender: The biological sex of an individual might affect weight-related behavior and knowledge (Crosnoe, 2007). In addition, gender could affect the consequences of obesity, for example, women have a higher risk on obesity-related cancers (Crosnoe, 2007).
3) The number of household members: Household composition is probable to influence the chances of being overweight. The availability of resources could be different with a different household composition, to give an example. Household members often share food practices and habits and could thus be correlated with dietary knowledge as well (Barnes et al., 2013).
4) Civil status: A married couple has 'built-in' social support. Married couples could benefit from each other's support to contribute to physical activities (Wilson, 2012). Also, it could encourage healthy dietary choices. One could support their partner in learning more about healthy food if they value eating healthy a lot (Horwath, 1989). Or vice versa, if one partner values an unhealthy diet, it could lead to the other partner switching to an unhealthier diet.
5) Primary occupation: The primary occupation of an individual is a cause of disparities in activity levels. Working individuals tend to have more physical activity on average than non-working individuals. Economic opportunities can differ significantly between urban and rural areas. Urban areas generally provide a wider range of jobs, including jobs associated with higher education levels. This could in the long run affect dietary knowledge

and the probability of being overweight (Gu et al., 2014). The primary occupation is categorized as studying, housekeeping, retired, (partially) disabled, working, seeking work and other occupations.

6) Income: Individuals with a higher income generally have more purchasing power. They can afford a wider variety of healthy and nutritious food. In urban areas, supermarkets offer a wider variety of healthy and nutritious food, thus, in urban areas, high-income individuals can buy healthier nutritious food (Kim & von dem Knesebeck, 2018). In addition, unhealthy food options are generally cheaper than healthy food options (Temple & Steyn, 2011). There has been done a logarithmic transformation on the income variable in this study. The reason for this is big outliers of income.
7) Education: Highly educated Individuals may have had more exposure to health-related information, including information about nutrition. This could lead to a better understanding of maintaining a healthy weight and having a healthy diet. Furthermore, urban areas generally have more education availability, providing an opportunity to advance knowledge, including dietary knowledge (Devaux et al., 2011).

   Education is categorized into four levels in this research paper. Primary education consists of a combination of uneducated, primary school and individuals who did not yet start an education. Secondary education consists of a combination of 'vmbo' and 'havo'/'vwo' ('vmbo', 'havo', and 'vwo' are different levels of secondary education in the Netherlands). Vocational Education is a combination of 'mbo' (intermediate vocational education) and 'hbo' (higher vocational education). And lastly, higher education consists only of individuals with a university degree.
8) Ethnicity: Ethnicity often comes together with cultural dietary practice. Different cultures have different dietary practices and patterns, which influence dietary knowledge and behavior (Caprio et al., 2008). Also, urban, and rural areas often have different compositions in terms of the ethnicity of the populations. Ethnicity in this study is defined as fully Dutch, first-generation Western background, first-generation non-Western background, second-generation Western background, and second-generation non-Western background.
9) Smoking: Smoking has an association with various metabolic changes, which could affect weight management (Dare et al., 2017). Smokers have different health awareness than non-smokers (Woodward et al., 1994), and this could impact dietary knowledge. Lastly, income and education levels differ between urban and rural areas, this could contribute to disparities in smoking rates between these areas.
10) Alcohol: Alcoholic beverages often contain a high number of calories, so excessive alcohol consumption contributes to weight gain (Schröder et al., 2007). Also, when under the influence of alcohol, individuals may be more likely to choose unhealthy food options

(Obst et al., 2018). Alcohol consumption differs between different socio-economic factors such as income and education level. Income and education levels differ as well between urban and rural areas. Urban areas generally offer more campaigns towards reducing excessive alcohol consumption, which influences dietary knowledge as well.

Alcohol consumption, in this study, is categorized into different groups: Frequent alcohol consumption drink five or more days a week, regular alcohol consumption (drinks one to four alcoholic drinks a week), occasional alcohol consumption (drinking approximately one alcoholic drink per month), and infrequent/no alcohol consumption (drinks once or twice a year or does not drink at all). The decision to categorize the data on alcohol consumption into four groups was made to simplify the analysis and to ensure each category contained a sufficient number of observations.

11) Physical activity: Physical activity helps to burn calories; this is essential to manage body weight. Urban and rural areas are expected to differ in terms of the availability of physical activities (Sandercock et al., 2010). Urban areas often have more access to health education, learning individuals about the importance of physical activity and dietary behavior and patterns. In this study, the level of physical activity is defined as how many days a week an individual performs sport. Four groups are defined consequently: No sport (the individual performs no sport), low activity (the individuals perform sports mostly two days a week), moderate activity (this category includes individuals who perform sports 3 to 5 days a week), and high-activity individuals (perform sports 6 or 7 days a week).
12) Diet: Individuals who actively follow a diet are likely to have better dietary knowledge (Lissner et al., 2000). They might have looked up information about nutrition, portion control or dietary approaches before committing to the diet. Following a diet that involves consuming healthier food, urban areas generally have better access to these kinds of healthier food options (Wang et al., 2010).
13) Happiness: Happiness is often associated with adopting healthier lifestyle choices. Individuals who are happy may have more motivation to engage in regular exercise and make better nutritional choices (Zhang & Chen, 2019). These healthier behaviors could contribute to weight management and a better understanding of dietary knowledge (Katsaiti, 2012). Every individual in the questionnaire gave their happiness a grade on a scale of 1 to 10. To improve the simplicity of the regression, this is converted as a binary variable. The individual is considered ‘happy’ if the grade is over 5 and ‘unhappy’ if the grade is 5 or lower.
14) Satisfaction of social contacts: The satisfaction of social contacts reflects the level of social support individuals receive from their social networks. Social support plays a crucial part in different health aspects, including weight management and dietary knowledge (Dierk et al., 2006). Furthermore, the satisfaction of social contacts could influence eating habits

(Higgs & Thomas, 2016). For example, through shared meals and social interaction. Finally, the network size and diversity of the network differ between urban and rural populations. The variable is defined in the same way as the 'Happiness' variable.

The variable list with the most important characteristics of every variable is included in Appendix B.

Data analysis

The data analysis began with the process of data cleaning and preparation. Given the complexity of the data, which involved five different datasets, a necessary step in the methodology was merging these distinct datasets into one comprehensive dataset for analysis. An inner join technique was used to consolidate four of the five datasets. The last dataset called 'Food Choice' was merged using an outer join technique. This process ensured that only the entries with complete records in all the datasets (except for the 'Food choice' dataset) were included in the final dataset.

Once the data was merged, because of the outer-join method for (only) the 'Food Choice' dataset, it became apparent that some variables for dietary knowledge, which is a variable from the 'Food Choice' dataset, had missing values. To ensure the integrity of the dataset and to lower the impact of the missing data, an imputation technique was used. The technique used is known as 'multiple imputations by chained equations'. Unlike simpler techniques such as mean imputation, which substitutes a single value for the missing data, multiple imputation recognizes the inherent uncertainty in the process of estimating missing values (White et al., 2011). This is accomplished by creating multiple copies of the dataset and replacing missing data with imputed values. A key assumption of multiple imputations is that the data are missing at random. Multiple imputations are widely accepted in statistical research, as it provides a robust method to address missing data. It may offer a considerable improvement over simpler methods and ensures that the missing data does not introduce a significant bias and was thus chosen as the imputation method in this study.

To assess the 'missing variables missing at random' assumption for multiple imputations by chained equations, the relationship between a missing value indicator ('1' if data is missing and '0' if not missing) and the other observed variables is examined (Seaman et al., 2013). The method of doing this is logistic regression, the missing data indicator is the dependent variable, and the other observed variables are the dependent variable. This is often referred to as the 'missingness model'. This method is essentially checking whether the missingness is dependent on other observed variables. The results of this logistic regression are summarized in Appendix C. Some of the coefficients are statistically significant ($p<0.05$). Finding a few statistically significant predictors in the missingness model does not necessarily invalidate the 'missing at random' assumption. However, it will invalidate the stronger

assumption: ‘missing completely at random’. The ‘multiple imputations by chained equations method’ can still be used because the ‘missing at random’ assumption is still feasible. It is important to note that testing for the ‘missing at random’ assumption is generally impossible because the pattern of the data is not known. If the predictors of missingness are unobserved, then the data could be missing ‘not at random’ resulting in biased results. The logistic regression method is, in this case, just used to provide valuable insights into the patterns of missingness in the data and cannot validate or invalidate the ‘missing at random’ assumption.

Besides the missing values for the variable for dietary knowledge, the income variable contained missing values as well. These values are replaced with the median of the logarithmic income due to its simplicity and ease to understand. But while this method is easily applicable, it could introduce bias. Individuals with missing values for other variables were removed from the sample.

The primary statistical method used in this research was logistic regression analysis. Logistic regression provides interpretability in terms of odds and probabilities. Being overweight, in the case of this research, is transformed into a binary variable, logistic regression is a good option for this instance. Also, logistic regressions seem to be relatively robust to outliers and require smaller sample sizes in comparison to other machine learning algorithms (MENARD, 1995). Lastly, logistical regressions are well-established. There are many recourses available to understand and apply logistic regressions. The logistical regression model incorporated a comprehensive list of control variables to minimize the potential for confounding effects and to get as close as possible to estimating a causal effect.

Mediation

Mediation analysis was employed using an additional regression model to elucidate the indirect effect of urban-rural differences on overweight and obesity through dietary knowledge. This analysis was critical in testing the hypothesis that dietary knowledge mediates the relationship between urban-rural areas and overweight rates. This methodology, therefore, allowed for a nuanced understanding of the complex interplay between geographical location, dietary knowledge and overweight.

The method employed was based on the procedures outlined by Baron and Kelly (Zhao et al., 2010). Their approach typically follows three steps:

I. First, the association between the independent variable (urban-rural differences) and the dependent variables (overweight rates) is examined and should be significant. This is done with the use of a logistical regression analysis.

II. Next, the relationship between the independent variable and the mediator variable (dietary knowledge) is assessed. This stage involves regression analysis and should be significant.

III. Finally, when both the independent variable and the mediator variable are used for a prediction of the dependent variable using regression, the mediator should remain significant. Furthermore, the correlation between the independent and the dependent variable is important in this step. This path between the independent variable and the dependent variable is greatly reduced or insignificant if the mediator variable carries the influence of the independent variable on the dependent variable.

These steps allow for a statistical confirmation of the hypothesized mediation role of dietary knowledge in the relationship between urban-rural differences in overweight. It is still important to note that the Baron and Kelly method for mediation, while still often used, has been supplemented by other methods (Collins et al., 1998). These methods provide more power and flexibility. Nonetheless, the Baron and Kelly method provides a straightforward view of mediation.

Model validation and Software

Various model diagnostics were carried out to ensure the validity and robustness of the results. This included checking multicollinearity, heteroscedasticity, and model fit. Furthermore, the statistical software being used in this research is STATA, developed by StataCorp. The version used is version 17.0.

A. Heteroskedasticity: Heteroskedasticity is a situation where the variance of errors of the model's residuals is not constant across all levels of independent variables. Because the statistical method used in this study is logistic regression, heteroscedasticity is not usually a concern. The assumption of equal variances (homoscedasticity) in the errors does not apply as it does for linear regression.

B. Model fit: To research the logistic regression model fit, a proportion of correct predictions (or classification accuracy rate) was calculated (ARONOFF, 1982). The predicted outcomes are compared to the actual outcomes. The model correctly predicted overweight status 59.11% of the time. Considering the complexity of the factors influencing obesity, this prediction accuracy could be considered reasonable. It is important to note the limitations of using classification accuracy as the measure of model fit. It does not consider how close the predicted outcomes are to the actual outcomes.

Overall, the data analysis strategy was designed to provide a comprehensive and accurate understanding of the research question: 'To what extent are urban-rural differences in overweight and obesity mediated by the variations in dietary knowledge and awareness among individuals living in the Netherlands?'

Quality assurance

The quality of the dataset and the quality of the logistic regression are crucial. Every single variable is checked on outlying observations. If such outliers exist, a fitting solution to not harm the internal validity of the sample will be chosen. The only, not plausible variables observed were two data points of the individual's height and some major outliers in the income variable. The values of the height were simply removed (before the imputation). The sample size is relatively large, and removing these two variables did not change any of the other variables significantly after looking into the data distribution before and after the removal. Histograms, boxplots, and summary statistics were used for this purpose. The income variable is transformed using a logistic transformation, minimizing the impact of the outliers.

With the merge of combining five datasets into one, the sample size is reduced to a smaller subset. It is important to check if this smaller sample is representative of the larger, randomized, sample of the individual datasets (Ramsey & Hewitt, 2005). Every variable of this study is checked using a t-test. The results of every t-test are included in the appendix (appendix DI). In addition to the t-test, the percentages of every category of the categorical variables are compared between the merged dataset and the full dataset (appendix DII). The results, in both the t-test and the percentages method, revealed significant differences between the original and the merged dataset. This suggests that the cleaning and merging of the datasets may have caused the merged dataset to not be representative of the bigger dataset. It is important to consider this limitation when interpreting the results of the study. Lastly, there was missing data in the dataset. These were treated using imputation or in some cases, the missing values were removed, as described in detail earlier.

Results

Regression

This study explores the relationship between urban-rural differences, dietary knowledge, and overweight among individuals in the Netherlands. The findings shed light on the intricate dynamics of these factors and contribute valuable insights on this topic.

The results for the logistic regression with the imputation technique are attached as Appendix E. Relationships between overweight and control variables, in this case, are considered significant for the

variables with a p-value less than 0.05 (indicated by*). The following list illustrates the relationships of these significant relationships.

I. Gender: The exponentiated coefficient, also known as the odds ratio (OR), is in this case 0.628. Thus, being female is associated with a 37.2% (1-0.628) decrease in the odds of being overweight compared to being male, keeping the other variables constant.
II. Age: When controlling for other variables in the model, the odds of the dependent variable (overweight) increase by approximately 1.7% for each unit increase in age.
III. Diet: The 'diet' variable has a relatively large negative coefficient. In terms of odds ratios, when an individual is on a diet, again, if the other variables remain constant, their chance of being overweight decreases by 37.8%.
IV. Civil state: The civil state dummies for 'married' and 'widow or widower' are significant. Both being married (OR: 1.303) and being a widow or widower (OR: 1.452) are positively correlated with being overweight, assuming other variables remain constant. 'Never been married' is in this case the reference category.
V. Occupation: 'Disability' (OR: 1.821), 'Working' (OR: 1.603) and 'Seeking for work' (OR: 1.568) also exhibit significant positive relationships with the dependent variable, holding other variables constant. Here, 'Other occupation' is used as the reference category.
VI. Education: The 'High education' (OR: 0.579) and 'Primary education' (OR: 1.676) variables are significantly associated with the dependent variable. Primary education has a negative relationship, while high education has a positive relationship if other variables are held constant. With this categorical variable, 'Vocational education' is used as a reference category.
VII. Drinking variables: 'Frequent drink' (OR: 0.622) and 'Regular drink' (OR: 0.787) are significantly associated with the dependent variable. Drinking frequently has a significant negative relationship with the dependent variable if the other variables are held constant.
VIII. Sport dummy variables: All the sports variables are significant. If the individual does not perform sports (OR: 1.970), this has a positive correlation with being overweight. If the individuals perform sports more frequently (Occasional sports OR: 1.599 & Regular sports OR: 1.476), the positive correlation decreases. This only holds when the other variables are held constant. 'Frequent sports' is, in this case, used as the reference category.
IX. Smoking variables: Lastly, if the individual never smoked, this had a negative correlation with being overweight, under the assumption all the other variables remain constant (OR: 0.654). This is also the case when the individual is a present smoker (OR: 0.574). 'Past smoke' is used as the reference category.

The first hypothesis, proposing a negative correlation between being overweight and urbanity is rejected based on the results. This is because the variable for 'urban' is not statistically significant. There is no

sufficient evidence to assert that a relationship between urbanity and overweight status exists in the population used in the study.

The second hypothesis proposed a negative relationship between dietary knowledge and overweight rates. However, this hypothesis could, just like the first hypothesis, not be confirmed and is thus rejected, due to the lack of a statistically significant relationship between dietary knowledge and overweight in the dataset.

Mediation

The mediation technique of Baron & Kenny is used to examine the role of dietary knowledge as a potential mediator in the relationship between urban-rural differences and overweight rates. The results of the performed mediation steps are included in the Appendices. First, after testing the relationship between the dependent variable (overweight) and the independent variable (urbanicity) (Appendix FI), it can be concluded that there is no statistically significant relationship between these two variables. Secondly, after testing the relationship between the independent variable and the mediator variable (dietary knowledge) (Appendix FII), it can, again, be concluded that there is no statistically significant relationship between these two variables. Lastly, in the logistic regression with the dependent, independent, and mediator variables included (Appendix E), the correlation coefficient between the independent and the dependent variable did decrease with the inclusion of the mediator variable in comparison to the correlation coefficient between the independent and the dependent variable without the inclusion of the mediator variable, although the coefficient is not considered statistically significant in both regressions ($P<0.05$).

Based on the results from the analysis, there is no evidence found of a mediation effect of dietary knowledge on the relationship between urban-rural differences and overweight. The first, second and third regressions used in the mediation analysis did not result in statistically significant relevant coefficients.

In conclusion, this study set out to explore the relationship between urban-rural differences, dietary knowledge, and overweight rates among individuals in The Netherlands. Despite extensive investigation and a comprehensive list of variables, the results did not provide sufficient evidence to support or reject the hypotheses. Nonetheless, these results still contribute valuable insights. The complex nature of these relationships is underscored and the need for further research is highlighted.

Discussion

In this study, the aim was to research the mediating role of dietary knowledge on the effect of urban-rural differences on overweight rates in individuals in the Netherlands. There was not enough evidence to support the hypotheses. This may suggest that interventions, focused solely on improving dietary

knowledge, to address urban-rural differences in overweight, may not be sufficient. The role of other factors, such as socioeconomic status, availability of healthy foods, and opportunities for physical activity, should also be considered in shaping public health policies and interventions.
Nevertheless, these findings still provide valuable insights into the complexity of the interplay of these variables and the potential influence of variables not included in the model. For instance, psychological effects and broader social-cultural factors may be important as well in the determination of an individual's weight status.

In the conclusion section of this paper, some interpretation was done on the remaining control variables used in the model. Although this could still provide some information, it is generally not possible to make valid conclusions because control variables for these relationships are missing. For example, the correlation between gender and overweight requires different control variables than the control variables used in this study for the relationship between urbanicity and overweight.

Many countries, including the Netherlands, continue to struggle with widespread overweight and obesity issues. Further research is still needed as the potential importance of education and awareness in promoting healthier behaviors is still not entirely clear. The absence of results in line with the initial hypothesis suggests that there may be no substantial mediation effect of dietary knowledge on the association between overweight and urbanicity. However, numerous other explanations are feasible. The variables are self-reported within the LISS panel. Self-reporting introduces the potential for biases, such as women tending to underestimate their weight while men may overestimate their weight. This will result in an under- or overrepresentation of overweight individuals in the dataset and will result in biased results. Moreover, the relatively short follow-up period employed in this study might have limited the ability to capture significant correlations. Although this study did not involve an intervention, exploring variables over an extended period of time may reveal more pronounced effects. Furthermore, the sample size was considered large enough in this study with the use of the rule of thumb, however, the desired sample size is dependent on a lot of factors and could still be too small. In addition, there was a relatively large number of missing values for the mediator variable (dietary knowledge), although this was assessed using an imputation method, this still introduces bias and could influence the mediating relationship of dietary knowledge, potentially resulting in non-statistically significant results. Lastly, multicollinearity could still be problematic. The method of checking for multicollinearity in this study was simple. There could still be multicollinearity which was not obtained from the Variance Inflation Factor test. It is crucial to acknowledge these limitations and explore them further in future research.

Existing literature on the association between urbanicity and overweight is mixed. Different studies have reported contrasting or inconclusive results (Congdon, 2019; Johnson & Johnson, 2015). Factors

like methodology, geographical location and differences in the sample used in the studies could be possible explanations for discrepancies. In the research by P. Congdon, urbanicity is more specified, with additional variables such as walkability, available amenities, and food environment. This could be a reason for the different results of P. Congdon and this research. Furthermore, the study from 2015 by Johnson & Johnson provided valid results. However, this study was conducted in the United States. The United States differs significantly in terms of food availability, built environment characteristics, and transportation infrastructure. These environmental factors play a crucial role in shaping dietary habits, potentially leading to varying associations between urbanicity and overweight.

A key limitation of this study pertains to the measures used to represent the variables of interest. For instance, the variable 'Origin' was classified only in terms of first- and second-generation status and Western or non-Western background. However, this might not capture all the complexities of one's cultural background. Simplification is often necessary for analytical purposes, although it is still important to consider possible implications with simplification.

In future studies, one critical aspect to consider is the sample size. Although the study benefited from a reasonable sample size, many individuals had missing data for dietary knowledge. The use of imputation can partially rectify this problem, but it is still a limiting factor for this study and is introducing bias. Furthermore, there is no proof the data is missing at random. As this is a key assumption for using multiple imputations by chained equations, this is an important implication to acknowledge. In addition, some individuals did not want to show their monthly income, resulting in missing variables. These missing variables were replaced with the median of the income. Although this technique is generally simple and easy to implement and understand, it could introduce bias. Because the different datasets were merged using an inner-join technique to limit the number of missing variables, the sample size became smaller and potentially unrepresentative. Consequently, even if a significant relationship was found in this research, it might not generalize to the broader Dutch population, given the specific characteristics of the sample. Future studies might benefit from incorporating strategies to maximize data completeness and to ensure a representative sample to provide more robust and generalizable findings.

Finally, although the current study did not find significant relationships between urban-rural differences, dietary knowledge, and overweight rates, it should not deter future research in this area. In the context of the wider field, this study adds to the mixed findings on the role of dietary knowledge in urban-rural differences in overweight and obesity. This underscores the need for further research to clarify these relationships, while also providing a foundation for further research.

Conclusion

This study aimed to answer the research question: 'To what extent are urban-rural differences in overweight and obesity mediated by variations in dietary knowledge and awareness among individuals living in the Netherlands?' Despite the relevance of this research question on economic, health and environmental aspects, the study did not find statistically significant evidence supporting the mediation role of dietary knowledge, rejecting the third hypothesis. Considering the findings from this study, it is necessary to acknowledge that all three of the proposed hypotheses have been refuted. Firstly, the anticipated higher prevalence of obesity in rural areas compared to urban areas in The Netherlands could not be statistically confirmed. Similarly, the proposed negative relation between dietary knowledge and overweight did not present statistically significant results.

The analysis did reveal significant correlations between overweight status and a range of factors, including gender, age, diet, civil state, occupation, education, alcohol consumption, physical activity, and smoking. However, these relationships were not controlled for potential confounding variables. These findings shed light on the multifactorial nature of overweight and obesity, highlighting the need for comprehensive strategies to address this public health issue. These results and findings could serve as the foundation of further investigation and provide valuable insights into the intricate dynamics of overweight and obesity, urban-rural differences, and dietary knowledge.

However, it is still important to acknowledge the limitations of the study. The sample size was smaller due to the use of an inner joint technique for the merging of some datasets. Additionally, this may not have provided a representative sample of the Dutch population. Additionally, missing data, which was addressed through multiple imputations or median imputations, might have influenced the results. Future studies would benefit from larger, more complete, and more diverse samples, as well as more detailed categorization of variables to better capture the complexity of the issues at hand. Furthermore, future studies could explore other mediators or confounding factors that were not considered in this study. Despite the limitations, this research represents a meaningful step forward in understanding the complex interrelations that influence overweight and obesity rates.

Reference list

Abu-Baker, N. N., Abusbaitan, H. A., Al-Ashram, S. A., & Alshraifeen, A. (2021). The Effect of Health Education on Dietary Knowledge and Practices of Pregnant Women in Jordan: A Quasi-Experimental Study. *International Journal of Women's Health*, *Volume 13*, 433–443. https://doi.org/10.2147/IJWH.S303568

Afshin, A., Forouzanfar, M. H., Reitsma, M. B., & Sur, P. (2017). Health Effects of Overweight and Obesity in 195 Countries over 25 Years. *New England Journal of Medicine*, *377*(1), 13–27. https://doi.org/10.1056/NEJMoa1614362

ARONOFF, S. (1982). CLASSIFICATION ACCURACY. A USER APPROACH. *PHOTOGRAMM. ENG. REMOTE SENSING*, *48*(8), 1299–1307.

Avila, C., Holloway, A. C., Hahn, M. K., Morrison, K. M., Restivo, M., Anglin, R., & Taylor, V. H. (2015). An Overview of Links Between Obesity and Mental Health. *Current Obesity Reports*, *4*(3), 303–310. https://doi.org/10.1007/s13679-015-0164-9

Barnes, M. G., Smith, T. G., & Yoder, J. K. (2013). Effects of household composition and income security on body weight in working-age men. *Obesity*, *21*(9), E483–E489. https://doi.org/10.1002/oby.20302

Bertakis, K. D., & Azari, R. (2005). Obesity and the Use of Health Care Services. *Obesity Research*, *13*(2), 372–379. https://doi.org/10.1038/oby.2005.49

Best, D., & Bhattacharya, S. (2015). Obesity and fertility. *Hormone Molecular Biology and Clinical Investigation*, *24*(1). https://doi.org/10.1515/hmbci-2015-0023

Bodor, J. N., Rice, J. C., Farley, T. A., Swalm, C. M., & Rose, D. (2010). The Association between Obesity and Urban Food Environments. *Journal of Urban Health*, *87*(5), 771–781. https://doi.org/10.1007/s11524-010-9460-6

Camilleri, M., Malhi, H., & Acosta, A. (2017). Gastrointestinal Complications of Obesity. *Gastroenterology*, *152*(7), 1656–1670. https://doi.org/10.1053/j.gastro.2016.12.052

Caprio, S., Daniels, S. R., Drewnowski, A., Kaufman, F. R., Palinkas, L. A., Rosenbloom, A. L., & Schwimmer, J. B. (2008). Influence of Race, Ethnicity, and Culture on Childhood Obesity: Implications for Prevention and Treatment. *Diabetes Care*, *31*(11), 2211–2221. https://doi.org/10.2337/dc08-9024

Carroll, P., Benevenuto, R., & Caulfield, B. (2021). Identifying hotspots of transport disadvantage and car dependency in rural Ireland. *Transport Policy*, *101*, 46–56. https://doi.org/10.1016/j.tranpol.2020.11.004

CBS. (n.d.-a). *Overweight and obesity*. Retrieved June 14, 2023, from https://www.cbs.nl/en-gb/news/2022/45/less-overweight-and-obesity-among-children-of-higher-educated-parents/overweight-and-obesity

CBS. (n.d.-b). *Rural areas*. Retrieved June 2, 2023, from https://www.cbs.nl/en-gb/news/2010/23/narrowing-gap-between-number-of-facilities-available-in-urban-and-rural-areas/rural-areas

CBS. (2022a). *How many adults are overweight?* https://longreads.cbs.nl/the-netherlands-in-numbers-2022/how-many-adults-are-overweight/#:~:text=In%202021%2C%20around%20half%20of,while%201%20percent%20were%20underweight

CBS. (2022b, May 16). *Tijd besteed aan beweegactiviteiten, 2017-2021*.

Centerdata. (2022). *Sample and recruitment*. https://www.lissdata.nl/about-panel/sample-and-recruitment#overlay-context=about-panel

Centers for Disease Control and Prevention (CDC). (2007). Prevalence of regular physical activity among adults--United States, 2001 and 2005. *MMWR. Morbidity and Mortality Weekly Report*, *56*(46), 1209–1212.

Cheah, & Lynette W. (2010). *Cars on a diet : the material and energy impacts of passenger vehicle weight reduction in the U.S.*

Christakis, N. A., & Fowler, J. H. (2007). The Spread of Obesity in a Large Social Network over 32 Years. *New England Journal of Medicine*, *357*(4), 370–379. https://doi.org/10.1056/NEJMsa066082

Cohen-Cole, E., & Fletcher, J. M. (2008). Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic. *Journal of Health Economics*, *27*(5), 1382–1387. https://doi.org/10.1016/j.jhealeco.2008.04.005

Collins, L. M., Graham, J. J., & Flaherty, B. P. (1998). An Alternative Framework for Defining Mediation. *Multivariate Behavioral Research*, *33*(2), 295–312. https://doi.org/10.1207/s15327906mbr3302_5

Congdon, P. (2019). Obesity and Urban Environments. *International Journal of Environmental Research and Public Health*, *16*(3), 464. https://doi.org/10.3390/ijerph16030464

Conner, M., & Norman, P. (2015). Predicting and Changing Health Behaviour: Research and Practice with Social Cognition Models. In *Predicting and Changing Health Behaviour: Research and Practice with Social Cognition Models* (third edtion, pp. 225–234). McGraw-Hill Education.

Craney, T. A., & Surles, J. G. (2002). Model-Dependent Variance Inflation Factor Cutoff Values. *Quality Engineering*, *14*(3), 391–403. https://doi.org/10.1081/QEN-120001878

Crosnoe, R. (2007). Gender, Obesity, and Education. *Sociology of Education*, *80*(3), 241–260. https://doi.org/10.1177/003804070708000303

Daniel, C. (2020). Is healthy eating too expensive?: How low-income parents evaluate the cost of food. *Social Science & Medicine*, *248*, 112823. https://doi.org/10.1016/j.socscimed.2020.112823

Dare, S., Mackay, D. F., & Pell, J. P. (2017). Correction: Relationship between Smoking and Obesity: A Cross-Sectional Study of 499,504 Middle-Aged Adults in the UK General Population. *PLOS ONE*, *12*(2), e0172076. https://doi.org/10.1371/journal.pone.0172076

Day, K., Alfonzo, M., Chen, Y., Guo, Z., & Lee, K. K. (2013). Overweight, obesity, and inactivity and urban design in rapidly growing Chinese cities. *Health & Place*, *21*, 29–38. https://doi.org/10.1016/j.healthplace.2012.12.009

de Heredia, F. P., Gómez-Martínez, S., & Marcos, A. (2012). Obesity, inflammation and the immune system. *Proceedings of the Nutrition Society*, *71*(2), 332–338. https://doi.org/10.1017/S0029665112000092

Dekker, L. H., Rijnks, R. H., Strijker, D., & Navis, G. J. (2017). A spatial analysis of dietary patterns in a large representative population in the north of The Netherlands – the Lifelines cohort study. *International Journal of Behavioral Nutrition and Physical Activity*, *14*(1), 166. https://doi.org/10.1186/s12966-017-0622-8

DeMaris, A. (1995). A Tutorial in Logistic Regression. *Journal of Marriage and the Family*, *57*(4), 956. https://doi.org/10.2307/353415

Devaux, M., Sassi, F., Church, J., Cecchini, M., & Borgonovi, F. (2011). Exploring the Relationship Between Education and Obesity. *OECD Journal: Economic Studies*, *2011*(1), 1–40. https://doi.org/10.1787/eco_studies-2011-5kg5825v1k23

Diclemente, R. J., Crosby, R. A., & Kegler, M. C. (2002). *Emerging Theories in Health Promotion Practice and Research: Strategies for Improving Public Health*.

Dierk, J.-M., Conradt, M., Rauh, E., Schlumberger, P., Hebebrand, J., & Rief, W. (2006). What determines well-being in obesity? Associations with BMI, social skills, and social support. *Journal of Psychosomatic Research*, *60*(3), 219–227. https://doi.org/10.1016/j.jpsychores.2005.06.083

Dunn, R. A. (2010). The Effect of Fast-Food Availability on Obesity: An Analysis by Gender, Race, and Residential Location. *American Journal of Agricultural Economics*, *92*(4), 1149–1164. https://doi.org/10.1093/ajae/aaq041

Erratum to: The urban–rural education gap: do cities indeed make us smarter? (2021). *Journal of Economic Geography*, *21*(5), 715–715. https://doi.org/10.1093/jeg/lbab020

Evans, B., Crookes, L., & Coaffee, J. (2012). Obesity/Fatness and the City: Critical Urban Geographies. *Geography Compass*, *6*(2), 100–110. https://doi.org/10.1111/j.1749-8198.2011.00469.x

Flegal, K. M. (2010). Prevalence and Trends in Obesity Among US Adults, 1999-2008. *JAMA*, *303*(3), 235. https://doi.org/10.1001/jama.2009.2014

Franck, C., Grandi, S. M., & Eisenberg, M. J. (2013). Agricultural Subsidies and the American Obesity Epidemic. *American Journal of Preventive Medicine*, *45*(3), 327–333. https://doi.org/10.1016/j.amepre.2013.04.010

Frank, A. P., de Souza Santos, R., Palmer, B. F., & Clegg, D. J. (2019). Determinants of body fat distribution in humans may provide insight about obesity-related health risks. *Journal of Lipid Research*, *60*(10), 1710–1719. https://doi.org/10.1194/jlr.R086975

Gangwisch, J. E., Malaspina, D., Boden-Albala, B., & Heymsfield, S. B. (2005). Inadequate Sleep as a Risk Factor for Obesity: Analyses of the NHANES I. *Sleep*, *28*(10), 1289–1296. https://doi.org/10.1093/sleep/28.10.1289

Gilbert, C. L. (2010). How to Understand High Food Prices. *Journal of Agricultural Economics*, *61*(2), 398–425. https://doi.org/10.1111/j.1477-9552.2010.00248.x

Goettler, A., Grosse, A., & Sonntag, D. (2017). Productivity loss due to overweight and obesity: a systematic review of indirect costs. *BMJ Open*, *7*(10), e014632. https://doi.org/10.1136/bmjopen-2016-014632

Gruber, J., & Frakes, M. (2006). Does falling smoking lead to rising obesity? *Journal of Health Economics*, *25*(2), 183–197. https://doi.org/10.1016/j.jhealeco.2005.07.005

Gu, J. K., Charles, L. E., Bang, K. M., Ma, C. C., Andrew, M. E., Violanti, J. M., & Burchfiel, C. M. (2014). Prevalence of Obesity by Occupation Among US Workers. *Journal of Occupational & Environmental Medicine*, *56*(5), 516–528. https://doi.org/10.1097/JOM.0000000000000133

Guo, G. (2007). Persistent Inequalities in Funding for Rural Schooling in Contemporary China. *Asian Survey*, *47*(2), 213–230. https://doi.org/10.1525/as.2007.47.2.213

Hammond, R., & Levine. (2010). The economic impact of obesity in the United States. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, 285. https://doi.org/10.2147/DMSOTT.S7384

Harris, J. L., Pomeranz, J. L., Lobstein, T., & Brownell, K. D. (2009). A Crisis in the Marketplace: How Food Marketing Contributes to Childhood Obesity and What Can Be Done. *Annual Review of Public Health*, *30*(1), 211–225. https://doi.org/10.1146/annurev.publhealth.031308.100304

Hecker, J., Freijer, K., Hiligsmann, M., & Evers, S. M. A. A. (2022). Burden of disease study of overweight and obesity; the societal impact in terms of cost-of-illness and health-related quality of life. *BMC Public Health*, *22*(1), 46. https://doi.org/10.1186/s12889-021-12449-2

Higgs, S., & Thomas, J. (2016). Social influences on eating. *Current Opinion in Behavioral Sciences*, *9*, 1–6. https://doi.org/10.1016/j.cobeha.2015.10.005

Horwath, C. C. (1989). Marriage and diet in elderly Australians: results from a large random survey. *Journal of Human Nutrition and Dietetics*, *2*(3), 185–193. https://doi.org/10.1111/j.1365-277X.1989.tb00023.x

Jackson, R. T., Rashed, M., & Saad-eldin, R. (2003). Rural urban differences in weight, body image, and dieting behavior among adolescent Egyptian schoolgirls. *International Journal of Food Sciences and Nutrition*, *54*(1), 1–11. https://doi.org/10.1080/0963748031000062047

Janz, N. K., & Becker, M. H. (1984). The Health Belief Model: A Decade Later. *Health Education Quarterly*, *11*(1), 1–47. https://doi.org/10.1177/109019818401100101

Johnson, J. A., & Johnson, A. M. (2015). Urban-Rural Differences in Childhood and Adolescent Obesity in the United States: A Systematic Review and Meta-Analysis. *Childhood Obesity*, *11*(3), 233–241. https://doi.org/10.1089/chi.2014.0085

Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental Pollution*, *151*(2), 362–367. https://doi.org/10.1016/j.envpol.2007.06.012

Katsaiti, M. S. (2012). Obesity and happiness. *Applied Economics*, *44*(31), 4101–4114. https://doi.org/10.1080/00036846.2011.587779

Kebriaeezadeh, A., Koopaei, N. N., Abdollahiasl, A., Nikfar, S., & Mohamadi, N. (2013). Trend analysis of the pharmaceutical market in Iran; 1997–2010; policy implications for developing countries. *DARU Journal of Pharmaceutical Sciences*, *21*(1), 52. https://doi.org/10.1186/2008-2231-21-52

Keith, S. W., Redden, D. T., Katzmarzyk, P. T., Boggiano, M. M., Hanlon, E. C., Benca, R. M., Ruden, D., Pietrobelli, A., Barger, J. L., Fontaine, K. R., Wang, C., Aronne, L. J., Wright, S. M., Baskin, M., Dhurandhar, N. V, Lijoi, M. C., Grilo, C. M., DeLuca, M., Westfall, A. O., & Allison, D. B. (2006). Putative contributors to the secular increase in obesity: exploring the roads less traveled. *International Journal of Obesity*, *30*(11), 1585–1594. https://doi.org/10.1038/sj.ijo.0803326

Kim, S., & Popkin, B. M. (2006). Commentary: Understanding the epidemiology of overweight and obesity—a real global public health concern. *International Journal of Epidemiology*, *35*(1), 60–67. https://doi.org/10.1093/ije/dyi255

Kim, T. J., & von dem Knesebeck, O. (2018). Income and obesity: what is the direction of the relationship? A systematic review and meta-analysis. *BMJ Open*, *8*(1), e019862. https://doi.org/10.1136/bmjopen-2017-019862

King, L. K., March, L., & Anandacoomarasamy, A. (2013). Obesity & osteoarthritis. *The Indian Journal of Medical Research*, *138*(2), 185–193.
Klaczynski, P. A., Goold, K. W., & Mudry, J. J. (2004). Culture, Obesity Stereotypes, Self-Esteem, and the "Thin Ideal": A Social Identity Perspective. *Journal of Youth and Adolescence*, *33*(4), 307–317. https://doi.org/10.1023/B:JOYO.0000032639.71472.19
Kleinbaum, D. G. (1994). *Logistic Regression*. Springer New York. https://doi.org/10.1007/978-1-4757-4108-7
Kluger, J. (2023, March 21). *Global obesity rates increasing*.
Kroneman, M., Boerma, W., van den Berg, M., Groenewegen, P., de Jong, J., & van Ginneken, E. (2016). Netherlands: Health System Review. *Health Systems in Transition*, *18*(2), 1–240.
Lazuka, R. F., Wick, M. R., Keel, P. K., & Harriger, J. A. (2020). Are We There Yet? Progress in Depicting Diverse Images of Beauty in Instagram's Body Positivity Movement. *Body Image*, *34*, 85–93. https://doi.org/10.1016/j.bodyim.2020.05.001
Lissner, L., Heitmann, B. L., & Bengtsson, C. (2000). Population studies of diet and obesity. *British Journal of Nutrition*, *83*(S1), S21–S24. https://doi.org/10.1017/S000711450000091X
Makker, H. (2010). Obesity and respiratory diseases. *International Journal of General Medicine*, 335. https://doi.org/10.2147/IJGM.S11926
Marcellini, F., Giuli, C., Gagliardi, C., & Papa, R. (2007). Aging in Italy: Urban–rural differences. *Archives of Gerontology and Geriatrics*, *44*(3), 243–260. https://doi.org/10.1016/j.archger.2006.05.004
Martin, S. L., Kirkner, G. J., Mayo, K., Matthews, C. E., Durstine;, J. L., & Hebert, J. R. (2005). Urban, Rural, and Regional Variations in Physical Activity. *The Journal of Rural Health*, *21*(3), 239–244. https://doi.org/10.1111/j.1748-0361.2005.tb00089.x
McLaren, L. (2007). Socioeconomic Status and Obesity. *Epidemiologic Reviews*, *29*(1), 29–48. https://doi.org/10.1093/epirev/mxm001
MENARD, S. (1995). Applied logistic regression analysis. *Sage University Paper Series on Quantitative Application in the Social Sciences*, *7*, 88. https://cir.nii.ac.jp/crid/1571135650228778496.bib?lang=en
Morris, S. (2007). The impact of obesity on employment. *Labour Economics*, *14*(3), 413–433. https://doi.org/10.1016/j.labeco.2006.02.008
Must, A. (1999). The Disease Burden Associated With Overweight and Obesity. *JAMA*, *282*(16), 1523. https://doi.org/10.1001/jama.282.16.1523
Ness-Abramof, R., & Apovian, C. M. (2005). Drug-induced weight gain. *Drugs of Today*, *41*(8), 547. https://doi.org/10.1358/dot.2005.41.8.893630
Obst, E., Schad, D. J., Huys, Q. J., Sebold, M., Nebe, S., Sommer, C., Smolka, M. N., & Zimmermann, U. S. (2018). Drunk decisions: Alcohol shifts choice from habitual towards goal-directed control in adolescent intermediate-risk drinkers. *Journal of Psychopharmacology*, *32*(8), 855–866. https://doi.org/10.1177/0269881118772454
Osler, M., & Hansen, E. T. (1993). Dietary knowledge and behaviour among schoolchildren in Copenhagen, Denmark. *Scandinavian Journal of Social Medicine*, *21*(2), 135–140. https://doi.org/10.1177/140349489302100214
Ponterio, E., & Gnessi, L. (2015). Adenovirus 36 and Obesity: An Overview. *Viruses*, *7*(7), 3719–3740. https://doi.org/10.3390/v7072787
Popkin, B. M. (1999). Urbanization, Lifestyle Changes and the Nutrition Transition. *World Development*, *27*(11), 1905–1916. https://doi.org/10.1016/S0305-750X(99)00094-7
Puhl, R., & Brownell, K. D. (2001). Bias, Discrimination, and Obesity. *Obesity Research*, *9*(12), 788–805. https://doi.org/10.1038/oby.2001.108
Puhl, R. M., & Brownell, K. D. (2003). Psychosocial origins of obesity stigma: toward changing a powerful and pervasive bias. *Obesity Reviews*, *4*(4), 213–227. https://doi.org/10.1046/j.1467-789X.2003.00122.x
Ramsey, C. A., & Hewitt, A. D. (2005). A Methodology for Assessing Sample Representativeness. *Environmental Forensics*, *6*(1), 71–75. https://doi.org/10.1080/15275920590913877
Ritchie, H., & Roser, M. (2017). *Obesity*.
Rolls, B. J. (2003). The Supersizing of America. *Nutrition Today*, *38*(2), 42–53. https://doi.org/10.1097/00017285-200303000-00004
Rosenberg, P., Kano, M., Ludford, I., Prasad, A., & Thomson, H. (2016). *Global Report on Urban Health: Equitable, Healthier Cities for Sustainable Development*.

Rutala, W. A., & Mayhall, C. G. (1992). Medical Waste. *Infection Control and Hospital Epidemiology*, *13*(1), 38–48. https://doi.org/10.1086/646421
Sandercock, G., Angus, C., & Barton, J. (2010). Physical activity levels of children living in different built environments. *Preventive Medicine*, *50*(4), 193–198. https://doi.org/10.1016/j.ypmed.2010.01.005
Sauer, C. M., Reardon, T., Tschirley, D., Liverpool-Tasie, S., Awokuse, T., Alphonce, R., Ndyetabula, D., & Waized, B. (2021). Consumption of processed food & food away from home in big cities, small towns, and rural areas of Tanzania. *Agricultural Economics*, *52*(5), 749–770. https://doi.org/10.1111/agec.12652
Schröder, H., Morales-Molina, J. A., Bermejo, S., Barral, D., Mándoli, E. S., Grau, M., Guxens, M., de Jaime Gil, E., Álvarez, M. D., & Marrugat, J. (2007). Relationship of abdominal obesity with alcohol consumption at population scale. *European Journal of Nutrition*, *46*(7), 369–376. https://doi.org/10.1007/s00394-007-0674-7
Schwartz, M. B., & Brownell, K. D. (2004). Obesity and body image. *Body Image*, *1*(1), 43–56. https://doi.org/10.1016/S1740-1445(03)00007-X
Seaman, S., Galati, J., Jackson, D., & Carlin, J. (2013). What Is Meant by "Missing at Random"? *Statistical Science*, *28*(2). https://doi.org/10.1214/13-STS415
Shimokawa, S. (2013). When does dietary knowledge matter to obesity and overweight prevention? *Food Policy*, *38*, 35–46. https://doi.org/10.1016/j.foodpol.2012.09.001
Sicular, T., Ximing, Y., Gustafsson, B., & Shi, L. (2007). THE URBAN?RURAL INCOME GAP AND INEQUALITY IN CHINA. *Review of Income and Wealth*, *53*(1), 93–126. https://doi.org/10.1111/j.1475-4991.2007.00219.x
Spiegel, K., Tasali, E., Penev, P., & Cauter, E. Van. (2004). Brief Communication: Sleep Curtailment in Healthy Young Men Is Associated with Decreased Leptin Levels, Elevated Ghrelin Levels, and Increased Hunger and Appetite. *Annals of Internal Medicine*, *141*(11), 846. https://doi.org/10.7326/0003-4819-141-11-200412070-00008
Steyn, N. P., Senekal, M., Brtis, S., & Nel, J. (2000). Urban and rural differences in dietary intake, weight status and nutrition knowledge of black female students. *Asia Pacific Journal of Clinical Nutrition*, *9*(1), 53–59. https://doi.org/10.1046/j.1440-6047.2000.00137.x
Sturmer, S., Simon, B., Loewy, M., & Jorger, H. (2003). The Dual-Pathway Model of Social Movement Participation: The Case of the Fat Acceptance Movement. *Social Psychology Quarterly*, *66*(1), 71. https://doi.org/10.2307/3090142
Swinburn, B. A., Kraak, V. I., Allender, S., Atkins, V. J., Baker, P. I., Bogard, J. R., Brinsden, H., Calvillo, A., De Schutter, O., Devarajan, R., Ezzati, M., Friel, S., Goenka, S., Hammond, R. A., Hastings, G., Hawkes, C., Herrero, M., Hovmand, P. S., Howden, M., … Dietz, W. H. (2019). The Global Syndemic of Obesity, Undernutrition, and Climate Change: The Lancet Commission report. *The Lancet*, *393*(10173), 791–846. https://doi.org/10.1016/S0140-6736(18)32822-8
Swinburn, B. A., Sacks, G., Hall, K. D., McPherson, K., Finegood, D. T., Moodie, M. L., & Gortmaker, S. L. (2011). The global obesity pandemic: shaped by global drivers and local environments. *The Lancet*, *378*(9793), 804–814. https://doi.org/10.1016/S0140-6736(11)60813-1
Tam, B. T., Morais, J. A., & Santosa, S. (2020). Obesity and ageing: Two sides of the same coin. *Obesity Reviews*, *21*(4). https://doi.org/10.1111/obr.12991
Taylor, S. I. (1999). Deconstructing Type 2 Diabetes. *Cell*, *97*(1), 9–12. https://doi.org/10.1016/S0092-8674(00)80709-6
Temple, N. J., & Steyn, N. P. (2011). The cost of a healthy diet: A South African perspective. *Nutrition*, *27*(5), 505–508. https://doi.org/10.1016/j.nut.2010.09.005
Valenzuela-Fernández, L., Merigó, J. M., & Nicolas, C. (2018). The most influential countries in market orientation. *International Journal of Engineering Business Management*, *10*, 184797901775148. https://doi.org/10.1177/1847979017751484
Vellinga, R. E., van Bakel, M., Biesbroek, S., Toxopeus, I. B., de Valk, E., Hollander, A., van 't Veer, P., & Temme, E. H. M. (2022). Evaluation of foods, drinks and diets in the Netherlands according to the degree of processing for nutritional quality, environmental impact and food costs. *BMC Public Health*, *22*(1), 877. https://doi.org/10.1186/s12889-022-13282-x
Vinke, P. C., Navis, G., Kromhout, D., & Corpeleijn, E. (2020). Age-and Sex-Specific Analyses of Diet Quality and 4-Year Weight Change in Nonobese Adults Show Stronger Associations in Young Adulthood. *The Journal of Nutrition*, *150*(3), 560–567. https://doi.org/10.1093/jn/nxz262

Visvikis-Siest, S. (2019). Rising rural body-mass index is the main driver of the global obesity epidemic in adults. *Nature*, *569*(7755), 260–264. https://doi.org/10.1038/s41586-019-1171-x

Wang, J., Williams, M., Rush, E., Crook, N., Forouhi, N. G., & Simmons, D. (2010). Mapping the availability and accessibility of healthy food in rural and urban New Zealand – Te Wai o Rona: Diabetes Prevention Strategy. *Public Health Nutrition*, *13*(7), 1049–1055. https://doi.org/10.1017/S1368980009991595

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, *30*(4), 377–399. https://doi.org/10.1002/sim.4067

Wilson, S. E. (2012). Marriage, gender and obesity in later life. *Economics & Human Biology*, *10*(4), 431–453. https://doi.org/10.1016/j.ehb.2012.04.012

Wilson Van Voorhis, C. R., & Morgan, B. L. (2007). Understanding Power and Rules of Thumb for Determining Sample Sizes. *Tutorials in Quantitative Methods for Psychology*, *3*(2), 43–50. https://doi.org/10.20982/tqmp.03.2.p043

Woodward, M., Boltonsmith, C., & Tunstallpedoe, H. (1994). Deficient Health Knowledge, Diet, and Other Life-Styles in Smokers: Is a Multifactorial Approach Required. *Preventive Medicine*, *23*(3), 354–361. https://doi.org/10.1006/pmed.1994.1049

Xu, Y., Zhu, S., Zhang, T., Wang, D., Hu, J., Gao, J., & Zhou, Z. (2020). Explaining Income-Related Inequalities in Dietary Knowledge: Evidence from the China Health and Nutrition Survey. *International Journal of Environmental Research and Public Health*, *17*(2), 532. https://doi.org/10.3390/ijerph17020532

Zhang, Z., & Chen, W. (2019). A Systematic Review of the Relationship Between Physical Activity and Happiness. *Journal of Happiness Studies*, *20*(4), 1305–1322. https://doi.org/10.1007/s10902-018-9976-0

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research*, *37*(2), 197–206. https://doi.org/10.1086/651257

Appendices

Appendix A: Characteristics of the studied sample

Variable name	Number of observations (1)	Mean (2)	Standard deviation (3)	Min (4)	Max (5)
Gender	4879	1.535356	0.498799	1	2
Age	4879	5.167719	1.708875	16	92
Number of household members	4879	2.508096	1.279465	1	8
Diet	4866	1.915947	.2774952	1	2
Grade	2366	3.847422	1.498966	0	9
Overweight	4877	.5080992	.4999857	0	1
Urban	4880	.6270492	.4836387	0	1
Studying	4879	.0778848	.2680177	0	1
Housekeeping	4879	.0776799	.2676946	0	1
Retired	4879	.2363189	.4248638	0	1
Disability	4879	.040992	.198292	0	1
Working	4879	.4894446	.4999398	0	1
Seeking for work	4879	.0422218	.2011153	0	1
Other occupation	4879	.0354581	.1849536	0	1
High education	4879	.092437	.2896715	0	1
Other education	4879	.0237754	.1523642	0	1
Primary education	4879	.0596434	.2368492	0	1
Secondary education	4879	.3500717	.477041	0	1
Vocational education	4879	.4740726	.4993785	0	1
Full Dutch	4677	.8593115	.3477371	0	1
First-generation foreign Western	4677	.0357066	.1855776	0	1
First-generation foreign non-Western	4677	.035279	.1845039	0	1
Second-generation Western	4677	.0481078	.2140168	0	1
Second-generation non-Western	4677	.021595	.1453727	0	1
Frequent drinking	4880	.2231557	.4164046	0	1
Regular drinking	4880	.3618852	.4805951	0	1
Occasional drinking	4880	.2135246	.4098369	0	1
No drinking	4880	.0797131	.2708763	0	1
No sports	4872	.5679392	.4954136	0	1
Occasionaly sports	4872	.2118227	.4086418	0	1
Regular sports	4872	.1750821	.3800763	0	1
Frequent sports	4872	.045156	.2076675	0	1
Log income	4880	7.723526	.530992	2.564949	1.2272533
Happy	4880	.9717213	.2813087	0	2
Never smoke	4880	.425	.4943937	0	1
Now smoke	4880	.1717213	.3771767	0	1
Past smoke	4880	.4020492	.490362	0	1

Appendix B: List of variable characteristics.

Variable name	Variable type	Definition	Units of measurement
Overweight	Binary	Is the individual overweight?	Yes/No: 0 for 'No' and 1 for 'Yes'
Dietary knowledge	Continuous	Individuals answered 12 true/false questions about dietary and health knowledge.	'0' is Bad dietary knowledge and '10' is the perfect dietary knowledge.
Self-reported dietary knowledge	Categorical	Individuals rated their health and dietary knowledge.	Very good: '5' Good: '4' Average: '3' Low: '2' Very low: '1'
Urban	Binary	Urbanity of the area the individual lives.	Urban if number of addresses exceeds 1000 per square kilometer. '1' if urban, '0' if not.
Age	Continuous	The self-reported age of the individual.	Ranging from '18' till '91'
Gender	Binary	Biological gender of the individual.	If male: '0' and female '1'
Number of household members	Continuous	Number of members of the household.	Ranging from '1' till '8'
Civil status	Categorical	What is the civil status of the individual?	Dummy variables for: 'Married', 'Separated', 'Divorced', 'Widow or widower' & 'Never been married'.
Primary occupation	Categorical (With the use of dummy variables)	What is the main occupation of the individual?	Dummy variables for: 'Studying', 'Pension', 'Working', 'Searching for work', 'Housekeeping' & '(partially) disabled'
Logarithmic income	Continuous	What is the gross income level of the individual?	Logarithmic scale, income in EUROS.
Education	Binary	Did the individual complete a higher education? HBO or WO.	If diploma for HBO or WO: '1', if not: '0'

Ethnicity	Categorical (With the use of dummy variables)	What is the origin of the individual?	Dummy variables for: 'Dutch background', 'First generation foreign, western background', 'First generation foreign, non-western background', Second generation foreign, western background' & Second generation foreign, non-western background'
Mental health	Binary	Individual rated if they felt depressed or gloomy on a scale from 1 till 5. If they rated 4 or 5, they are classified as 'Bad mental health'	Bad mental health: '1', good mental health '0'.
Smoking	Categorical (With the use of dummy variables)	Individuals indicated if they have ever smoked and if they smoke now.	Dummy variables for: 'Never smoked', 'Past smoker' & 'Smoker now'
Alcohol	Categorical (With the use of dummy variables)	Individuals indicated how many glasses of alcohol they drank over the past period.	Dummy variables for: 'Drinking almost every day', 'Five or six days a week', Three or four days a week', 'Once or twice a month', 'Once every two months' & 'Not this year'.
Physical activity	Categorical (With the use of dummy variables)	Individuals indicated how many minutes of sports they practiced a week.	Dummy variables for: 'No sport', 'Little sport', 'Average sport' & 'Lot of sports'
Diet	Categorical (With the use of dummy variables)	Does the individual actively follow a diet? If so, what kind of diet?	Dummy variables for: 'Diet with limited salt', 'Diet with limited cholesterol', 'diet with limited calories', 'Another diet', 'Multiple of these diets' & 'No diet'.
Happiness	Binary	The individual rates their happiness on a scale from 1 till 10. If the grade is sufficient 6-10, the individual will be classified as 'happy'.	If individual is happy: '1', if unhappy: '0'.

Appendix C: The results of the regression to test for 'Variables of grade are missing at random.'

Variable names	Coefficient (1)	Std. Err. (2)	p-value (3)
Gender	0.0408782	0.0660431	0.536
Age	-0.0080575	0.0030762	0.009
Number of household members	0.0257532	0.0262626	0.327
Diet	-0.0893692	0.1090386	0.412
Overweight	0.061801	0.0620926	0.320
Urban	0.0015203	0.0624495	0.981
Studying	-0.1090271	0.2304754	0.636
House keeping	-0.0973952	0.1924679	0.613
Retired	-0.2918491	0.1724517	0.091
Disabled	-0.3460027	0.2178934	0.112
Working	-0.1914541	0.1705849	0.262
Searching for work	-0.2682767	0.2180584	0.219
Other occupation	0 (omitted)		
High education	0.1937479	0.1085179	0.074
Other education	-0.3213098	0.2021637	0.112
Primary education	0.27928	0.1427154	0.050
Secondary education	0.0525893	0.0698879	0.452
Vocational education	0 (omitted)		
Dutch	-0.1509629	0.2113964	0.475
First-generation foreign western	-0.0835009	0.2620648	0.750
First-generation foreign non-western	-0.36298	0.2606265	0.164
Second-generation foreign western	-0.099825	0.2496689	0.689
Second-generation foreign non-western	0 (omitted)		
Frequent drinking	0.2313528	0.1110868	0.037
Regular drinking	0.2534077	0.1023952	0.013
Occasional drinking	0.1032112	0.1087421	0.343
No drinking	0.0810547	0.1350091	0.548
No sports	0.0705081	0.1573707	0.654
Occasional sports	0.2682141	0.3555838	0.451
Regular sports	0.1540341	0.2105381	0.464
Frequent sports	0 (omitted)		
Log income	-0.0204655	0.0611635	0.738
Happy	0.1036728	0.1098357	0.345
Satisfied of contacts	-0.172836	0.0879293	0.049
Constant	0.59461	0.7473328	0.426

Number of observations: 4,642

This table represents the regression results used to test whether the variables of grades are missing at random. The p-value represents the significance of each variable in the model.

Appendix DI: The results of the t-tests on continuous and binary variables.

The results of the t-test comparing the means of the original dataset with the inner joint for continuous and binary variables. The P-values represent the results of these t-tests. A P-value below 0.05 indicates a statistically difference in means between the two datasets (indicated with '*')

Dataset	Variable name:	p-value	Number of observations in the original dataset. (1)	Number of observations in the dataset with inner join. (2)	Mean original dataset. (3)	Mean dataset with inner join. (4)
Background variables	Gender	0.0002*	12,518	4,879	1.5091	1.5353
Background variables	Age	0.0000*	12,518	4,879	39.827	51.677
Background variables	Urbanity	0.8101	12,653	4,880	0.6254	0.6270
Background variables	Number of household members	0.0000*	12,518	4,879	3.0607	2.5081
Background variables	Income (logarithm)	0.0002*	4,751	2,476	7.6109	7.6656
Health	BMI	0.8434	6,001	4,877	34.0479	36.0718
Health	Never smoke	0.0644	6,001	4,880	0.4381	0.4323
Health	Now smoke	0.5270	6,001	4,880	0.1751	0.1717
Health	Past smoke	0.0192	6,001	4,880	0.3856	0.4020
Health	Happy	0.1967	6,004	4,880	0.9665	0.9717
Health	Diet	0.2531	5,982	4,866	1.9114	1.9159
Social integration & leisure	Satisfied with social contacts	0.6582	6,571	4,856	0.9218	0.9240

Appendix DII: The percentages for the size of every category for categorical variables.

This appendix shows the proportions of each category for categorical variables in the original dataset and the dataset with an inner join. Each percentage indicates the proportion of the total data represented by a particular category.

Dataset	Variable name & Categories	Percentages dataset inner joint (1)	Percentages Original dataset (2)
Background variables	**Civil status:**		
	Married	56.38	41.96
	Separated	0.70	0.42
	Divorced	9.61	6.68
	Widow or widower	5.98	3.56
	Never been married	27.32	47.38
Background variables	**Primary occupation**		
	Studying	7.79	23.74
	House keeping	7.77	5.78
	Retired	23.63	13.93
	Disability	4.10	2.86
	Working	48.94	43.07
	Seeking for work	4.22	3.60
	Other occupation	3.55	7.02
Background variables	**Education**		
	High education	9.24	8.29
	Vocational education	47.41	37.15
	Secondary education	35.01	28.85
	Other education	2.38	2.42
	Primary education	5.97	23.28
Background variables	**Ethnicity**		
	Dutch background	85.93	83.04
	First-generation foreign, Western background	3.57	3.67
	First generation foreign, non-Western background	3.53	4.92
	Second-generation foreign, Western background	4.81	5.46
	Second-generation foreign, non-Western background	2.16	2.90
Health	**Sports**		
	No activity	56.79	55.26
	Low activity	21.18	21.49
	Moderate activity	17.51	18.66
	High activity	4.52	4.59
Health	**Alcohol**		
	No drinking	20.04	20.74
	Occasional drinking	21.37	21.90
	Regular drinking	36.24	36.79
	Frequent drinking	22.34	20.57

Appendix E: Results of the multiple imputations estimates logistic regression.

Appendix D shows the results of a multiple imputations estimates logistic regression model that analyses the impact of various demographics, lifestyles, and health factors on being overweight. A factor with significant influence on the outcome variable is indicated with an '*'.

Multiple imputation estimates
Logistic regression
DF-adjustment: large sample

Number of observations: 4,642
Average relative variance increase: 0.0314
largest fraction of missing information: 0.5913
F-value: 11.53

Variable name	Coefficient (1)	Standard error (2)	p-value (3)
Urban	-.1039654	.065488	0.112
Grade of dietary knowledge	.014651	.0307731	0.641
Gender	-.4642173	.0704603	0.000*
Age	.0168581	.0037241	0.000*
Number of household members	-.0141482	.0309758	0.648
Diet	-.9703412	.1217501	0.000*
Married	.2651236	.1010164	0.009*
Separated	-.1297149	.3942753	0.742
Divorced	.2056134	.1298696	0.113
Widow or widower	.3733272	.1659885	0.025*
Never been married	0 (omitted)		
Studying	-.4174983	.2561125	0.103
Housekeeping	.2787891	.2016052	0.167
Retired	.1747523	.1801133	0.332
Disability	.5989768	.2305687	0.009*
Working	.4723356	.17937	0.008*
Seeking for work	.4499006	.2291129	0.050*
Other occupation	0 (omitted)		
High education	-.5471877	.1158193	0.000*
Other education	.1184339	.209152	0.571
Primary education	.5161602	.1583614	0.001*
Secondary education	.1080031	.0732398	0.140
Vocational education	0 (omitted)		
Full Dutch	-.3524717	.226348	0.119
First gen foreign western	-.0097793	.2810139	0.972
First gen foreign non western	-.155387	.2768616	0.575
Second gen foreign western	-.3675146	.2655419	0.166
Second gen foreign non western	0 (omitted)		
Frequent drinking	-.474797	.1186638	0.000*
Regular drinking	-.2398463	.1087042	0.027*
Occasional drinking	.0413942	.1152453	0.719
No drinking	.0547029	.1421001	0.700

No sports	.6780308	.1573186	0.000*
Occasional sports	.4689341	.1658262	0.005*
Regular sports	.3891765	.1685471	0.021*
Frequent sports	0 (omitted)		
Log income	.0910411	.0666597	0.172
Happy	-.1657045	.1152541	0.151
Never smoke	-.4231292	.0732506	0.000*
Now smoke	-.5540734	.0926159	0.000*
Past smoke	0 (omitted)		
Satisfied of contacts	.0598316	.0922155	0.516
Constant	.8857352	.7248731	0.222

(* if $p<0.05$)

Appendix FI: Mediation; Relationship between overweight and urbanicity

Dependent variable in this regression: overweight

Multiple imputation estimates | Number of observations: 4,665
Logistic regression | Average relative variance increase: 0.0000
DF-adjustment: large sample | largest fraction of missing information: 0.0000
F-value: 13.57

Variable name	Coefficient (1)	Std. err. (2)	p-value (3)
Urban	-.1191627	.0652039	0.068
Gender	-.4601284	.0689996	0.000*
Age	.0219248	.0032795	0.000*
Number of household members	.0206751	.0271313	0.446
Diet	-.9520288	.1253763	0.000*
Studying	-.4524848	.255856	0.077
Housekeeping	.2913621	.2038609	0.153
Retired	.164763	.1812933	0.363
Disability	.5779189	.2293664	0.012*
Working	.4698061	.1801607	0.009*
Seeking for work	.4254532	.2261695	0.060
Other occupation	0	(omitted)	
High education	-.5368528	.1133755	0.000*
Other education	.1180867	.2064463	0.567
Primary education	.5176204	.1565479	0.001*
Secondary education	.1077577	.0731424	0.141
Vocational education	0	(omitted)	
Full Dutch	-.3371682	.2354692	0.152
First-generation foreign Western	.0432599	.2875363	0.880
First-generation foreign non-Western	-.1377703	.283023	0.626
Second-generation foreign Western	-.3373429	.2721153	0.215
Second-generation foreign non-Western	0	(omitted)	
Frequent drinking	-.4837026	.1181697	0.000*
Regular drinking	-.2478145	.1079675	0.022*
Occasional drinking	.0355205	.1155619	0.759
No drinking	.0508386	.1438115	0.724
No sports	.6964036	.1576527	0.000*
Occasional sports	.4811691	.1658905	0.004*
Regular sports	.4032554	.1682116	0.017*
Frequent sports	0	(omitted)	
Log income	.0927603	.0668811	0.165
Happy	-.1507369	.1137007	0.185
Never smoke	-.4153124	.0725979	0.000*
Now smoke	-.557942	.0928214	0.000*
Past smoke	0	(omitted)	
Constant	.7599743	.7253775	0.295

Significant (*) if P<0.05

Appendix FII: Mediation; Relationship between dietary knowledge and urbanicity.

Dependent variable in this regression: Dietary knowledge

Multiple imputation estimates | Number of observations: 4,665
Linear regression | Average relative variance increase: 0.5667
DF-adjustment: small sample | largest fraction of missing information: 0.6725
F-value: 1.98

Variable name	Coefficient (1)	Std. err. (2)	p-value (3)
Urban	-.006701	.0536217	0.901
Gender	.1160942	.0720033	0.128
Age	.0102191	.0025904	0.000*
Number of household members	.0301389	.0303373	0.342
Diet	.0237766	.0934221	0.800
Studying	0	(omitted)	
Housekeeping	-.3480424	.1930091	0.088
Retired	-.3626989	.1844179	0.060
Disability	-.2462598	.1853216	0.192
Working	-.3345464	.1284785	0.012
Seeking for work	-.4403694	.1706112	0.011*
Other occupation	-.3490817	.247889	0.180
High education	.2161979	.2113783	0.313
Other education	0	(omitted)	
Primary education	-.2595126	.2319538	0.272
Secondary education	.0845178	.1998776	0.675
Vocational education	.1622058	.1944217	0.409
Full Dutch	-.0954561	.1921206	0.622
First-generation foreign Western	-.2467712	.2093346	0.239
First-generation foreign non-Western	-.2248515	.2415426	0.357
Second-generation foreign Western	-.0823521	.1885855	0.662
Second-generation foreign non-Western	0	(omitted)	
Frequent drink	.2950225	.11869	0.022*
Regular drink	.1553311	.129339	0.256
Occasional drink	.1462949	.1108224	0.200
No drink	.0634726	.1175525	0.591
No sports	-.0513907	.0709579	0.472
Occasional sports	.0199936	.0783124	0.799
Regular sports	0	(omitted)	
Frequent sports	-.0152645	.1269355	0.904
Log income	.1122071	.0621984	0.074
Happy	-.0150029	.0986012	0.880
Never smoke	-.0638163	.0694153	0.369
Now smoke	-.062055	.0892212	0.495
Past smoke	0	(omitted)	
Constant	2.380.435	.5968913	0.000*

Significant (*) if P<0.05