The impact of poverty on obesity in Belgium

Auteur : Challal, Hajar
Promoteur(s) : Lefevre, Mélanie
Faculté : HEC-Ecole de gestion de l'Université de Liège
Diplôme : Master en sciences économiques, orientation générale, à finalité spécialisée en economic, analysis and policy
Année académique : 2024-2025
URI/URL : http://hdl.handle.net/2268.2/24148

The impact of poverty on obesity in Belgium

Jury:
Supervisor:
Mélanie LEFEVRE
Reader(s):
Jérôme SCHOENMAECKERS
Xavier FLAWINNE

Master thesis presented by
Hajar CHALLAL
To obtain the degree of
MASTER IN ECONOMICS
with a specialization in
Economic, Analysis and Policy

Academic year 2024/2025

Ecole de Gestion de l'Université de Liège

Acknowledgments

I would like to express my gratitude to my thesis supervisor, Professor Mélanie Lefevre, for her guidance and valuable feedback throughout the making of this thesis.

I am also thankful to Professor Jérôme Schoenmaeckers for his help with the methodological part of this work. I am grateful to him, as well as to Mr. Xavier Flawinne, for taking the time to read and evaluate my thesis.

In addition, I would like to thank Professor Lionel Artige for helping me define and refine the subject of this research last year.

Finally, I am deeply grateful to my family for their constant support and to my friends and classmates, especially Chirine, for making these two years of master's studies more enjoyable and for helping me get through this challenge.

Table of Contents

Acknowledgments i

Introduction 1

Literature Review 2

Data & descriptive statistics 6

1.1. Presentation of the database 6

1.2. Theoretical model 6

1.3. Descriptive statistics 8

Methodology 14

1.1. Naive estimation 14

1.2. A solution to endogeneity 14

1.2.1. Linear probability model 15

1.2.2. The Control function approach 16

1.2.3. Limitations 17

Empirical results 18

1.1. Preliminary tests 18

1.2. Discussion of results 19

Complementary analysis & Robustness checks 22

1.1. Obesity among the youth 22

1.2. Health literacy 24

Conclusion and discussion 26

References 27

Appendix 31

Appendix 1: table of variables 31

EXECUTIVE SUMMARY 32

Introduction

Obesity is rapidly increasing in Europe and across the world. According to the World Health Organization, "*Worldwide adult obesity has more than doubled since 1990, and adolescent obesity has quadrupled. In 2022, 2.5 billion adults were overweight, of whom 890 million were living with obesity*"***(WHO, 2024).*** Obesity has serious health consequences, including an increased risk of diabetes, cardiovascular diseases, and other chronic conditions ***(Kelishadi, 2014)***. It is one of the leading causes of premature death worldwide ***(Ritchie & Roser, 2024; Members' Research Service, 2025)***. In 2019 alone, approximately 5 million premature deaths were attributed to obesity, representing around 10% of deaths in high-income countries and about 15% in middle-income countries ***(Ritchie & Roser, 2024).***

In Europe, the situation is also concerning. Nearly 51% of the EU's population aged 16 and over are overweight, and 17% are obese. Rates are rising not only among adults but also among children. These alarming figures highlight the urgency of addressing the obesity epidemic.

To counter this trend, the WHO set a goal in 2018 to halt the rise in obesity by 2025 ***(Mahase, 2022; WHO, 2025).*** Unfortunately, no EU country is currently on track to meet this goal ***(Members' Research Service, 2025)***.

Obesity patterns also vary by income level. In high-income countries, research has shown that obesity is concentrated among economically disadvantaged households, whereas in low-income countries, it tends to be more common among higher-income groups ***(Dinsa et al., 2012)***. In the United States, the link between low income and higher obesity risk is well documented, with numerous studies highlighting the so-called *poverty–obesity paradox*, where rising poverty and food insecurity correlate with higher obesity rates ***(Levine, 2011).***

Although several studies have examined the relationship between poverty and obesity in European countries (***Żukiewicz-Sobczak et al.,2014; Salmasi & Celidoni, 2017)***, specific research on Belgium remains limited. Therefore, this thesis aims to address this gap by answering the following question:

Does poverty cause obesity in Belgium?

The objective is to assess whether poverty has any impact on obesity among the Belgian population and to explore the mechanisms through which it may influence it. The analysis is based on data from *the Belgian Health Interview Survey (HIS) from Sciensano.* The thesis is structured as follows: first, a review of the existing literature on the topic is presented, and then the empirical methodology and model specification follow. Third, the presentation and discussion of results are presented, and finally, additional analyses to reinforce the findings are presented, followed by the conclusion.

Literature Review

Obesity is a chronic disease characterized by an excessive accumulation of fat in the body. The body mass index (BMI), calculated as weight (kg) divided by height squared (m^2), is widely used to assess someone's weight. ***(World Health Organization: WHO, 2024).*** Based on BMI, adult individuals are typically classified into four categories: underweight (BMI < 18.5), normal weight (BMI 18.5–24.9), overweight (BMI 25–29.9), and obese (BMI ≥ 30) ***(Dwyer et al., n.d.).***

Belgium is not spared from the global rise in obesity. In the Food Consumption Survey 2022–2023 conducted by Sciensano, a total of 3,777 individuals randomly selected from the Belgian population were interviewed about their food intake, health status, and behaviors. The study found that 49% of the population aged 3 years and over was overweight, including 18% who were classified as obese ***(Sciensano, 2023).*** Obesity was measured using the Body Mass Index (BMI), but the results were confirmed using other indicators such as waist circumference and the waist-to-height ratio.

They also highlighted that the prevalence of overweight and obesity increases with age. Adults, particularly those who are middle-aged or older, are more affected than younger age groups. Furthermore, adult men have a higher probability of being overweight or obese compared to women. However, among children and adolescents, no significant gender differences were observed ***(Sciensano, 2023).***

Obesity can have serious consequences on health, such as type 2 diabetes, cardiovascular diseases, and hypertension ***(Kelishadi, 2014)***. Beyond its health consequences, obesity also represents a significant financial burden for the healthcare system and the economy. Based on data from the Belgian National Health Interview (BHIS) and individual health insurance data, ***Gorasso et al. (2022)*** estimated that every year, at least €4.5 billion were spent to cover the direct (such as medical care and treatment) and indirect costs (cost of absenteeism, loss of productivity) related to overweight and obesity.

Many studies have been conducted to better understand the causes of obesity. While unhealthy eating behaviors and lack of physical activity are well-known contributors ***(NHLBI, 2022)***, research has also revealed the important role of poverty and socioeconomic conditions.

Studies have been done in the United States to analyze the relationship between poverty and obesity, highlighting a "*poverty-obesity paradox*" where rising poverty and food insecurity correlate with higher obesity rates ***(Dhurandhar, 2016)***. This paradox is not limited to the United States. For example, in Poland, ***Żukiewicz-Sobczak et al. (2014)*** found that lower-income families had higher rates of overweight and obesity.

Poverty is broadly defined by the World Bank as "*pronounced deprivation in well-being*," typically understood in terms of a lack of income and limited access to essential services such as food, education, healthcare, and housing ***(World Bank, 2023).*** In low-income countries, poverty is often defined in absolute terms, referring to the inability to meet basic needs. In contrast, in higher-income countries, it is seen as a relative concept: individuals are considered poor if they lack the resources needed to participate adequately in society as it evolves ***(World Bank, 2023).***

In the European Union, poverty is measured using the AROPE indicator (At Risk of Poverty or Social Exclusion), which takes into account three dimensions ***(Eurostat):***

- *Monetary poverty*: Living in a household with disposable income below 60% of the national median.

- *Low work intensity*: Living in a household where working-age members worked less than 20% of their total work potential over the past year.
- *Severe material and social deprivation*: Inability to afford essential goods or services, such as paying bills on time, adequately heating the home, or participating in basic social activities.

A person is considered at risk of poverty or social exclusion if they meet at least one of these criteria.

In Belgium, *Statbel*, the national statistics institute, reported that in 2024, 18.2% of the population was at risk of poverty or social exclusion, slightly lower than in 2023 (18.6%). The highest rate was observed in the Brussels-Capital Region (37%), followed by Wallonia (24%) and Flanders (12.2%). Focusing only on income poverty (AROP), 11.5% of Belgians had an income below the poverty threshold, which in 2024 corresponded to €18,235 per year for a single adult ***(Statbel, 2025)***.

To address these issues, the European Commission has set a target to reduce the number of people at risk of poverty or social exclusion by 15 million by 2030, including 5 million children. However, projections suggest that Belgium is unlikely to meet this target ***(Federal Planning Bureau, 2024).***

Several reasons explain why poverty can impact weight. Firstly, healthy foods tend to be more expensive, making them less accessible to low-income families. Additionally, some low-income neighborhoods, known as "food deserts", lack affordable options for nutritious food, making healthy eating more difficult. Meanwhile, high-fat, calorie-dense foods are often cheaper and more widely available, making them a common choice for those with limited resources. There is also an increasing trend among the youth to consume sweet snacks and sodas, as they are cheaper than healthy snacks like fruits. This leads to poor dietary habits and eventually adolescent obesity ***(Finkelstein & Strombotne, 2010; Żukiewicz-Sobczak et al., 2014).***

The living environment also plays a significant role in influencing weight. A study conducted in Turkey ***(Pirgon & Aslan, 2015)*** highlighted a strong correlation between children's living environments and their likelihood of being overweight or obese. Childhood obesity is found to be more prevalent in economically disadvantaged urban areas, where low-income households are more likely to live in neighborhoods with limited access to green spaces, recreational facilities, and safe walking paths.

In addition, rapid urbanization has contributed to this issue. Many families, especially in under-resourced areas, tend to keep their children indoors due to safety concerns or a lack of suitable outdoor spaces. This often leads to increased screen time and reduced physical activity, further contributing to childhood obesity ***(Pirgon & Aslan, 2015).***

Further research reinforces the importance of neighborhood design in shaping health outcomes. Specifically, individuals living in more walkable environments tend to have lower body weight. For example, ***Kowaleski-Jones et al. (2017),*** using propensity score matching on cross-sectional data from Salt Lake County, found that mothers in more walkable neighborhoods had lower BMI. Similarly, ***Wang et al. (2022),*** using data from the U.S. National Health Interview Survey, confirmed a negative association between walkability and BMI.

Beyond physical activity, the food environment also contributes to obesity disparities. In Flanders, ***Smets and Vandevijvere (2022)*** found that the number of fast-food outlets and convenience stores near schools increased between 2008 and 2020. Using spatial data, they assessed the food environment around schools and linked it to school-level obesity rates. Their results showed a positive correlation between the density of fast-food outlets and the weight of children under 12, with stronger effects among children from economically disadvantaged families.

Another important channel through which poverty may influence obesity is education. Children growing up in low-income households often face barriers to educational attainment, including limited access to quality schools and fewer academic resources. According to ***Ferguson et al. (2007)***, such disadvantages in early life can have lasting effects on cognitive development and academic achievement.

Lower levels of education are associated with reduced health literacy, less healthy diets, and a higher risk of obesity. A study using sibling data from the Wisconsin Longitudinal Study ***(Kim, 2016)*** found that higher levels of education were associated with lower BMI, as educated individuals are more likely to adopt healthier behaviors. An analysis of four OECD countries (Australia, Canada, England, and Korea) suggests a strong negative correlation between education and obesity, particularly among women. However, the causal relationship remains uncertain. For instance, using changes in compulsory education laws in Denmark, ***Arendt (2005)*** found mixed results regarding the impact of education on BMI. In contrast, ***Spasojevic (2003, as cited in Devaux et al., 2011)*** demonstrated that an additional year of education in Sweden has a causal effect on maintaining a healthy BMI.

In Belgium, a study using data of a representative random sample of individuals aged 20-69 years old in Liège revealed that abdominal obesity is associated with lower educational level and economic disadvantages ***(Streel et al., 2015).*** Furthermore, another study conducted in Belgium, using a representative sample of 10,084 individuals aged 2 to 17 years, concluded that the increase in childhood obesity was more pronounced among children with less educated parents ***(Drieskens et al., 2024)***.

Education is often linked to higher health knowledge, also referred to as '*health literacy'*. Health literacy is defined as "*the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions"* ***(Ratzan & Parker, as cited in Van Der Heide et al., 2013)****.* Using regression analysis based on the 2013 Danish health and morbidity survey**, *Friis et al. (2016)*** identified health literacy as a mediator that explains the relationship between education and health outcomes, including obesity. Similarly, a study conducted in the United States ***(Chari et al., 2013)*** revealed that adolescent obesity was negatively correlated with health literacy. The study also showed that obesity in school-aged children was associated with both the health literacy and obesity levels of their parents, suggesting that individuals with greater health knowledge are more likely to adopt healthier lifestyles.

In the context of Belgium, based on the Health Interview Survey, ***Berete et al. (2024)*** highlighted that health literacy partially mediates the relationship between educational attainment and income, and health behaviors. Health literacy was found to be positively associated with physical activity and diet. However, the mediating role of health literacy remains limited.

While our paper will analyze the causal impact of poverty on obesity, it is important to discuss the reverse relationship between obesity and poverty. ***Komlos and Rashad (2014)*** wrote that obesity can limit earnings by three ways: when the employer has a distaste for obese workers ("*cosmetic discrimination*"), when obesity is perceived as a proxy for low productivity by employers so they would be willing to employ them at a lower wage ("*statistical discrimination*") and finally when employers provide fewer job-trainings to obese workers, which results in a lower wage over time ("*job promotion discrimination*"). They found that obesity had a negative impact on wages for women in the US and Southern Europe, but not in the North of Europe (like Belgium). Moreover, they did not find any significant effect of obesity on wages for men. ***Brunello and D'Hombres (2006)*** found similar conclusions: the negative impact of the weight on the wage is larger in the South of Europe.

Another important point to mention is that all the studies mentioned above used the Body Mass Index (BMI) as a measure of obesity. However, BMI has some limitations. It does not differentiate between fat, muscle, and bone mass, which can result in misclassifying muscular individuals as 'obese' when they are not ***(Burkhauser & Cawley, 2007)***. It is also argued that BMI is a poor measure of obesity, as it is weakly correlated with actual body fat. Alternative measures, such as waist circumference or body fat percentage, provide a more accurate assessment of fatness. Given the limitations of our dataset, which includes only height and weight, BMI represents the most viable measure of obesity for this study. Future research could benefit from incorporating more precise measures of body weight.

To conclude, while studies like ***Salmasi and Celidoni (2017)*** have examined the poverty obesity paradox at a European level, there remains a lack of research focused specifically on Belgium. Additionally, while previous analyses have employed instrumental variables to address the reverse causality, identifying a valid instrument can be challenging given our database. This study aims to address these gaps by using micro-level data from the Belgian Health Interview Survey and applying alternative methods, such as the *Lewbel Model,* to establish the causal relationship between poverty and obesity in Belgium.

Data & descriptive statistics

1.1. Presentation of the database

This study uses data from *the Health Interview Survey (HIS)*, conducted by Sciensano, the Belgian Institute for Health. The HIS is a nationally representative cross-sectional survey carried out in six waves between 1997 and 2018, collecting comprehensive information on Belgian residents' health behaviors, socio-economic conditions, living environments, and financial situations.

Each wave consists of a net sample of approximately 10,000 individuals, interviewed at both the individual and household levels. At the individual level, two types of questionnaires are used: a face-to-face questionnaire, conducted by an interviewer, and a self-administered questionnaire, completed independently by respondents aged 15 and above, as this type of questionnaire includes more sensitive topics such as mental health and substance use.

The HIS was repeated several times throughout time (1997 to 2018), but this analysis only focuses on the 2013 and 2018 waves, as some variables (such as nutritional habits and access to parks) are only available in those years.

Our main sample consists of adults aged 18 and over, since some indicators (such as educational attainment) do not apply to minors. However, to complement the main analysis, we also include a robustness check based on individuals under 18, to explore whether childhood poverty is associated with obesity.

After merging the individual and household datasets, the final sample includes 12,552 individuals aged 2 to 102. The adult sample forms the main analysis, while the child and adolescent data serve as a secondary analysis.

It is important to note that Sciensano provides a weighting procedure to make the survey data more representative of the Belgian population. These weights correct for unequal selection probabilities, non-response, and demographic imbalances. However, for the sake of simplicity, those weights are not applied in this study. As a result, the findings presented may not fully reflect the Belgian population and should be interpreted with that limitation in mind.

1.2. Theoretical model

To evaluate the impact of poverty on obesity, the following model will be estimated :

$$\begin{aligned} Obese_i = \beta_0 + \beta_1 Income\ quintiles_i + \beta_2 Education_i + \beta_3 Age_i + \beta_4 Gender_i \\ + \beta_5 Region\ of\ residence_i + \beta_6 Urbanization_i + \beta_7 Access\ to\ parks_i \\ + \beta_8 Physical\ activity_i + \beta_9 Nutrition\ habits_i + \varepsilon_i \end{aligned}$$

Dependent Variable

Obesity is measured using the *Body Mass Index* (BMI). The dependent variable is a binary indicator equal to 1 if the individual's BMI is greater than or equal to 30 (classified as obese according to WHO standards), and 0 otherwise. In the HIS dataset, a pre-constructed variable already identifies obesity among adults based on this threshold.

For children and adolescents, obesity is not defined using a fixed BMI cut-off. Instead, classification relies on age- and sex-specific thresholds developed by ***Cole and Lobstein (2012).*** Because children's BMI varies naturally with age and differs between boys and girls, adult BMI classifications cannot be directly applied. Using adult cut-offs would risk misclassifying many children's weight status. To address this, ***Cole et al. (2000)*** aggregated data from six countries (Brazil, Great Britain, Hong Kong, the Netherlands, Singapore, and the United States) to establish international age- and sex-specific BMI centile curves. These curves were designed to adjust BMI thresholds across childhood and adolescence while aligning with the adult definitions of overweight (BMI ≥ 25) and obesity (BMI ≥ 30) at age 18. This provides standardized, internationally comparable, and less arbitrary thresholds for defining overweight and obesity in children.

Independent Variables

Poverty is measured using household income after social transfers, categorized into quintiles. These range from 1 (lowest income group) to 5 (highest), reflecting each household's position within the national income distribution in Belgium. In the analysis, the richest quintile (quintile 5) is used as the reference category, allowing us to compare the probability of being obese for individuals in the lowest income group relative to those in the highest.

Education is measured based on the highest diploma obtained by the individual. It is a categorical variable ranging from 1 (no diploma or only primary school) to 4 (higher education, like a bachelor's or master's degree).

Age and gender are included as demographic control variables. Based on the literature, the probability of obesity is expected to increase with age and with being male.

The *region of residence* is included as a categorical variable distinguishing between Flanders, Brussels, and Wallonia.

The urbanization level corresponds to the type of area where the respondent lives, ranging from 1 (rural) to 4 (big cities and dense agglomerations). As discussed in the literature, the living environment may influence dietary and physical activity behaviors, and thus obesity.

Access to parks is measured using a dummy variable indicating whether the respondent has access to green or recreational public spaces (1 = yes, 0 = no). This reflects environmental constraints that may limit physical activity.

Physical activity is captured using the indicator for leisure-time physical activity, based on WHO guidelines. The indicator is available at the individual level for respondents aged 15 and above. This variable ranges from 1 (sedentary) to 3 (high-intensity physical activity or more than 4 hours of training).

Nutritional habit is measured by a variable indicating the frequency of sweet and salty snack consumption, as the literature has shown a strong link between high snack consumption and increased obesity risk. The variable ranges from 1 (never) to 5 (once or more a day).

All variables included in the model are summarized in ***Appendix 1***. The table provides the variable names, value ranges or categories, and the survey years for which the data are available.

1.3. Descriptive statistics

This section presents some descriptive statistics to provide an initial overview of the key variables in the analysis, with a particular focus on the relationship between obesity and income. As a reminder, the sample used in this study might not be representative of the Belgian population due to the non-use of sampling weights. Therefore, the results apply only to our sample and should be interpreted with caution.

General overview of the database

The database consists of 12,552 individuals, including 3,156 individuals under the age of 18, and covers two survey years: 2013 and 2018.

Table 1 presents a general overview of the sample's characteristics, including age, gender, region of residence, and degree of urbanization and income quintiles. The sample is relatively balanced across key variables. The distribution between male and female respondents is nearly equal, with 48.9% men and 51.1% women. The two survey waves are relatively balanced, though 2018 represents a slightly larger share of the sample (54%) compared to 2013 (46%).

In terms of age, the largest age group consists of middle-aged adults (40–64), who constitute 36% of the sample. Regionally, most respondents live in Flanders (39.5%) and Wallonia (36%), while Brussels accounts for 24.6% of the sample. As for the degree of urbanization, 46.2% of individuals live in large cities or dense agglomerations, while only 15.4% live in rural areas.

Regarding income quintiles, 24% of the sample belongs to households classified in the fifth income quintile (the highest), while 18% falls into the first quintile (the lowest). Overall, the sample shows a relatively balanced distribution across the five income groups

Table 1 : General overview

	Observations	Proportion (%)
2013	5,779	46.04
2018	6,773	53.96
Total	***12,552***	***100***
Male	6,138	48.90
Female	6,414	51.10
Total	***12,552***	***100***
Flemish Region	4,954	39.47
Brussels' Region	3,084	24.57
Walloon Region	4,514	35.96
Total	***12,552***	***100***
Big cities and dense agglomerations	5,795	46.17
Suburban / banlieus	1,797	14.32
Urbanized municipalities	3,028	24.12
Rural	1,932	15.39
Total	***12,552***	***100***
2-17 years old	3,156	25.14
18-39 years old	2,739	21.82
40-64 years old	4,521	36.02
65+ years old	2,136	17.02
Total	***12,552***	***100***
1st income quintile	2,247	17.90
2nd income quintile	1,971	15.70
3rd income quintile	2,497	19.89
4th income quintile	2,816	22.43
5th income quintile	3,021	24.07
Total	***12,552***	***100.00***

Source: Own computations based on the data from the HIS database

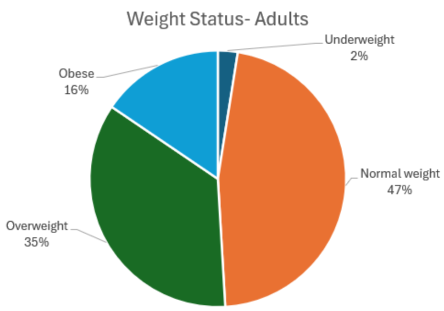
Distribution of the weight classes

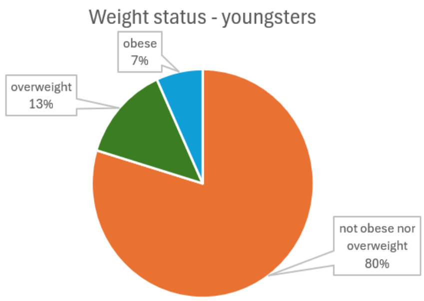
Figures 1 and 2 display the distribution of weight status among adults and youngsters, respectively. As said earlier, the classification of weight categories differs between the two groups: adult BMI categories are based on standard WHO cutoffs, while for children and adolescents, the classification is age- and sex-specific.

Among adults, nearly half of the respondents (47%) have a normal weight. However, a considerable proportion is overweight (35%), and 16% are classified as obese. Underweight is less common, affecting only 2% of the adult population.

Among youngsters, 80% are classified as neither overweight nor obese, while 13% are overweight and 7% are obese. It is important to note that within the group labeled as “not overweight nor obese,” the database does not allow us to distinguish between children with normal weight and those who are underweight. For children, only the variables “overweight” and “obese” are available, while BMI values and the corresponding categories of underweight and normal weight are not provided.

Figure 1: Weight status – adults *Figure 2: Weight status - youngsters*



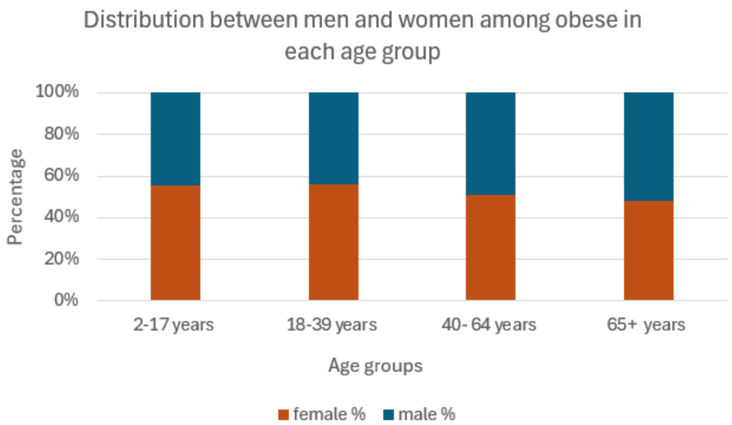


Source: Own computations based on the data from the HIS database

Distribution of obesity by gender and age

Figure 3 presents the distribution of obese individuals by gender across age groups. This analysis is restricted to individuals classified as obese. Among children and adolescents (2–17 years), as well as young adults (18–39 years), obesity is more prevalent among females than males. However, this trend reverses in the older age groups: from age 65 onward, the proportion of obese individuals is slightly higher among males. In the 40–64 age group, the gender gap narrows, with nearly equal shares of obese men and women.

Figure 3: Distribution between men and women among obese by age group

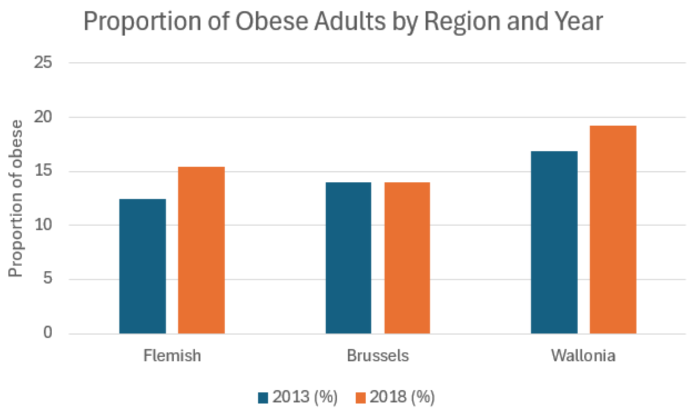


Source: Own computations based on the data from the HIS database

Prevalence of obesity by region and year

Figure 4 presents the proportion of obese adults by region for the years 2013 and 2018. We observe an increase in obesity prevalence in Flanders and Wallonia over this period, while the proportion remains relatively stable in Brussels. In both years, Wallonia shows the highest obesity rates, with 18% of its adult population classified as obese in 2018.

Figure 4: Proportion of obese adults by region and year



Source: Own computations based on the data from the HIS database

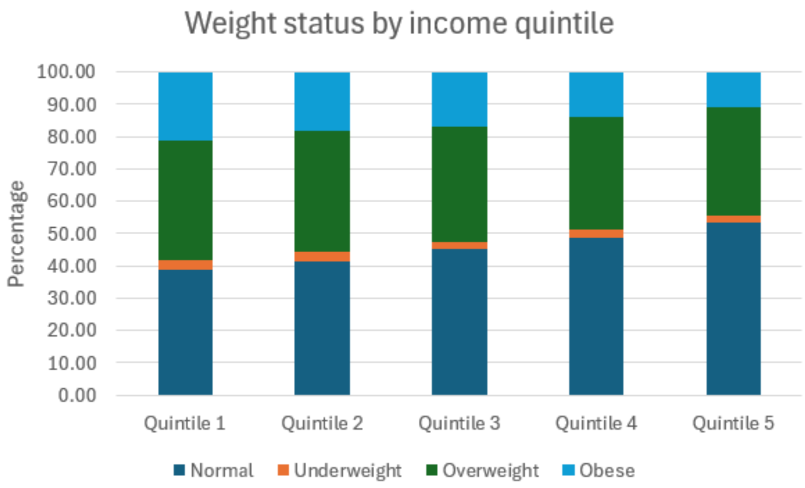
Relationship between income quintile and weight

The central focus of this study is the relationship between poverty and obesity. To explore this, both the distribution of weight classes and the average BMI across income quintiles are examined. Weight status is defined based on BMI, using the WHO classification. *Figure 5* displays the distribution of weight categories among adults aged 18 and over, where the first quintile represents the lowest-income group and the fifth the highest.

As income increases, a clear pattern emerges: the proportion of individuals with normal weight rises from 39% in the lowest quintile to 50% in the highest, while obesity prevalence drops from 21% to 11%. The share of overweight individuals also decreases slightly with income.

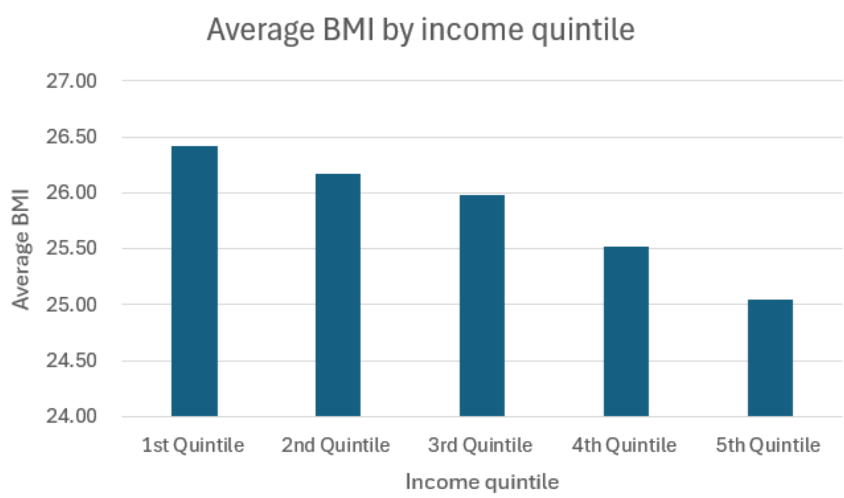
In addition, *Figure 6* presents the average BMI for adults by income quintile, showing a similar downward trend: individuals in the lowest quintile have an average BMI of 26.4, compared to 25.0 in the highest quintile. Together, these findings suggest a negative relationship between income and body weight among Belgian adults.

Figure 5: Weight status by income quintile - adults



Source: Own computations based on the data from the HIS database

Figure 6 : average BMI by income quintile - adults



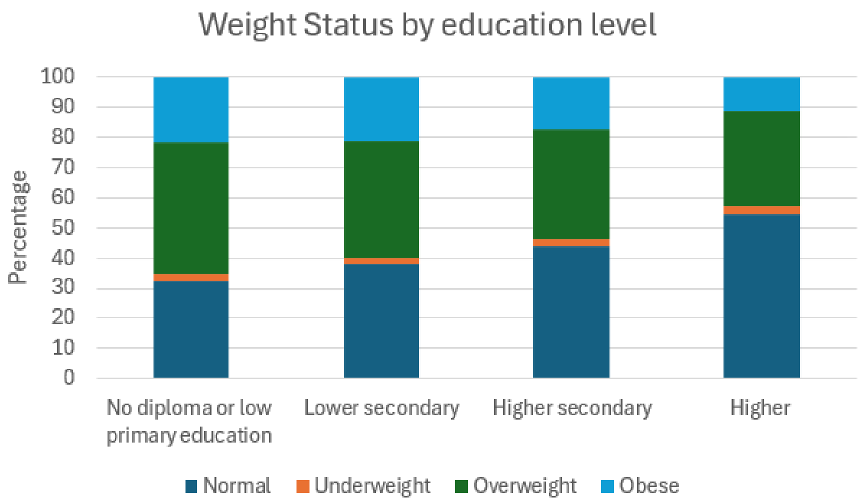
Source: Own computations based on the data from the HIS database

Education and weight status

The relationship between weight status and education level is also examined. *Figure 7* shows the distribution of weight categories by educational attainment for adults. The trend is similar to what was observed with income: as education level increases, the proportion of individuals with normal weight rises, while the share of obese individuals decreases, from 21% among those with no or low formal education to 11% among those with higher education.

These results appear to align with the idea that higher socioeconomic status is associated with healthier weight outcomes.

Figure 7: Weight status by education level - adults



Source: Own computations based on the data from the HIS database

This section provided a first look into the potential relationship between income (and by extend poverty) and obesity. As suggested by the descriptive graphs, there appears to be a correlation between lower income levels and higher obesity rates. However, descriptive statistics alone cannot establish causality. To assess the impact of poverty on obesity, more advanced statistical methods are required.

The next section defines the methodological approach used in this analysis.

Methodology

1.1. Naive estimation

As a first step to understand the relationship between poverty and obesity, a logit model is used. Given the binary nature of the dependent variable (obesity), this approach allows us to estimate the probability of being obese based on different income quintiles.

In a first specification, a simple model is used, including only the income quintiles as explanatory variables. In a second step, control variables were added, such as age, gender, education, and others mentioned in the theoretical model.

However, it is important to acknowledge the limitations of this naive estimation. In particular, the model may suffer from endogeneity**.** As discussed in the literature review, reverse causality is possible: obesity can influence employment status and, by extension, income. If income is endogenous, the coefficients from a logit model are biased and inconsistent, and the marginal effects cannot be interpreted causally. Therefore, another method is used to address this issue.

1.2. A solution to endogeneity

As discussed earlier, our model is susceptible to endogeneity, primarily due to the potential for reverse causality between income and obesity. This means that the coefficients from a naïve logit estimation may be biased. A common way to address endogeneity is to use instrumental variables (IV) in a two-stage regression. However, no variable available in the HIS dataset satisfies both the relevance and validity conditions required for an instrument.

To overcome this limitation, the ***Lewbel (2012)*** method is used. This approach enables us to generate instruments from the data itself by exploiting heteroskedasticity in the first-stage error term. The key idea is to use interactions between mean-centered exogenous variables and the heteroskedastic component of the error to construct instruments that satisfy IV assumptions. This method is useful when external instruments are available.

The method is based on three main assumptions ***(Baum & Lewbel, 2019)*** :

- ***Endogeneity through common unobserved factors:*** The unobserved factors in the first-stage equation (income) and the outcome equation (obesity) are correlated, meaning some hidden variables affect both, creating endogeneity.
- ***Exogeneity of the constructed instruments:*** The Lewbel instruments must not be correlated with the unobserved error term in the second stage equation. The instruments can affect obesity only indirectly, through income. This is a difficult assumption to verify in practice.
- ***Presence of heteroskedasticity in the first-stage residuals:*** The error term from the first-stage regression must display heteroskedasticity. This is what allows Lewbel's method to generate instruments from the data. If the error variance is constant, the method cannot be applied. This is why we need to test for heteroskedasticity (e.g., using Breusch–Pagan or White tests) before applying the method.

The Lewbel method was originally designed for linear models estimated by Ordinary Least Squares (OLS). In our case, the outcome variable (obesity) is binary, making the model nonlinear and preventing direct application of the standard two-stage least squares (2SLS) framework. To adapt Lewbel's idea to this context, we consider two alternatives:

- **Linear Probability Model (LPM):** Treat the binary outcome as a linear probability and apply Lewbel IV estimation directly using OLS.
- **Control Function Approach (CFA):** Incorporate the residuals from the first-stage Lewbel regression into a nonlinear model (e.g., logit), which corrects for endogeneity while maintaining the binary nature of the dependent variable.

1.2.1. Linear probability model

The first alternative consists of treating the model as linear and applying the Lewbel heteroskedasticity-based IV method using Ordinary Least Squares (OLS). This leads to the *Linear Probability Model (LPM)*, where the dependent variable equals 1 if the individual is obese and 0 otherwise.

The estimation follows the two-stage least squares (2SLS) logic within the Lewbel framework:

First stage (instrumenting income):

We first estimate income quintiles as a function of exogenous covariates (age, gender, education, region, urbanization, etc.) and obtain the residuals from this regression. Then, following Lewbel, we construct instruments by interacting the mean-centered exogenous variables with these residuals. These generated instruments capture variation in income that comes from heteroskedasticity in the data.

Next, we re-estimate the first-stage regression of income quintiles on the exogenous covariates and the Lewbel instruments. This provides fitted values of income that are purged of endogeneity.

$$
\begin{aligned}
&Income\_Quintile_i \\
&\quad = \beta_0 + \beta_1 Age_i + \beta_2 Gender_i + \beta_3 Education_i + \beta_4 Region\ of\ residence_i \\
&\quad + \beta_5 Urbanization_i + \beta_6 Income\_work_i + \beta_7 Unemployement\_benefits_i \\
&\quad + \beta_8 Old\_age\_benefits_i + \beta_9 Sickness\_benefits_i + \beta_{10}\, Social\_support_i \\
&\quad + \beta_{11} Family\_allowances_i + \beta_{12} Other\_benefits_i + \beta_{13} Household\_size_i + \varepsilon_{\mathrm{i}}
\end{aligned}
$$

More details about the variables are available in ***Appendix 1***

Second stage (LPM for obesity):

In the second stage, the binary variable *obese* is regressed on the instrumented income quintiles and additional exogenous variables:

$$
\begin{aligned}
Obese_i = \ & \beta_0 + \beta_1 \widehat{Income\_quintile_i} + \beta_2 Education_i + \beta_3 Age_i + \beta_4 Gender_i \\
& + \beta_5 Region\ of\ residence_i + \beta_6 Urbanization_i + \beta_7 Access\ to\ parks_i \\
& + \beta_8 Physical\ activity_i + \beta_9 Nutrition\ habits_i + \varepsilon_{\mathrm{i}}
\end{aligned}
$$

Here, $\widehat{Income\_quintile}$ represents the fitted values from the first stage, ensuring correction for the potential endogeneity of income.

The LPM offers several advantages: it is simple to implement, coefficients are easy to interpret, and it is naturally compatible with instrumental variables ***(Wooldridge, 2010)***. More importantly, it works well with the Lewbel approach, which is designed for linear IV estimation.

However, the LPM has well-known drawbacks. It can predict probabilities outside the [0,1] range and assumes a linear relationship between the explanatory variables and the probability of the outcome. This may not hold in our case: for instance, the change in obesity probability when moving from the

1st to the 2nd quintile may differ from the change between the 4th and 5th quintiles. Assuming constant marginal effects when the true relationship is nonlinear can lead to biased estimates ***(Deke, 2014)***.

Researchers typically use the LPM not for prediction but to estimate causal effects ***(Chen et al., 2023). Deke (2014)*** shows that in randomized experiments, the LPM produces treatment effect estimates that are just as accurate as those obtained from logistic regression. This highlights its practical value in policy evaluation and supports its use as a simple and transparent alternative to nonlinear models.

Nevertheless, given the nonlinear nature of the relationship between obesity and income quintiles, a more appropriate specification is needed. Therefore, in the next section, we turn to the Control Function Approach, which accounts for nonlinearity.

1.2.2. The Control function approach

To address the nonlinearity of our outcome variable (obesity), we use the *control function approach* (also known as the two-stage residual inclusion (2SRI)). This method adapts instrumental variables to nonlinear models. This method applies the same logic as a two-stage regression in the 2SLS. However, the difference lies in the fact that the residuals obtained from the first stage regression are included in the second stage and do not replace the endogenous variable. This allows for control of the endogeneity present and enhances the consistency of the estimates ***(Terza et al., 2008).***

Terza et al. (2008) demonstrated that, although the control function and the 2SLS produce equivalent results in linear models, they generally diverge in nonlinear models. In particular, the control function approach (CFA) remains statistically consistent in nonlinear models such as probit or logit, while 2SLS does not. This makes the control function approach suitable for models with binary dependent variables.

Applied to our case with Lewbel instruments, the procedure consists of the following steps:

In the first stage regression**,** household income quintiles are regressed on exogenous variables likely to affect income. As in the LPM specification, the Lewbel instruments are constructed by exploiting heteroskedasticity in the residuals of this regression. The residuals from the first stage are then added to the second stage equation, capturing the unobserved factors influencing income that may also affect obesity.

In the second stage regression, a logit model is then estimated for obesity, where the main equation includes the endogenous variable (income quintiles), the exogenous controls, and the residuals from the first stage:

$$\begin{aligned} Obese_i = {} & \beta_0 + \beta_1 Income\_quintile_i + \beta_2 Education_i + \beta_3 Age_i + \beta_4 Gender_i \\ & + \beta_5 Region\ of\ residence_i + \beta_6 Urbanization_i + \beta_7 Access\ to\ parks_i \\ & + \beta_8 Physical\ activity_i + \beta_9 Nutrition\ habits_i + \beta_{10} Residuals + \varepsilon_\mathrm{i} \end{aligned}$$

Compared to the LPM, which assumes a linear relationship and uses fitted values in the second stage, the CFA accommodates the nonlinear nature of the obesity outcome and includes the first-stage residuals directly in the logit model.

1.2.3. Limitations

The Lewbel model is useful to address endogeneity when no other external instruments are available, but it relies on some important assumptions and comes with some limitations. First, the Lewbel method relies on heteroskedasticity in the first-stage regression to generate instruments. If there is no heteroskedasticity, the method cannot be applied. Therefore, it is essential to test for heteroskedasticity in the first-stage model before applying this approach.

Second, in order for the LPM and CFA to yield consistent estimates, the first-stage equation must be well specified. This means including all relevant variables that may affect income. In our case, we used all variables available in the HIS database that are likely to influence income, but it is still possible that some important factors (like personal motivation) are missing and unobserved. This could affect the quality of the results.

Third, because the instruments are generated from the model itself, their quality depends heavily on the variables chosen in the first stage. If the exogenous variables are weakly correlated with income, the instruments may also be weak. Therefore, it is important to test the validity and relevance of these generated instruments using standard diagnostics, such as the first-stage F-statistic.

A more difficult assumption to assess is the exclusion restriction, which requires that the instruments affect the outcome variable (obesity) only indirectly through the endogenous variable (income quintile), and not directly. This is particularly problematic when using internally generated instruments, such as those derived from Lewbel's method, since they are not grounded in theory or derived from external sources. Because they are constructed from residuals and exogenous variables, it is not possible to formally test whether they satisfy the exclusion restriction. As a result, the validity of these instruments relies on untestable assumptions, and any conclusions drawn from this approach must therefore be interpreted with caution ***(Bastardoz et al., 2023)***.

The next section presents the empirical results.

Empirical results

In this section, the results from the different models are presented, preceded by some preliminary tests.

1.1. Preliminary tests

Heteroskedasticity :

The Lewbel's method relies on the presence of heteroskedasticity in the first-stage regression errors to generate valid instruments. To test this, a Breusch-Pagan test was performed. The null hypothesis of homoskedasticity was rejected ($p < 0.001$), confirming the presence of heteroskedasticity and supporting the application of Lewbel's method.

Relevance of instruments:

To ensure that the endogeneity of the income variable is properly addressed, the quality of the instruments must be assessed in terms of both relevance and validity.

For the Control Function Approach (CFA), the first-stage regression of income on Lewbel-generated instruments yielded an F-statistic of 346.09 ($p < 0.001$), well above the conventional threshold of 10. This confirms that the instruments are relevant and suitable for use in the CFA.

To test for the relevance of the instruments in the LPM, the Cragg-Donald Wald F-statistic was given. The Cragg-Donald Wald F-statistic for the LPM was only 1.164, far below the threshold. This indicates the instruments are weak in this specification, which can lead to biased and inconsistent estimates. While the Hansen J test failed to reject the null hypothesis of valid overidentifying restrictions *($p = 0.366$)*, the weak instrument problem suggests that results from the LPM specification should be interpreted with caution.

Validity of the instrument:

An important condition is that the instruments must be uncorrelated with the structural error term in the obesity equation. As explained earlier, because these instruments are internally generated from the data rather than grounded in external theory, this assumption cannot be directly tested. Consequently, the validity of the results relies on untestable assumptions, and conclusions should be interpreted with caution.

Model specification:

For the Control function to yield consistent and unbiased estimates, both the first-stage and second-stage equations must be correctly specified. To assess potential misspecification in the first-stage equation, a Ramsey RESET test was conducted. The results indicated a rejection of the null hypothesis (*Prob > chi2 = 0.0001*), suggesting the presence of misspecification in the first-stage model.

For the second-stage logit regression, a link test was performed. The squared predicted value was statistically significant ($p = 0.000$), indicating possible misspecification in the outcome model as well. This suggests that some relevant factors influencing obesity may not be captured, even though a wide range of control variables was included.

Because the first-stage regression is identical for both the CFA and the LPM, the same concerns about specification apply to the LPM.

Endogeneity :

In the control function**,** Endogeneity is tested via the significance of the residual from the first-stage regression, included in the second-stage logit regression. The residual was found to be statistically significant ($p < 0.10$), indicating the presence of endogeneity and confirming the need for correction through instrumental variables.

For the LPM, a Durbin–Wu–Hausman test failed to reject the null of exogeneity ($p = 0.3846$). However, due to the weakness of the instruments in this specification, this result may not be reliable.

1.2. Discussion of results

As a reminder, the goal of this thesis is to analyze whether poverty has an impact on obesity. To investigate this, income quintiles are used as a proxy for poverty, with a particular focus on comparing individuals in the first income quintile (the poorest) to those in the fifth quintile (the richest), which serves as the reference group. By observing how the probability of being obese changes across income levels, especially between these two extremes, we aim at identifying whether and how poverty increases the risk of obesity.

Table 2 presents the results from the 4 model specifications, based on a sample of 9,396 observations.

Column (1) shows the average marginal effects of the naive logit model with no control variables. All income quintiles show a positive and statistically significant association with obesity at the 1% level. In particular, individuals in the first income quintile are, on average, 10.2 percentage points more likely to be obese than those in the fifth quintile. However, this model does not account for any other variables.

Column (2) presents the average marginal effects of the logit regression with controls (e.g., education, age, gender, region, urbanization, lifestyle). After controlling for these factors, the income coefficients remain positive and significant. The marginal effect of the first quintile drops to 5.5 percentage points, suggesting that some of the income-obesity relationship is explained by observable characteristics.

However, this model may still suffer from endogeneity. Therefore, *column (3)* presents the results from the control function approach using Lewbel instruments. The first-stage residual is statistically significant at the 10% level, confirming the presence of endogeneity. The average marginal effect of the lowest income quintile increases to 9.9 percentage points, implying that the logit model of the specification (2) likely underestimated the true impact of poverty on obesity.

Column (4) displays the results of the 2SLS regression using the linear probability model (LPM) with Lewbel instruments. Here, none of the income quintiles are statistically significant, suggesting that income and, by extension, poverty, have no significant impact on obesity. Moreover, education is also insignificant in this specification, which further diverges from both the literature and the other models' results. These inconsistencies are likely due to the presence of weak instruments in the LPM specification, as confirmed by a low Cragg-Donald F-statistic. As weak instruments can produce biased and unreliable estimates, the results from the LPM should be interpreted with caution.

Given the strength of the instruments and the consistency of results with theory and prior research, the control function approach remains the most reliable specification.

Regarding the control variables, education is significant across all models (except the LPM): individuals with higher education levels are less likely to be obese, consistent with findings in the literature. Age shows a small but statistically significant positive effect, and gender also matters, with males slightly more likely to be obese than females.

Interestingly, the level of urbanization and access to parks or recreational public spaces are not statistically significant. This contradicts findings in the literature and suggests that these environmental factors may not play a significant role in obesity within the Belgian context.

Physical activity and snack consumption are both significant at the 1% level. Physical activity has the expected negative effect on obesity. However, the negative coefficient on snack frequency is unexpected. This would suggest that more frequent snacking is associated with a lower probability of obesity. This result may be due to misreporting in the original dataset and should be further investigated.

To summarize, after controlling for both observed factors and endogeneity, the results suggest that poverty, proxied by low income, significantly increases the likelihood of being obese. While the findings are consistent with previous research, they must be interpreted with caution due to potential limitations in the data and identification strategy.

Table 2: Table of results - Adults

	(1) Obese	(2) Obese	(3) Obese	(4) Obese
Income quintiles				
5th income quintile	Ref.	Ref.	Ref.	Ref.
4th income quintile	0.028***	0.018*	0.026**	-0.198
	(0.010)	(0.010)	(0.011)	(0.155)
3rd income quintile	0.060***	0.034***	0.054***	0.123
	(0.011)	(0.011)	(0.015)	(0.143)
2nd income quintile	0.070***	0.027**	0.057***	-0.018
	(0.012)	(0.012)	(0.021)	(0.120)
1st income quintile	0.102***	0.055***	0.099***	0.066
	(0.013)	(0.013)	(0.029)	(0.138)
Highest diploma		-0.017***	-0.010*	-0.008
		(0.004)	(0.006)	(0.019)
Age		0.001***	0.001***	0.001*
		(0.000)	(0.000)	(0.000)
Gender (ref = female)		0.016**	0.018**	0.019**
		(0.007)	(0.007)	(0.009)
Region of residence				
Flemish region		*Ref.*	*Ref.*	*Ref.*
Brussels region		-0.002	-0.005	-0.014
		(0.011)	(0.011)	(0.016)
Walloon region		0.022***	0.019**	0.019
		(0.009)	(0.009)	(0.013)
Level of urbanization		-0.004	-0.005	-0.006
		(0.004)	(0.004)	(0.005)
Access to parks and/or recreational public spaces		-0.024	-0.022	-0.021
		(0.016)	(0.016)	(0.021)
Frequency of sweet and salty snacks		-0.010***	-0.010***	-0.008**
		(0.003)	(0.003)	(0.004)
leisure time physical activity		-0.061***	-0.060***	-0.057***
		(0.006)	(0.006)	(0.009)
Residuals			-0.013*	
			(0.007)	
Observations	9,396	9,396	9,396	9,396

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Source: Own computations based on the data from the HIS database

Complementary analysis & Robustness checks

In this section, we extend the main analysis to explore whether the observed relationship between poverty and obesity holds in two additional contexts: among children and when accounting for health literacy. These robustness checks help verify the stability of the results and provide further insights into how poverty affects different population subgroups.

1.1. Obesity among the youth

A large part of the literature review focuses on the impact of poverty on childhood obesity. Therefore, as a complementary analysis, it is relevant to examine whether growing up in an economically disadvantaged household affects the probability of obesity among children and adolescents in the Belgian sample. Prior studies have found that children from lower-income households face a higher risk of obesity.

We estimate a model similar to the one applied in the main analysis, but with some adjustments:

$$\begin{aligned} Obese_i = \; & \beta_0 + \beta_1 Income\ quintiles_i + \beta_2 highest\_Education_i + \beta_3 Age_i + \beta_4 Gender_i \\ & + \beta_5 Region\ of\ residence_i + \beta_6 Urbanization_i + \beta_7 Nutrition\ habits_i + \varepsilon_i \end{aligned}$$

As explained earlier, obesity in children is not defined the same way as for adults. Obesity is defined using age-specific cutoffs for children and adolescents. The education variable here reflects the highest diploma obtained within the household. Finally, the variables *physical activity* and *access to parks* are excluded, as they are only available for respondents aged 15 and above.

Contrary to the previous analysis focusing on adults, the risk of reverse causality is lower in this case. Household income is determined primarily by the earnings of adults, which are unlikely to be affected by the child's weight status. However, unobserved factors could still confound the relationship. To account for this possibility, we apply both the control function approach (CFA) and the linear probability model (LPM) with Lewbel instruments.

Table 3 presents the results from the same four model specifications using a sample of 3,156 children aged 2 to 17. The same preliminary tests were performed and gave us the same conclusions as in the main analysis. The results must also be interpreted with caution.

Across all models, the coefficients for the 3rd and 4th income quintiles are not statistically significant, suggesting no measurable difference in obesity risk compared to children in the highest income quintile. By contrast, the 1st income quintile is positive and statistically significant in every model, indicating that children in the poorest households are more likely to be obese.

In model (3), which applies the control function approach, being in the 1st income quintile increases the probability of being obese by 8.8 percentage points on average compared to children in the 5th quintile. However, the first-stage residuals are not significant, implying no evidence of endogeneity in this subsample, unlike in the adult analysis.

In the LPM (model 4), only the 1st and 2nd income quintiles are statistically significant, in line with the other models. However, the magnitude of the coefficients is unusually large. For instance, the coefficient is nearly 20 percentage points for the 1st quintile. This raises concerns about bias due to weak instruments and a limited sample size. Additionally, education remains statistically insignificant in the LPM specification, which is unexpected given its well-established link with health outcomes.

As with the main analysis, it is important to note that some relevant variables, such as physical activity in this case, are not included in the model, especially in the first-stage regression. Their omission may limit the accuracy and completeness of the findings.

Table 3: Table of results - Children

	(1) Obese	(2) Obese	(3) Obese	(4) Obese
Income quintiles				
5th income quintile	*Ref.*	*Ref.*	*Ref.*	*Ref.*
4th income quintile	-0.005 (0.010)	-0.009 (0.012)	-0.006 (0.011)	0.296* (0.165)
3rd income quintile	0.013 (0.011)	0.009 (0.013)	0.015 (0.013)	-0.051 (0.078)
2nd income quintile	0.038*** (0.014)	0.030** (0.015)	0.041** (0.019)	0.202** (0.098)
1st income quintile	0.105*** (0.014)	0.067*** (0.016)	0.088*** (0.027)	0.196** (0.093)
Highest education level within household		-0.021*** (0.005)	-0.016** (0.007)	-0.024 (0.024)
Age		-0.006*** (0.001)	-0.006*** (0.001)	-0.005*** (0.001)
Gender (Ref = female)		0.020** (0.009)	0.020** (0.009)	0.009 (0.011)
Region of residence				
Flemish region		*Ref.*	*Ref.*	*Ref.*
Brussels region		-0.017 (0.012)	-0.020 (0.013)	-0.008 (0.021)
Walloon region		0.002 (0.013)	-0.001 (0.013)	0.002 (0.016)
Level of urbanization		0.011** (0.005)	0.010** (0.005)	0.010 (0.007)
Frequency of eating sweet and salty snacks		-0.011*** (0.004)	-0.011*** (0.004)	-0.012** (0.005)
Residuals			-0.007 (0.006)	
Observations	3,156	3,156	3,156	3,156

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Source: Own computations based on the data from the HIS database

Despite some limitations, the results consistently show that children from low-income families are significantly more likely to be obese than those from high-income families. This confirms previous research done and the main analysis, and highlights the importance of addressing poverty early in life to reduce long-term health inequalities.

1.2. Health literacy

As discussed in the literature review, education may influence obesity partly through health literacy. The Health Interview Survey (HIS) includes variables on health literacy, but they are only available for the 2018 wave. For this reason, they were not included in the main analysis. In this section, we restrict the sample to the 2018 data to control for health literacy and examine whether the results from the main analysis remain consistent after controlling for health literacy. This analysis is limited to adults above 18 years old.

We replicate the same four model specifications used in the main analysis, with the addition of a health literacy variable. This variable is defined as a dummy equal to 1 if the respondent has a sufficient level of health literacy and 0 otherwise. This variable is based on a six-question survey in which respondents indicate how easy or difficult it is for them to perform specific health-related actions. Responses are scored as follows: *very easy* = 4; *fairly easy* = 3; *fairly difficult* = 2; *very difficult* = 1.

The six questions asked are:

1. *Judge when you may need to get a second opinion from another doctor*
2. *Use the information the doctor gives you to make decisions about an illness*
3. *Find information on how to manage certain mental health problems, like stress or depression*
4. *Judge if the information on health risks in the media is reliable? (Examples: TV, Internet, or other media)*
5. *Find out about activities that are good for your mental well-being? (Examples: meditation, sport, walking,...)*
6. *Understand information in the media on how to get healthier? (Examples: Internet, newspapers, magazines).*

The final score is calculated by taking the average of the six responses. Based on this score, three levels of health literacy are defined: insufficient (1 ≤ score ≤ 2), limited (2 < score < 3), and sufficient (3 ≤ score ≤ 4). For this analysis, we use the binary indicator ***"Sufficient level of health literacy,"*** which equals 1 if the respondent has a sufficient level of health literacy and 0 otherwise.

Table 4 presents the results from the four model specifications, including the control function approach (CFA) and the linear probability model (LPM) to account for potential endogeneity. The sample consists of 5,024 individuals aged 18 to 102. In model (3), the residuals from the first-stage regression are statistically significant, confirming the presence of endogeneity in this setting.

The coefficients for the income quintiles remain positive and statistically significant in the CFA specification, indicating that poverty continues to be associated with a higher probability of obesity, even after accounting for health literacy. However, the coefficient of the effect is larger than in the main analysis: individuals in the first income quintile are, on average, 12.9 percentage points more likely to be obese than those in the highest quintile (compared to 9.9 percentage points in the main sample).

In contrast, none of the income quintiles are statistically significant in the LPM, as was also the case in the main analysis. Again, this likely reflects problems with weak instruments and potential bias in this model. Health literacy is also not significant in the LPM.

Surprisingly, in the CFA, the coefficient on health literacy is positive and significant, suggesting that individuals with a sufficient level of health literacy are more likely to be obese. This is counterintuitive and may reflect reporting issues in the HIS health for the health literacy variable. A similar concern was previously noted with the snack consumption variable. Moreover, once health literacy is included, the education variable loses statistical significance, which may suggest that part of education's effect on obesity operates through health literacy or due to misspecification.

In summary, after controlling for health literacy, the results remain broadly consistent with the main analysis: poverty, as proxied by income quintiles, continues to have a positive and statistically significant effect on obesity risk.

Table 4: Marginal effects (including health literacy)

	(1) Obese	(2) Obese	(3) Obese	(4) Obese
Income quintiles				
5th income quintile	*Ref.*	*Ref.*	*Ref.*	*Ref.*
4th income quintile	0.031**	0.023	0.034**	-0.155
	(0.014)	(0.015)	(0.014)	(0.232)
3rd income quintile	0.059***	0.033**	0.061***	0.196
	(0.015)	(0.016)	(0.020)	(0.196)
2nd income quintile	0.072***	0.032*	0.076***	-0.130
	(0.016)	(0.017)	(0.029)	(0.191)
1st income quintile	0.106***	0.064***	0.129***	0.145
	(0.018)	(0.019)	(0.040)	(0.186)
Highest diploma		-0.014**	-0.005	-0.008
		(0.006)	(0.008)	(0.023)
Age		0.001***	0.001***	0.001*
		(0.000)	(0.000)	(0.001)
Gender (Ref = female)		0.027***	0.030***	0.027**
		(0.010)	(0.010)	(0.012)
Region of residence				
Flemish region		*Ref.*	*Ref.*	*Ref.*
Brussels region		-0.025*	-0.030**	-0.038*
		(0.014)	(0.014)	(0.020)
Walloon region		0.020	0.015	0.016
		(0.013)	(0.013)	(0.018)
Level of urbanization		0.003	0.002	0.001
		(0.005)	(0.005)	(0.008)
Access to parks and/or recreational public spaces		-0.023	-0.021	-0.029
		(0.022)	(0.022)	(0.028)
Frequency of eating sweet and salty snacks		-0.012***	-0.012***	-0.009
		(0.004)	(0.004)	(0.006)
Leisure time physical activities		-0.060***	-0.059***	-0.051***
		(0.009)	(0.009)	(0.015)
sufficient level of health literacy		0.030***	0.030***	0.023
		(0.011)	(0.011)	(0.014)
Residuals			-0.019*	
			(0.010)	
Observations	5,024	5,024	5,024	5,024

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Source: Own computations based on the data from the HIS database

Conclusion and discussion

To conclude, this thesis aimed to examine whether poverty has an impact on obesity in Belgium. Using income quintiles as a proxy for poverty, the analysis showed that households in the lowest quintiles have a significantly higher probability of being obese compared to those in the highest quintile. This pattern is also true among children: those living in low-income households are more likely to be obese than those in wealthier households. These findings confirm that poverty plays an important role in shaping obesity outcomes in Belgium and highlight the need to address social inequalities as part of obesity prevention strategies.

This study contributes to the limited evidence on the poverty–obesity relationship in Belgium by applying methods that address endogeneity, including the Control Function Approach with Lewbel-generated instruments. This approach is useful when no valid external instruments are available, but it comes with certain limitations.

First, the Belgian Health Interview Survey (HIS) relies on sampling weights to ensure representativeness, but these were not applied in this study. As a result, the findings may not fully reflect the Belgian population. Future research should re-estimate the models with weights to assess the robustness of the results.

Second, the chosen model specifications also have limitations. The *Control Function Approach* (CFA) with Lewbel-generated instruments successfully detected endogeneity in the income variable and proved the impact of poverty on obesity. However, the CFA relies on several assumptions, and potential misspecification may have affected the estimates. In contrast, the *Linear Probability Model* (LPM) produced weak instrument results, indicating that the instruments were not strong enough in this specification. This weakness, combined with the risk of misspecification, limits the reliability of the LPM estimates. Overall, while the Lewbel method is useful when external instruments are lacking, it should not be seen as a substitute for strong external instruments when these are available ***(Baum et al., 2019)***.

Despite these limitations, the results have clear policy implications. Addressing poverty and inequality should be a central component of efforts to limit the obesity epidemic in Belgium. Policies that improve access to affordable and healthy food, create supportive environments for physical activity, and reduce economic barriers to healthy living could help mitigate the higher obesity risk faced by low-income households. Moreover, interventions targeted at children in disadvantaged households may be especially important to break the intergenerational cycle of poverty and poor health.

References

Arendt, J. N. (2004). Does education cause better health? A panel data analysis using school reforms for identification. *Economics of Education Review*, *24*(2), 149–160. https://doi.org/10.1016/j.econedurev.2004.04.008

Bastardoz, N., Matthews, M. J., Sajons, G. B., Ransom, T., Kelemen, T. K., & Matthews, S. H. (2023). Instrumental variables estimation: Assumptions, pitfalls, and guidelines. *The Leadership Quarterly*, *34*(1), 101673. https://doi.org/10.1016/j.leaqua.2022.101673

Baum, C. F., & Lewbel, A. (2019). Advice on using heteroskedasticity-based identification. *The Stata Journal*, 19(4), 757-767.

Berete, F., Gisle, L., Demarest, S., Charafeddine, R., Bruyère, O., Van Den Broucke, S., & Van Der Heyden, J. (2024). Does health literacy mediate the relationship between socioeconomic status and health-related outcomes in the Belgian adult population? *BMC Public Health*, *24*(1). https://doi.org/10.1186/s12889-024-18676-7

Brunello, G., & D'Hombres, B. (2006). Does body weight affect wages? *Economics & Human Biology*, *5*(1), 1–19. https://doi.org/10.1016/j.ehb.2006.11.002

Burkhauser, R. V., & Cawley, J. (2007). Beyond BMI: The value of more accurate measures of fatness and obesity in social science research. *Journal of Health Economics*, *27*(2), 519–529. https://doi.org/10.1016/j.jhealeco.2007.05.005

Chari, R., Warsh, J., Ketterer, T., Hossain, J., & Sharif, I. (2013). Association between health literacy and child and adolescent obesity. *Patient Education and Counseling*, *94*(1), 61–66. https://doi.org/10.1016/j.pec.2013.09.006

Chen, K., Martin, R. S., & Wooldridge, J. M. (2023, August 29). *Another look at the linear probability model and nonlinear index models*. arXiv.org. https://arxiv.org/abs/2308.15338

Cole, T. J., Bellizzi, M. C., Flegal, K. M., & Dietz, W. H. (2000). Establishing standard definition for child overweight and obesity worldwide: International survey. *ResearchGate*. https://www.researchgate.net/publication/12520427_Establishing_standard_definition_for_child_overweight_and_obesity_worldwide_International_survey

Cole, T. J., & Lobstein, T. (2012). Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. *Pediatric Obesity*, *7*(4), 284–294. https://doi.org/10.1111/j.2047-6310.2012.00064.x

Deke, J. (2014). Using the Linear Probability Model to Estimate Impacts on Binary Outcomes in Randomized Controlled Trials. *HHS Office of Adolescent Health.*

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

Dhurandhar, E. J. (2016). The food-insecurity obesity paradox: A resource scarcity hypothesis. *Physiology & Behavior*, *162*, 88–92. https://doi.org/10.1016/j.physbeh.2016.04.025

Dinsa, G. D., Goryakin, Y., Fumagalli, E., & Suhrcke, M. (2012). Obesity and socioeconomic status in developing countries: a systematic review. *Obesity Reviews*, *13*(11), 1067–1079. https://doi.org/10.1111/j.1467-789x.2012.01017.x

Drieskens, S., Charafeddine, R., Vandevijvere, S., De Pauw, R., & Demarest, S. (2024). Rising socioeconomic disparities in childhood overweight and obesity in Belgium. *Archives of Public Health*, *82*(1). https://doi.org/10.1186/s13690-024-01328-y

Dwyer, J. T., Melanson, K. J., Sriprachy-Anunt, U., Cross, P., & Wilson, M. (n.d.). *Table 4, Classification of weight status by body mass Index (BMI) - Endotext - NCBI Bookshelf*. https://www.ncbi.nlm.nih.gov/books/NBK278991/table/diet-treatment-obes.table4clas/

Eurostat. (n.d.). *At risk of poverty or social exclusion (AROPE)*. Statistics Explained. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:At_risk_of_poverty_or_social_exclusion_(AROPE)

Federal Planning Bureau. (2024, November 29). *Risque de pauvreté ou d'exclusion sociale (i01)*.https://www.indicators.be/fr/i/G01_PSE/Risque_de_pauvret%C3%A9_ou_d%27exclusion_sociale_%28i01%29

Ferguson, H., Bovaird, S., & Mueller, M. (2007). The impact of poverty on educational outcomes for children. *Paediatrics & Child Health*, *12*(8), 701–706. https://doi.org/10.1093/pch/12.8.701

Finkelstein, E. A., & Strombotne, K. L. (2010). The economics of obesity. *American Journal of Clinical Nutrition*, *91*(5), 1520S-1524S. https://doi.org/10.3945/ajcn.2010.28701e

Friis, K., Lasgaard, M., Rowlands, G., Osborne, R. H., & Maindal, H. T. (2016). Health literacy mediates the relationship between educational attainment and health behavior: a Danish Population-Based study. *Journal of Health Communication*, *21*(sup2), 54–60. https://doi.org/10.1080/10810730.2016.1201175

Gorasso, V., Moyersoen, I., Van Der Heyden, J., De Ridder, K., Vandevijvere, S., Vansteelandt, S., De Smedt, D., & Devleesschauwer, B. (2022). Health care costs and lost productivity costs related to excess weight in Belgium. *BMC Public Health*, *22*(1). https://doi.org/10.1186/s12889-022-14105-9

Kelishadi, R. (2014). Health impacts of Obesity. *Pakistan Journal of Medical Sciences*, *31*(1). https://doi.org/10.12669/pjms.311.7033

Kim, Y. (2016). The long-run effect of education on obesity in the US. *Economics & Human Biology*, *21*, 100–109. https://doi.org/10.1016/j.ehb.2015.12.003

Komlos, J., & Rashad, K. (2014). *The Oxford Handbook of Economics and Human Biology*. Oxford University Press. https://books.google.be/books?hl=fr&lr=&id=h-smDAAAQBAJ&oi=fnd&pg=PA317&dq=poverty+and+obesity+correlation&ots=IOBGaKGjjJ&sig=yV97ZA23HJA2wxhXIyNbdiw9P60&redir_esc=y#v=onepage&q=poverty%20and%20obesity%20correlation&f=false

Kowaleski-Jones, L., Zick, C., Smith, K. R., Brown, B., Hanson, H., & Fan, J. (2017). Walkable neighborhoods and obesity: Evaluating effects with a propensity score approach. *SSM - Population Health*, *6*, 9–15. https://doi.org/10.1016/j.ssmph.2017.11.005

Levine, J. A. (2011). Poverty and obesity in the U.S. *Diabetes*, 60(11), 2667–2668. https://doi.org/10.2337/db11-1118

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business and Economic Statistics*, *30*(1), 67–80. https://doi.org/10.1080/07350015.2012.643126

Mahase, E. (2022). Obesity: No European country is on track to halt rising levels by 2025, WHO warns. *BMJ*, o1107. https://doi.org/10.1136/bmj.o1107

Members' Research Service. (2025, March 4). Obesity in the EU: An ongoing epidemic. *Epthinktank*. https://epthinktank.eu/2025/03/04/obesity-in-the-eu-an-ongoing-epidemic/

News-Medical. (2023, January 23). *What is Obesity Paradox*? https://www.news-medical.net/health/What-is-obesity-paradox.aspx

NHLBI, NIH. (2022, March 24). *Causes and risk factors*. https://www.nhlbi.nih.gov/health/overweight-and-obesity/causes

O'Neill, D. (2015). Measuring obesity in the absence of a gold standard. *Economics & Human Biology*, *17*, 116–128. https://doi.org/10.1016/j.ehb.2015.02.002

Pirgon, Ö., & Aslan, N. (2015). The role of urbanization in childhood obesity. *Journal of Clinical Research in Pediatric Endocrinology*, *7*(3), 163–167. https://doi.org/10.4274/jcrpe.1984

Ritchie, H., & Roser, M. (2017, August 17). *Obesity*. Our World in Data. https://ourworldindata.org/obesity

Salmasi, L., & Celidoni, M. (2017). Investigating the poverty-obesity paradox in Europe. *Economics & Human Biology*, *26*, 70–85. https://doi.org/10.1016/j.ehb.2017.02.005

Sciensano.(2020, July 1). Determinants of Health: Weight status, Health Status Report, Brussels, Belgium, https://www.healthybelgium.be/en/health-status/determinants-of-health/weight-status

Sciensano (2023). Food consumption survey 2022-2023: Weight status and weight related behaviors in the Belgian population, Brussels, Belgium.

Sciensano. (2023, March 6). Factsheets: Cost of overweight and obesity, Health Status Report, Brussels, Belgium, https://www.belgiqueenbonnesante.be/fr/etat-de-sante/factsheets/cost-of-overweight-and-obesity

Sciensano. (2023, March 21). Factsheets: Past, present, and future trends of overweight and obesity in Belgium, Health Status Report, Brussels, Belgium, https://www.belgiqueenbonnesante.be/fr/etat-de-sante/factsheets/past-present-and-future-trends-of-overweight-and-obesity-in-belgium

Sciensano. (2024, June) Weight status and eating disorders: Overweight and obesity (BMI), Food Consumption Survey 2022-2023, Brussels, Belgium, https://www.sciensano.be/en/results-food-consumption-survey-2022-2023/weight-status-and-eating-disorders/overweight-and-obesity-bmi

Smets, V., & Vandevijvere, S. (2022). Changes in retail food environments around schools over 12 years and associations with overweight and obesity among children and adolescents in Flanders, Belgium. *BMC Public Health*, *22*(1). https://doi.org/10.1186/s12889-022-13970-8

Statbel. (2025, January 29). *Risque de pauvreté ou d'exclusion sociale*. https://statbel.fgov.be/fr/themes/menages/pauvrete-et-conditions-de-vie/risque-de-pauvrete-ou-dexclusion-sociale#news

Streel, S., Donneau, A., Hoge, A., Majerus, S., Kolh, P., Chapelle, J., Albert, A., & Guillaume, M. (2015). Socioeconomic impact on the prevalence of cardiovascular risk factors in Wallonia, Belgium: a

Population-Based study. *BioMed Research International*, *2015*, 1–10. https://doi.org/10.1155/2015/580849

Terza, J. V., Basu, A., & Rathouz, P. J. (2008). Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics*, *27*(3), 531–543. https://doi.org/10.1016/j.jhealeco.2007.09.009

Van Der Heide, I., Wang, J., Droomers, M., Spreeuwenberg, P., Rademakers, J., & Uiters, E. (2013). The relationship between health, education, and health literacy: Results from the Dutch Adult Literacy and Life Skills Survey. *Journal of Health Communication*, *18*(sup1), 172–184. https://doi.org/10.1080/10810730.2013.825668

Wang, M. L., Narcisse, M., & McElfish, P. A. (2022). Higher walkability associated with increased physical activity and reduced obesity among United States adults. *Obesity*, *31*(2), 553–564. https://doi.org/10.1002/oby.23634

Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. *MIT press*.

World Bank. (2023, August 2). *Where do the poor live?* World Development Indicators. https://datatopics.worldbank.org/world-development-indicators/stories/where-do-the-poor-live.html

World Health Organization: WHO. (2024, March 1). *Obesity and overweight*. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

Żukiewicz-Sobczak, W., Wróblewska, P., Zwoliński, J., Chmielewska-Badora, J., Adamczuk, P., Krasowska, E., Zagórski, J., Oniszczuk, A., Piątek, J., & Silny, W. (2014). Obesity and poverty paradox in developed countries. *Annals Of Agricultural And Environmental Medicine*, *21*(3), 590-594. https://doi.org/10.5604/12321966.1120608

Data used

This thesis was based on the data from the Health Interview Survey (HIS) provided by Sciensano.

Sciensano, OD Public health and surveillance (2020). Health Interview Survey 2018 [Data file and code book]. Obtainable under the condition from the Sciensano Web site: https://www.sciensano.be/fr/projets/enquete-de-sante/enquete-de-sante-procedure-de-demande-de-microdonnees

Appendix

Appendix 1: table of variables

Variable Name	*Type of variable*	*Possible values*	*Years available*
Obese	Binary	1 = Obese 0= Not obese	1997-2018
Reported equivalent household income (Belgian weighted quintiles)	Categorical	1= First income quintile 2= Second income quintile 3= Third income quintile 4= Fourth income quintile 5= Fifth income quintile	1997-2018
Education: highest diploma	Categorical	1= No diploma or primary education 2= Lower secondary 3= Higher secondary 4= Higher education	1997-2018
Highest level of education within the household	Categorical	1= No diploma or primary education 2= Lower secondary 3= Higher secondary 4= Higher education	1997-2018
Age	Continuous	Value between 2 and 102 years old	1997-2018
Gender	Binary	1= Male 0 = Female	1997-2018
Region of residence	Categorical	1= Flemish region 2= Brussels region 3= Walloon region	1997-2018
Level of urbanization	Categorical	1= Rural 2= Urbanized municipalities 3= Suburban / banlieues 4= Big cities and dense agglomerations	1997-2018
Frequency of eating sweet and salty snacks	Categorical	1= Never 2= Less than once a week 3=1 to 3 times a week 4= 4 to 6 times a week 5= Once or more a day	2013-2018
Leisure time Physical activity	Categorical	1= Sedentary activities 2=Sport < 4 hours / light activities 3= Intensive training / sport ≥ 4 hours per week	1997-2018
Access to parks or other green or recreational public places	Binary	1= Yes 0= No	2013-2018
Income from work	Binary	1=Yes 0=No	2008-2018
Unemployment benefits	Binary	1=Yes 0=No	2008-2018
Old-age pension or survivor’s benefits	Binary	1=Yes 0=No	2008-2018
Sickness or disability benefits	Binary	1=Yes 0=No	2008-2018
Social support (OCMW/CPAS)	Binary	1=Yes 0=No	2008-2018
Family - child(ren) related allowances	Binary	1=Yes 0=No	2008-2018
Other regular benefits (rental income, annuity)	Binary	1=Yes 0=No	2008-2018
Overcrowded household (household size)	Binary	1=Yes 0=No	1997, 2013-2018
Sufficient level of Health literacy	Binary	1=Yes 0=No	2018

Source: Health Interview Survey

EXECUTIVE SUMMARY

Obesity is rising rapidly in Europe and represents a major public health and social challenge. While the link between poverty and obesity is well-documented in the United States, evidence for Belgium remains limited. This thesis investigates whether poverty has an impact on obesity in Belgium and through which mechanisms it influences it.

The analysis uses data from the Belgian Health Interview Survey (HIS) of Sciensano, with poverty measured through household income quintiles and obesity assessed using the Body Mass Index (BMI). To address endogeneity concerns, Lewbel's method was applied.

The results show that individuals in the lowest income quintiles have a significantly higher probability of being obese compared to those in the highest quintile. This pattern is also observed among children. These findings highlight the need for integrated policies addressing both poverty and public health.

KEYWORDS: Obesity, poverty, Lewbel model, control function approach, linear probability model, endogeneity, Belgium

WORD COUNT: 10,433

Ecole de Gestion de l'Université de Liège