URBAN-RURAL DIFFERENCES IN OVERWEIGHT AND OBESITY IN UKRAINE

by

Ivan Shcherbaniuk

A thesis submitted in partial fulfillment of
the requirements for the degree of

MA in Economic Analysis

Kyiv School of Economics

2025

Thesis Supervisor: ______________________________ Professor Maksym Obrizan

Approved by ____
Head of the KSE Defense Committee, Professor

____

____

____

Date ____

Kyiv School of Economics

Abstract

URBAN-RURAL DIFFERENCES
IN OVERWEIGHT AND OBESITY
IN UKRAINE

by Ivan Shcherbaniuk

Thesis Supervisor: Professor Maksym Obrizan

Obesity is one of the major public health concerns of the 21st century. In Ukraine, around 60% of adults are overweight, with nearly 25% of them being obese – it causes a huge burden on the healthcare system and economy overall. This thesis investigates urban-rural differences in obesity and evaluates upstream determinants contributing to those differences. It uses the 2019 WHO STEPS survey in Ukraine and both Linear Probability Model and Logistic Regression to retrieve insights from the data.

The results confirm a statistically significant association: urban residents are almost 5 percentage points less likely to be obese if to compare with rural residents – even after adjusting for health behaviors and mental health. This rural-urban gap aligns with findings from neighboring countries like Poland.

The paper discusses policy recommendations to address these disparities. It outlines such intervention strategies as improving nutrition literacy, expanding access to healthcare, and promoting policies tailored for rural populations. The ultimate goal of this work is to provide valuable insights for public health policymakers responsible for socioeconomic well-being in Ukraine.

TABLE OF CONTENTS

LIST OF FIGURES

Number *Page*

LIST OF TABLES

ACKNOWLEDGEMENTS

I want to express my deep gratitude to all KSE professors for their challenging yet rewarding courses – if not that foundation in economics and econometrics, it won't be possible to accomplish this thesis.

I extend special thanks to my thesis advisor Professor Maksym Obrizan who helped me to formulate the concept for this work, directed me to the best available data sources and provided valuable feedback on every stage. I'm also sincerely grateful to all research workshop professors who patiently reviewed my work and diligently helped me to improve it.

Finally, I want to thank the entire KSE community, my groupmates, and everyone involved. Truly, the process of studying and discovering something new together is magic and I'm happy and thankful for this journey.

LIST OF ABBREVIATIONS

AME: Average Marginal Effects.

BMI: Body Mass Index.

BP: Blood Pressure.

COSI: Childhood Obesity Surveillance Initiative.

DBP: Diastolic Blood Pressure.

ILO: International Labour Organization.

LMIC: Low- and Middle-Income Countries.

LPM: Linear Probability Model.

NCD: Non-Communicable Disease.

NHANES: National Health and Nutrition Examination Survey.

NGO: Non-Governmental Organization.

OLS: Ordinary Least Squares.

OR: Odds Ratio.

PSU: Primary Sampling Unit.

PURE: Prospective Urban and Rural Epidemiology Study.

SAGE: Study on Global Ageing and Adult Health.

SBP: Systolic Blood Pressure.

SE: Standard Error.

SEBS: Socioeconomic Burden Score.

SSU: Secondary Sampling Unit.

STEPS: STEPwise Approach to NCD Risk Factor Surveillance.

TackSHS: Tackling Second-Hand Tobacco Smoke Survey.

TSU: Tertiary Sampling Units.

VIF: Variance Inflation Factor.

WC: Waist Circumference.

WHR: Waist-to-Hip Ratio.

WHtR: Waist-to-Height Ratio.

WHO: World Health Organization.

Chapter 1

INTRODUCTION

Health is not everything, but without health, everything is nothing.

Arthur Schopenhauer

Health is the most valuable asset we have at our disposal. Not only is it the most valuable per se, but it is also a prerequisite for everything else. One cannot fully enjoy consumption if constrained by health conditions, nor can they utilize their production capabilities to the full extent (Zweifel et al. 2009).

The question asking what can impair our health naturally arises here, and among the long list of reasons, obesity occupies one of the top positions. Obesity is a chronic disease defined as an abnormal or excessive fat accumulation which may impair health (World Health Organization 2024).

The Ministry of Health reports that almost 60% of adults in Ukraine are overweight, with nearly 25% of them being obese (WHO STEPS 2019). The rate is one of the highest among Eastern European countries. What's even more striking is that only 50 years ago, in 1975 just a little more than 12% of Ukrainians suffered from obesity (World Health Organization 2020).

There is an upward trend not just in Ukraine but globally as well – it is predicted that by 2030 one in 5 women and one in 7 men will be living with obesity, equating to over 1 billion people worldwide. This makes obesity one of the major public health concerns of the 21st century (World Obesity Federation 2022).

What caused the proliferation of obesity at such speed? Both neoclassical (Phillipson et al.) and behavioral theory (Cutler et al.) link the spread of obesity to technological progress (Specchia et al. 2015). On one hand, it led to decreased physical activity and a higher number of sedentary jobs, on the other hand, it dropped food prices, and the time spent on food preparation.

But perhaps even more important for us is not the cause of obesity but what obesity itself can cause and the list here is alarming. Physicians have established long time ago the link between obesity and high blood pressure, stroke, type 2 diabetes, osteoarthritis, heart and liver diseases (Dixon et al. 2019). It is also confirmed that excess body weight can increase the risk of 13 types of cancer (World Health Organization 2022).

Obesity harms not just our body but also causes depression, anxiety, and negatively affects mental well-being in general (Stival et al. 2022). In some cases, people suffer from obesity stigma and discrimination, which hinders job seeking and makes them feel less confident and comfortable – hence productive – in the workplace (Diamantis et al. 2022).

These are some of the clinical consequences, but within the scope of this paper we're of course more interested in the economic aspect. In terms of numbers, overweight and obesity are responsible for about USD 1 trillion in annual healthcare costs worldwide. The breakdown of the costs is highly complex, as shown in Figure 1. In Europe, this translates to approximately USD 220 billion, representing 13.6 % of total healthcare expenditures (Diamantis et al. 2022). In the case of Ukraine, researchers from the World Obesity Federation forecast that by 2060 Ukraine might lose $21 billion due to obesity – almost 6% of its GDP (Okunogbe et al. 2022).

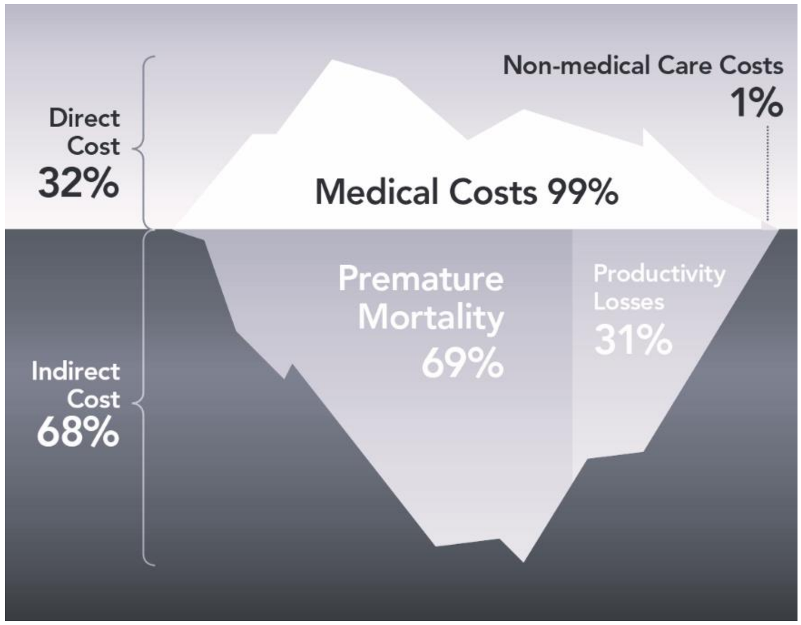

Figure 1. Economic costs of the consequences of overweight and obesity
Source: World Obesity Federation 2022

These devastating consequences can be mitigated by effective public health policies. First, policymakers need to identify the population groups with increased risk of obesity, and second, they must act upon socioeconomic drivers leading to the problem.

Hence, the objective of this paper is twofold. First, to *estimate rural-urban disparities* in obesity prevalence in Ukraine, and second, to *explore upstream determinants* for overweight and obesity in cities and villages. These upstream determinants operate at a higher level if to compare with the individual characteristics (e.g., genetics, diet, and physical activity behaviors) and have proven to be more effective and sustainable as they deal with underlying mechanisms (root causes) contributing to obesity. They can include:

- Living conditions: Access to healthy food, reliable transportation, and stable housing;

- Economic stability: Income, wealth, and working conditions;
- Social environment: Discrimination, and unequal treatment;
- Government policies: Laws and regulations that can affect health;
- Physical environment: The environment where people live, learn, work, and play (Vandevijvere et al. 2023).

If we understand the *specific groups* we need to target and the *underlying factors*, we should be able to design effective interventions – this is the ultimate goal.

The preliminary hypothesis suggests that the rural population is more obese if to compare with urban residents. Two recent studies done in Poland (Stos et al. 2022, Zatonska et al. 2021) provide evidence for this, but for the case of Poland. In Ukraine, to the best of the author's knowledge, specific studies on urban-rural disparities are scarce.

As a baseline approach to test the hypothesis, Linear Probability Model (LPM) will be used. It allows to estimate the probability of being obese or not based on rural/urban residence, controlling for sociodemographic factors, smoking, drinking, and exercise attributes. Logistic Regression (Logit) will also be employed for more in-depth analysis to ensure the consistency of results.

This paper will use the 2019 WHO STEPS survey in Ukraine. The resulting set for Ukraine includes 4 409 observations which will be further refined and adjusted to the purposes of this study.

The structure of the study is as follows: Chapter 2 presents the literature review. Chapter 3 describes the data and how it was prepared for the analysis. The methodology is described in Chapter 4. Estimation results along with testing outcomes are all discussed in Chapter 5. Finally, Chapter 6 summarizes the main findings, discusses possible policy instruments for addressing the urban-rural gap, and outlines possible areas for further research in this field.

Chapter 2

LITERATURE REVIEW

Urban-rural perspective of obesity has been researched in different settings. We will build our literature review in a top-down manner starting from global powers, then observing what's going on in Europe overall, and in neighboring countries in particular. This flow was chosen just to streamline the review. Separately we'll stop on WHO reports, which also constitute an important component of the knowledge base.

2.1 Global perspective

Expectedly, obesity has been researched perhaps the most in the United States. Economic evaluations and studies that include children under the age of 18 are more common. Meanwhile, "the first study that directly examines individual- and neighborhood-level mediators [education, income, etc.] of overweight/obesity disparities between rural and urban adult Americans" – as Wen and colleagues describe their paper – was published just in 2017.

The authors worked with the results of the 2003 – 2008 National Health and Nutrition Examination Survey (NHANES), encompassing more than 10 000 participants aged 20 – 64 years old. They applied multilevel logistic regression and discovered that, all else being equal, rural individuals are 36% more likely to be obese than urban individuals. This is an important insight in itself, but they then decomposed this raw association and began controlling for education, income, and built environment features (e.g., access to green spaces or healthy food options). As a result, the rural-urban obesity gap reduces by 94%. This means that not the location, but rather systematic differences in socioeconomic status (SES) contribute to obesity (Wen et al. 2018).

"A Social-Ecological Review of the Rural versus Urban Obesity Disparity" – much like the paper above – addresses these factors across multiple levels: individual, interpersonal/relationship, community/physical environment, and societal/policy levels. Refer to Figure 2.

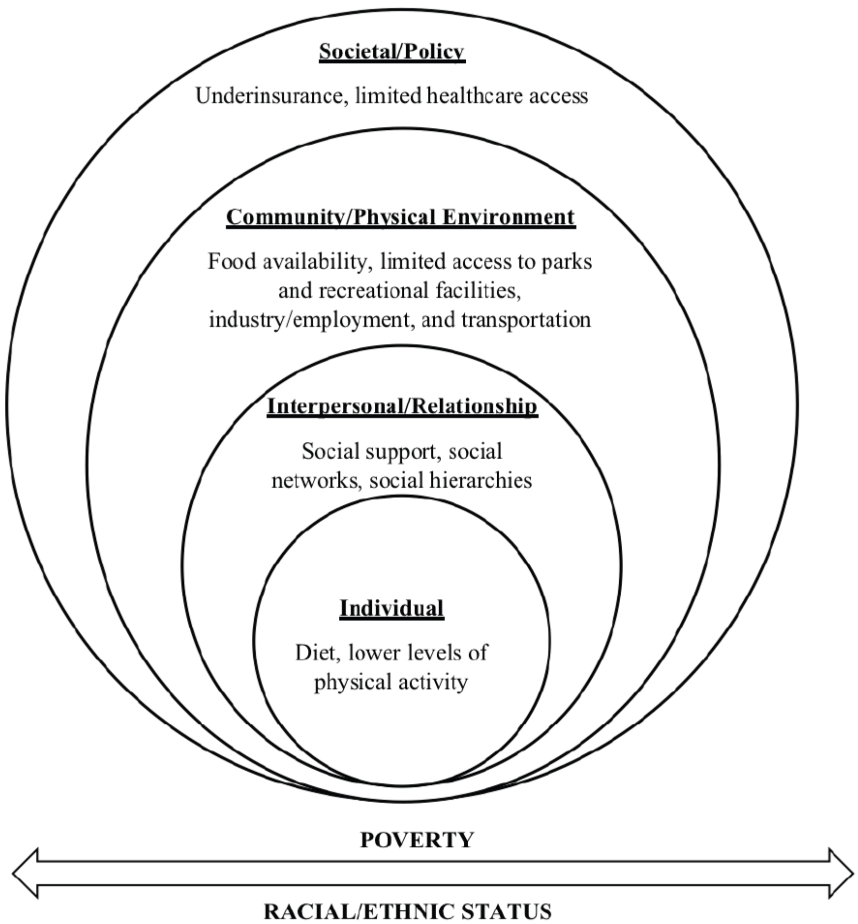

Figure 2. The US Centers for Disease Control and Prevention's model of the factors contributing to the rural/urban disparity in obesity

Source: Dixon et al. 2019

This is a striking fact, but you can pick any of these 4 levels and rural communities will always be worse off: more fats and sugars, less fruits, no regular physical activities, no social support of beneficial health behaviors, lack of nearby grocery stores, built environment components, like walking or biking trails.

Two factors interact across levels: poverty and racial/ethnic status. Poor individuals and those from racial/ethnic minority groups are disproportionally burdened by obesity. For example, both in rural and urban communities almost half of non-Hispanic black adults are obese (48%) (Dixon et al. 2019).

Summing up what these two papers show us through the lens of current research, we can highlight the fact that in the US rural population is more obese, but it is SES which primarily drives the prevalence of obesity.

Meanwhile, in China – the other world's largest economy – urban obesity prevails (NCD-RisC 2019). In July 2021, Wang et al. presented detailed analyses of 6 consecutive national health surveys done between 2004 and 2018. They were especially interested in the breakpoint which happened in 2010 when Chinese authorities introduced a set of national non-communicable disease prevention programs. Since then, obesity prevalence has slowed down substantially for urban men and women, only moderately for rural men, and continued steadily for rural women – the most problematic category to deal with.

Another interesting insight is that more educated women are less obese compared to less educated, but the inverse is true for men. Overall, after questioning more than 746 000 participants, researchers estimated that 85 million adults aged 18 – 69 years to be obese in China in 2018, three times (!) as many as in 2004 (Wang et al. 2021).

Not just global leaders like the US but also low- and middle-income countries (LMIC) have been researched recently to deduce if there any urban-rural differences in overweight and obesity. We won't cite them here in details for the sake of time and space, but what is crucial for the current research is that all these studies provide evidence that in less-developed Asian and African countries the urban population is more obese, which is inverse to the United States (Thapa et al. 2021, Ajayi et al. 2016).

2.2 European perspective

Recent pan-European studies, to the best of our knowledge, do not use urban-rural attribute. Yet, what they've discovered might be of interest for the current study because they discuss related obesity determinants.

Stival and others used data from the TackSHS (Tackling second-hand tobacco smoke) survey conducted in 2017 – 2018 in 12 European countries. After interviewing almost 11 000 participants, researchers estimated that half of the adult population in Europe is overweight (53%, to be precise) and almost 1 in every 8 is obese (13%, to be precise).

As shown in Figure 3, obesity prevalence is the lowest in Italy (7.5%) and France (8.8%) – the authors attributed this to the Mediterranean diet.

Although being a part of Mediterranean basin, Spain and Greece both show higher values of obesity due to a shift to a more Western-type diet characterized by a higher consumption of meat and dairy products.

As we can see, there is a large variability across countries, but what is important for the current paper is that we can derive a strong trend for Eastern Europe, with Romania (21,1%) being the most obese (still 2 times less than in the USA with its 43% of obesity prevalence).

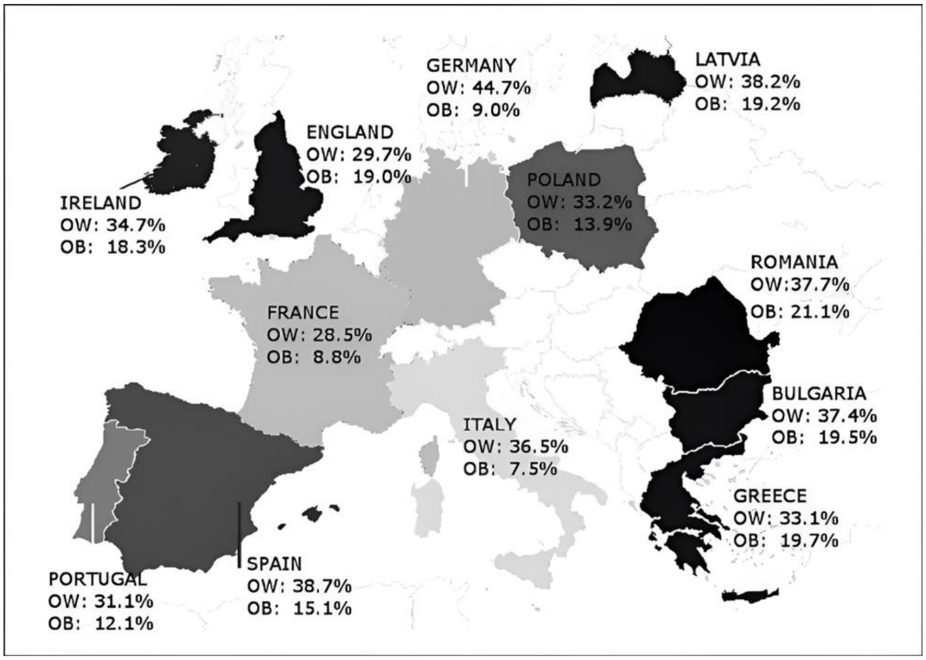

Figure 3. Percent prevalence of overweight and obesity among adults from 12 European countries

Source: Stival et al. 2022

The cause-and-effect chain roughly looks like this – in Eastern Europe, there is a lower level of economic growth, which consequently leads to lower socioeconomic status, resulting in:

- Greater difficulties in providing proper access to healthy food.
- Poor level of health literacy, health education and education overall.
- Higher probability of living in areas where non-occupational physical activity is not promoted (Stival et al. 2022).

These conclusions are also backed by another more recent study by Diamantis et al., published in October 2022. This study utilized the Feel4Diabetes survey, which included more than 19 000 adults across 6 European countries: Bulgaria, Hungary, Finland, Belgium, Greece, and Spain.

What is particularly interesting about this paper is that the authors brought the discussion around socioeconomic factors to another level and introduced SEBS – the socioeconomic burden score. The building blocks for SEBS are:

- Occupational status: employed (both full- and part-time, being retired or student), unemployed.
- Educational level: duration of studies less than 12 years or more than 12 years.
- Income insecurity: ease in covering household cost, difficulty in covering household costs.

The score itself was calculated by adding 1 point every time a participant indicated 1) unemployment or 2) less than 12 years of education or 3) difficulty in securing one's income, with a minimum score of 0 and a maximum of 3. Key fact is that the increase of SEBS has been constantly associated with increased overweight/obesity likelihood, with scores 1, 2 and 3 having OR 1.43 (95% CI 1.33, 1.54), 1.76 (1.62, 1.92) and 1.99 (1.76, 2.24). So, if you are unemployed, uneducated, and poor, there are 2 times more chances that you'll end up being obese.

In the previous sub-section when talking about the US and LMICs we highlighted different tendencies in obesity. Interestingly, Diamantis et al. show that the association between higher SEBS and higher likelihood of obesity work across all examined countries – even thought they were categorized differently. Bulgaria and Hungary are low- to middle income countries, Finland and Belgium are high-income, and Greece and Spain are high-income under austerity measures (Diamantis et al. 2022).

2.3 Studies from neighboring countries

Among neighboring countries, Poland probably would be the most interesting for us. In March 2021, Zatonska et al. published their analysis of the PURE (Prospective Urban and Rural Epidemiology) study. It covers more than 2000 people from Lower Silesian voivodeship. What was unique about this study is the recruitment of these 2000 participants – both urban and rural population were carefully included, as one of the objectives of the study was to address health inequalities. The authors found that 1) rural place of residence, 2) age, and 3) educational level were significantly associated with increased odds for obesity. A logistic regression model showed that the odds for obesity were almost 2 times higher among rural residents if compared with urban residents (odds ratio (OR) = 1.79, 95% CI = 1.48 – 2.16) (Zatońska et al. 2021).

Traczyk and others analyzed the results of another interesting study conducted in Poland in 2017 – 2020. Again, 2000 individuals were evaluated. However, what's special about this study is that it was assessing the prevalence of not just general but also abdominal obesity and overweight – therefore, weight and height were measured, but also waist circumference. It is especially important that researchers paid attention to the distribution of fat in the body because it is much more dangerous when it surrounds the organs in the abdominal cavity, worsening their functioning. Excess body weight was found in 51% of respondents, abdominal overweight – in 21.2%, and abdominal obesity in 27.2% of respondents. What's interesting is that men are more likely to develop excess body weight, but women are more likely to develop abdominal obesity (women – 39.6%, men – 14.1%) (Traczyk et al. 2023).

But probably the most fascinating and controversial study on obesity in Poland was published in January 2022. Stos and others claim that most of their predecessors were focused on selected towns and regions (we saw it on the example of Lower Silesian voivodeship), whereas they carried out their research on a representative nationwide sample of 1800 individuals aged 18+. The prevalence of overweight was 42.2%, of obesity – 16.4%. What is particularly interesting is that out of 11 factors they analyzed in the restricted multivariate logistic regression, only 5 turned out to be significant – so greater odds of overweight/obesity had:

- Males (OR = 2.44, 95% CI = 2.00 – 2.99; p < 0.001);
- Individuals with at least one chronic disease (OR = 1.51; 95% CI = 1.11 – 2.07; p = 0.009);
- Occupationally active individuals (OR = 1.50; 95% CI = 1.17 – 1.93; p <0.001);
- Those living in rural areas (OR = 1.32; 95% CI = 1.07 – 1.63; p = 0.008);
- Older participants (OR = 1.04; 95% CI = 1.03 – 1.04; p <0.01).

Contrary to all the aforementioned papers, Stos et al. did not find a statistically significant association between odds of overweight/obesity and 1) educational level, 2) financial situation, and 3) physical activity level.

The study also revealed that marital status, having children under 18 years of age, and living alone were not important (Stoś et al. 2022). This is something that we will also try for the Ukrainian context within this paper.

2.4 WHO reports

As a concluding part of literature review, we'd like to mention some important WHO's and other international organizations' reports.

Probably the most interesting recent WHO contribution is WHO European Regional Obesity Report 2022. The authors extensively discussed obesogenic environments and how they drive obesity. Traditionally, these environments include food and built environments, but the authors also raised the question of digital food environments and how via digital marketing, online supermarkets and modern meal delivery apps they affect end users.

Within the context of rural obesity, the authors discuss health literacy as an unrecognized determinant. They agree that income and wellbeing are important, but at the same time claim that the obesity problem often boils down to trivial public unawareness, though it may seem strange to many readers.

The authors conclude the report by recommending a suite of population-level interventions – they agree that tackling obesity in individuals is important, yet preventing obesity in populations should be governments' objective (World Health Organization 2022).

We must also mention the WHO STEPS Survey, which will be discussed in detail in the next chapter, as it served as a source of obesity data (World Health Organization 2020).

Chapter 3

DATA

3.1 Data source

The study used data from the 2019 WHO STEPS survey in Ukraine. STEPS is the survey of NCD (non-communicable disease) risk factors, so it is a health examination study – not just an ordinary demographic study with health-related questions – which makes it a perfect data source.

Although two other large studies conducted in Ukraine in 2019 – GATS and the Health Index Survey – also collected data on NCD risk factors, they had only the interview component and hence self-reported measures. In contrast, STEPS incorporates examination studies (BP, height, weight and BMI, hip and waist circumference, and blood glucose and lipids) and objective health measures.

The target population consisted of men and women aged 18 to 69 years, urban and rural, resident in the country. The results of the study can be generalized to the entire population of Ukraine, thanks to the elaborate multistage cluster sampling approach with random selection of units at each stage, which was applied as follows:

- 66 electoral *districts* were selected randomly out of 199 as a primary sampling unit (PSU).
- A list of all electoral *units* (secondary sampling units, SSUs) within the selected districts was created – approximately 10 000 units in total. From this list, 449 urban electoral units and 193 rural electoral units were selected (the population of Ukraine is spread unevenly (urban, 69.4%; rural, 30.6% in 2019)).

- A list of *households* (tertiary sampling units, TSUs) in the 642 selected electoral units was drawn up, and 12 households were selected randomly from each unit.
- Finally, within each selected household *resident participants* were chosen using the eSTEPS mobile application.

As a result, 4 409 respondents participated in the survey.

3.2 Dataset cleaning and preparation

In the first step, we excluded pregnant women and underweight cases (BMI < 18.5 kg/m^2) to focus just on the contrast between excessive weight and normal weight. This is a standard step across all similar studies (Wen et al. 2018, Stos et al. 2022, Diamantis et al. 2022).

To further clean the Body Mass Index ("*mbmi*" variable), we calculated the upper bound using a simple percentile method. All observations above the 99th percentile were removed as outliers, but there were actually just a few of them.

This percentile-based approach, which defines upper and lower bounds (1st and 99th percentiles) and removes data points outside these bounds as outliers, was applied several times:

- Household earnings: a few households earning hundreds of thousands of hryvnias per year were excluded.
- Years of schooling: a few respondents with only preschool education or doctoral degrees were removed.
- Living conditions: a few households with 15 to 17 people living under one roof were excluded.

Additionally, whenever observations were missing covariate information (e.g., "Refused to answer" or "Don't know") and it was impossible to handle missing data through imputation (e.g., missing income and health indicators replaced with medians) – then such rows were removed from the dataset. This ensures that the dataset only includes valid and interpretable responses.

3.3 Dependent variables

The dependent variable for the study is *"isobes",* which equals 1 if obese and 0 otherwise. It was determined based on the *"mbmi"* variable (ratio of weight/height^2) and set to 1 if BMI ≥ 30 kg/m^2, following WHO recommendations (World Health Organization 2022).

BMI as a measure of obesity has been criticized by many because it does not indicate the distribution of the adipose tissue in the abdominal area. For this purpose, alternative indicators such as waist circumference (WC), waist-to-hip ratio (WHR), and waist-to-height ratio (WHtR) may be more accurate.

Since the survey also reports height, hip, and waist circumference, we decided to calculate all three alternative measures:

- Waist circumference (WC);
- Waist-to-hip ratio (WHR);
- Waist-to-height ratio (WHtR).

It is widely accepted that WHtR is the best one for predicting central obesity – it uses two measures and because of that it has greater predictive power than waist circumference used alone in screening tests: values ≥ 0.5 indicate an increased risk of cardiovascular disease, diabetes, as well as obesity (Traczyk et al 2023).

We analyzed the correlation between BMI and WHtR and plotted a scatterplot. The diagram indicates a strong positive relationship: higher WHtR aligns with higher BMI.

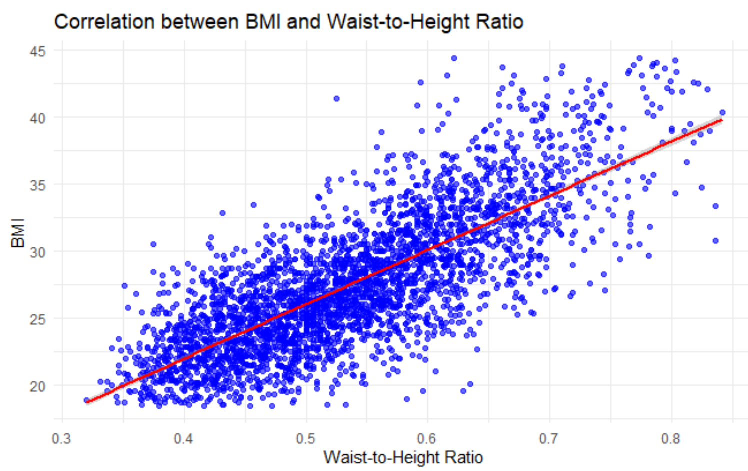

Figure 4. Correlation between BMI and Waist-to-Height Ratio

Source: author's calculations, World Health Organization 2020

Since we do not lose any accuracy, we've decided to stick with BMI – it is more standard and allows for comparison with other similar studies, placing our findings within a broader research context and yielding more valuable insights.

3.4 Independent variables

Independent measures are grouped into several categories depending on their nature. The idea is to use a stepwise approach and introduce different categories into a regression one by one.

Demographic Information

The simplest variables – yet, as we will see later, the most decisive ones – are *female* and *age*. Female is a binary indicator (1 for women, 0 for men). Age is measured continuously in years.

The same applies to years of schooling (*yearschool*). It is worth mentioning that STEPS also provides data on the highest level of education completed by the respondent, making it possible to create a set of dummy variables like "secondary school", "high school", etc. However, in regression analysis (e.g. Mincerian wage functions) the opposite approach is often taken, and the highest level of education is being converted to a continuous variable. For example, "secondary school" corresponds to 11 years of schooling (Gorodnichenko 2005). Therefore, we dropped the highest level of education and retained just years of schooling.

For conciseness, six options available for marital status (*maritstat*) were collapsed into four: "unmarried", "married" (which also includes "living together"), "divorced" (which also includes "separated"), and "widowed". However, in regression models we will likely use a simpler *"livalone"* variable: 1 for living alone, 0 for living with one or more persons. It should capture the same effect that marital status does.

STEPS provides a detailed breakdown of work statuses:

1. Employee of a governmental organization/enterprise
2. Employee of a non-governmental organization/enterprise
3. Employee of a private enterprise
4. Self-employed/Private Entrepreneur
5. Non-paid
6. Student
7. Homemaker
8. Retired

9. Unemployed (able to work)
10. Unemployed (unable to work)

However, we do not require this level of granularity – we are predominantly interested in whether a given individual belongs to the economically active population (labor force) or to the economically inactive population (outside the labor force) during a reference period. That is why in line with ILO recommendations we classify statuses 1 through 4 as economically active, while the rest are considered economically inactive. We introduce binary indicator *econactive*: 1 if economically active, 0 otherwise. This approach is not new. Stos et al. in their study also classified occupational status as either active (currently employed) or passive (currently unemployed) (Stos et al. 2022).

Another important characteristic is household earnings (*hhearn*). Respondents were asked to report their income either as per week, or per month, or per year. That is why on the very 1st step we normalized all time periods to a monthly format (weekly income was multiplied by 4, and annual income was divided by 12). After that – on the 2nd step – the 1st and 99th percentiles were calculated and observations outside this range were removed as outliers.

Expectedly enough respondents were reluctant to disclose their income. However, removing all the rows with missing values was not an option as it would narrow down our sample significantly and weaken predictive power. That is why on the 3rd step the median was calculated and populated to all empty cells.

Unlike some other surveys, STEPS asks to report household earnings rather than individual earnings, which is an advantage and economically more accurate since individuals live in households, not individually in isolation.

The last but not least in this subset is the "*urban*" variable denoting the place of residence: 1 if urban, 0 if rural.

Behavior Measurements

Smoke is a binary indicator: 1 for smokers, 0 for non-smokers. STEPS reports not only smoked tobacco use but also the prevalence of tobacco heating systems (e.g. iQOS), electronic cigarettes, and even snuff and chewing tobacco. However, for the purposes of this study we focus only on smoked tobacco (i.e. cigarettes, cigars).

The *alcconsump* variable represents alcohol consumption. It is measured in standard alcoholic drinks on a monthly basis using mean intervals. For example, if the participant responded drinking 5-6 days per week, then the value of *alcconsump* would be 22, calculated as follows: (5 episodes * 4 weeks + 6 episodes * 4 weeks)/2.

The concept of a standard alcoholic drink is a rather interesting one and worth briefly discussing. According to the STEPS survey, a standard drink is any drink that contains about 10 g of pure alcohol. It has been determined that approximately 30 ml of spirits, 120 ml of wine or 285 ml of beer contain this amount of alcohol (World Health Organization 2020).

Something similar in the spirit of standard alcoholic drinks is the notion of the standard serving of fruits or vegetables. Each serving is approximately 80 grams. The WHO recommends a combined intake of at least 5 servings per day, making the daily amount of about 400 grams total of fruits and vegetables. So, we took fruits servings, vegetables servings, summed them, and if the resulting number equals 5 or more, then *eatfruitveg* (eating fruits and vegetables) was assigned a value of 1, meaning that the consumption was sufficient, and 0 otherwise.

Another variable responsible for diet is *eatprocfood* (eating processed food). It is categorized into "always", "often", "sometimes", "rarely", or "never".

The final variable in the set of behavior measurements is *physact* (physical activity). WHO recommendations are not demanding, making them relatively easy to meet. If an adult does at least 10 minutes of vigorous-intensity physical activity continuously per day (whether at work, during travel time or leisure time), he is classified as having a “high” level of *physact*. If we still talk about 10 minutes but of moderate, not vigorous-intensity activity, it corresponds to “moderate” level. If there is no activity at all, it is a “low” level.

On a positive note, levels of physical activity in Ukraine – among the highest in the WHO European Region. Only 10% of the population did not meet the WHO recommendation (WHO 2024).

Physical Measurements and Biochemical Indicators

The WHO classifies STEPS participants based on their blood pressure readings in the following categories:

- normal if their SBP (Systolic Blood Pressure) and DBP (Diastolic Blood Pressure) readings were < 140 mmHg and < 90 mmHg, respectively;
- raised if their SBP was ≥ 140 mmHg and/or their DBP was ≥ 90 mmHg.

To reflect this, we introduced the *hypertension* binary indicator which equals 1 for raised BP, 0 otherwise.

Continuous readings for blood glucose and cholesterol were converted into categorical variables using standard thresholds from Table 1.

Table 1. Biochemical blood indicator cut-off points

Biochemical indicator	**Normal**	**At risk**	**Increased**
Plasma glucose	< 6.1 mmol/L	≥ 6.1 mmol/L and < 7.0 mmol/L	≥ 7.0 mmol/L or using glucose-lowering drugs
Total cholesterol	< 5.0 mmol/L	≥ 5.0 mmol/L to < 6.1 mmol/L	≥ 6.2 mmol/L or using cholesterol-lowering drugs

Source: World Health Organization 2020

Hence, the *gluc* and *chol* variables classify glucose and cholesterol levels into "Normal", "Increased", or "At risk" based on standard cutoffs.

It is worth noting that a significant number of participants refused blood tests, with some even declining a hypertension checkup. To preserve as many observations as possible, missing values were replaced with the median.

Mental Health

Yet another advantage of the STEPS survey is detailed tracking of depression symptoms. Our *mhealth* (mental health) variable is categorized into "good", "moderate", or "poor" based on three STEPS mental health indicators:

- "For the past 12 months have you been feeling sad, devastated or depressed for a few days in a row?"
- "For the past 12 months have you experienced a period during a few days when you had lost interest in the majority of things that bring pleasure to you (e.g., relations, job or hobby/rest)?"
- "For the past 12 months have you experienced a period during a few days when you felt a decrease in energy or permanent fatigue?"

Whenever all three questions receive a response "Yes", *mhealth* is poor. If three No's, *mhealth* is good. If there is a mix, *mhealth* is moderate. This algorithm is consistent with the one used in WHO's Study on Global Ageing and Adult Health (SAGE).

3.5 Sample characteristics

Table 2 presented below describes the characteristics of the sample, exploring all the dependent and independent variables one by one while focusing on urban-rural differences and potential insights.

We would like to stop on the top rows describing obesity (BMI ≥ 30 kg/m²) and combined obesity and overweight statuses (BMI ≥ 25 kg/m²).

Although urban areas have a higher absolute number of cases – both for obesity (568 vs. 364) and obesity and overweight (1414 vs. 776) – this is primarily due to the larger urban sample size. At the same time, if we compare obesity rates, then those living in rural environments are worse off – 364/1144 or roughly 32% is rural obesity and 568/2210 or 26% is urban obesity.

This goes in line with the WHO findings (World Health Organization 2020): *"The mean BMI of an adult was 26.8 kg/m² and increasing with age. Only two fifths (39.6%) of the population in Ukraine had normal weight (BMI 18.5–24.9 kg/m²). Almost three fifths (59.1%) were overweight (BMI ≥ 25 kg/m²), including a quarter of the population (24.8%) who were obese (BMI ≥ 30 kg/m²). Both overweight and obesity increased sharply with age, and obesity was more prevalent among women (men: 20.1%; women: 29.8%)."*

Table 2. Characteristics of the sample by residence status

			Residence status			
Variable	**Total Sample n=3354**		**Rural n = 1144**		**Urban n = 2210**	
	n	**n (%)**	**n**	**n (%)**	**n**	**n (%)**
Obesity						
No	2422	72.2%	780	32.2%	1642	67.8%
Yes	932	27.8%	364	39.1%	568	60.9%
Obesity and overweight						
No	1164	34.7%	368	31.6%	796	68.4%
Yes	2190	65.3%	776	35.4%	1414	64.6%
Gender						
Male	1277	38.1%	410	32.1%	867	67.9%
Female	2077	61.9%	734	35.3%	1343	64.7%

Variable	Total Sample n=3354		Residence status: Rural n = 1144		Residence status: Urban n = 2210	
	n	n (%)	n	n (%)	n	n (%)
Age (years)						
18-29	405	12.1%	113	27.9%	292	72.1%
30-44	865	25.8%	279	32.3%	586	67.7%
45-59	1095	32.6%	408	37.3%	687	62.7%
60-69	989	29.5%	344	34.8%	645	65.2%
Years of schooling (years)						
8	71	2.1%	42	59.2%	29	40.8%
9	104	3.1%	47	45.2%	57	54.8%
10	347	10.3%	118	34.0%	229	66.0%
11	424	12.6%	173	40.8%	251	59.2%
12	571	17.0%	225	39.4%	346	60.6%
13	618	18.4%	216	35.0%	402	65.0%
14	303	9.0%	91	30.0%	212	70.0%
15	543	16.2%	144	26.5%	399	73.5%
16	255	7.6%	57	22.4%	198	77.6%
17	74	2.2%	19	25.7%	55	74.3%
18	44	1.3%	12	27.3%	32	72.7%
Marital status						
Divorced/separated	638	19.0%	236	37.0%	402	63.0%
Married/living together	1799	53.6%	601	33.4%	1198	66.6%
Unmarried (never married)	391	11.7%	101	25.8%	290	74.2%
Widowed	526	15.7%	206	39.2%	320	60.8%
Occupational status						
Economically inactive	1593	47.5%	643	40.4%	950	59.6%
Economically active	1761	52.5%	501	28.4%	1260	71.6%
Living conditions						
Living with one or more persons	2442	72.8%	819	33.5%	1623	66.5%
Living alone	912	27.2%	325	35.6%	587	64.4%

Variable	Total Sample n=3354		Residence status Rural n = 1144		Urban n = 2210	
	n	n (%)	n	n (%)	n	n (%)
Household earnings						
Less than 3 000 UAH	352	10.5%	166	47.2%	186	52.8%
3 001 - 4 500 UAH	334	10.0%	152	45.5%	182	54.5%
4 501 - 6 000 UAH	1758	52.4%	590	33.6%	1168	66.4%
6 001 - 8 000 UAH	273	8.1%	94	34.4%	179	65.6%
8 001 - 10 000 UAH	187	5.6%	62	33.2%	125	66.8%
10 001 - 15 000 UAH	246	7.3%	54	22.0%	192	78.0%
15 001 - 20 000 UAH	100	3.0%	16	16.0%	84	84.0%
20 001 - 25 000 UAH	56	1.7%	5	8.9%	51	91.1%
More than 25 000 UAH	48	1.4%	5	10.4%	43	89.6%
Tobacco use						
No	2447	73.0%	864	35.3%	1583	64.7%
Yes	907	27.0%	280	30.9%	627	69.1%
Alcohol consumption (standard drink(s) per month)						
0	760	22.7%	262	34.5%	498	65.5%
1	1270	37.9%	446	35.1%	824	64.9%
2	686	20.5%	217	31.6%	469	68.4%
6	395	11.8%	122	30.9%	273	69.1%
14	161	4.8%	62	38.5%	99	61.5%
22	38	1.1%	15	39.5%	23	60.5%
30	44	1.3%	20	45.5%	24	54.5%
Sufficient fruits and vegetables consumption						
No	1973	58.8%	658	33.4%	1315	66.6%
Yes	1381	41.2%	486	35.2%	895	64.8%
Processed food consumption						
Always	103	3.1%	37	35.9%	66	64.1%
Never	453	13.5%	138	30.5%	315	69.5%
Often	650	19.4%	230	35.4%	420	64.6%

Variable	Total Sample n=3354		Residence status: Rural n = 1144		Residence status: Urban n = 2210	
	n	n (%)	n	n (%)	n	n (%)
Rarely	1232	36.7%	450	36.5%	782	63.5%
Sometimes	916	27.3%	289	31.6%	627	68.4%
Physical activity						
High	3010	89.7%	1051	34.9%	1959	65.1%
Moderate	126	3.8%	39	31.0%	87	69.0%
Low	218	6.5%	54	24.8%	164	75.2%
Hypertension						
No	2105	62.8%	654	31.1%	1451	68.9%
Yes	1249	37.2%	490	39.2%	759	60.8%
Blood glucose						
Normal	2920	87.1%	985	33.7%	1935	66.3%
Increased	217	6.5%	76	35.0%	141	65.0%
At risk	217	6.5%	83	38.2%	134	61.8%
Total cholesterol						
Normal	2288	68.2%	763	33.3%	1525	66.7%
Increased	231	6.9%	74	32.0%	157	68.0%
At risk	835	24.9%	307	36.8%	528	63.2%
Mental health						
Good	2144	63.9%	733	34.2%	1411	65.8%
Moderate	843	25.1%	292	34.6%	551	65.4%
Poor	367	10.9%	119	32.4%	248	67.6%

Source: author's calculations, World Health Organization 2020

The table displays all the variables of interest – even though not all of them were used in the regression analysis, they were still considered when drafting the conclusions and policy implications sections.

C h a p t e r 4

METHODOLOGY

Regression specification

Three sets of independent variables defined in the previous section will be added incrementally to the model to isolate the contribution of each factor. The starting point is the model with the *base set* of independent variables (1), which is then expanded to Set 1 with the inclusion of *health behaviors* (2). Finally, Set 1 is further expanded to Set 2 with the addition of the *mental health* variable to the model (3):

$$\begin{gathered} isobes = \beta^0 + \beta^1 female + \beta^2 age + \beta^3 yearschool + \beta^4 econactive \\ + \beta^5 livalone + \beta^6 ln(hhearn) + \beta^7 urban + \ \varepsilon \end{gathered} \tag{1}$$

$$\begin{gathered} isobes = \beta_0 + \beta_1 female + \beta_2 age + \beta_3 yearschool + \beta_4 econactive + \\ \beta_5 livalone + \beta_6 ln(hhearn) + \beta_7 urban + \beta_8 smoke + \beta_9 ln(alcconsump) + \\ \beta_{10} eatfruitveg + \beta_{11\mathrm{n}} \sum_{n=1}^{5} eatprocfood + \varepsilon \end{gathered} \tag{2}$$

$$\begin{gathered} isobes = \beta_0 + \beta_1 female + \beta_2 age + \beta_3 yearschool + \beta_4 econactive + \\ \beta_5 livalone + \beta_6 ln(hhearn) + \beta_7 urban + \beta_8 smoke + \beta_9 ln(alcconsump) + \\ \beta_{10} eatfruitveg + \beta_{11\mathrm{n}} \sum_{n=1}^{5} eatprocfood + +\beta_{12\mathrm{m}} \sum_{m=1}^{3} mhealth + \varepsilon \end{gathered} \tag{3}$$

The choice of these estimation models was inspired by prior studies, including:

- Zatońska et al.: Utilized predictors such as Age, Gender, Place of residence, Educational level, and Marital status (Zatońska et al. 2021).

- Stoś et al.: Included the same predictors as Zatońska but also added Occupational status (active/passive), Financial situation (good/moderate/bad), Physical activity level (low/moderate/high), and dummies for Having children under 18 years of age, Living alone, and Presence of at least one chronic disease (Stoś et al. 2022).
- Wen et al.: Did not include marital status (compared to Zatońska) but worked with Financial situation and Physical activity level (similar to Stoś). Additionally, Wen et al. used ethnicity-related dummies (e.g., Non-Hispanic white, Non-Hispanic black, Hispanic, US-born), which are likely less relevant in the context of Ukraine (Wen et al. 2018).

Each of the three aforementioned models will be estimated using both the Linear Probability Model (LPM) and the Logit Model, which was also utilized in the studies by Zatońska, Stoś, and Wen.

Our goal is to estimate the probability of being obese using both linear (LPM) and non-linear models (Logit) to check if they tell a consistent story, e.g. the signs of the coefficients are the same, the same variables are statistically significant in each model, etc.

We see that Set 1 and Set 2 introduce categorical variables *eatprocfood* (eating processed food) and *mhealth* correspondingly. They'll be automatically converted into dummy variables where let's say in the case of *eatprocfood* "always" serves as a reference category, and the coefficients for "often", "sometimes", "rarely", "never" will represent the effect of those categories compared to "always".

Another detail, *maritstat* (marital status) is not used in the regression as we believe it correlates with the *livalone* (living alone) variable. Also, we did not incorporate *physact* into the model as almost the entire sample reports being physically active – this could be due to WHO's low activity thresholds and the fact that *physact* is self-reported, making it prone to overestimation.

Hypertension, blood glucose, and total cholesterol are also excluded from the regression due to the potential for reverse causality – chances are that let's say glucose affects obesity the same way obesity affects glucose, so it is hard to distinguish which depends on which. Nonetheless, we will retain these variables in the thesis as they may provide valuable insights in further analyses.

Another crucial detail regarding specification is that household income (*hhearn*) and alcohol consumption (*alcconsump*) were log-transformed – a standard practice to correct for skewness. On the later stages, it will help to avoid marginal effect inflation and make interpretation more meaningful.

To conclude the discussion about specification it is worth mentioning that some other functional forms were evaluated which included not just logarithms, but also variables in levels and squares (to possibly capture a diminishing effect) or interaction terms (allowing binary qualitative variables to interact with continuous quantitative variables) (Wooldridge 2019). But they did not bring that much value, just introduced another level of complexity.

Multicollinearity check

After defining these three sets of independent variables, Variance Inflation Factors (VIFs) were calculated for each model. VIFs are used to assess multicollinearity – i.e. whether all predictors are sufficiently independent to be included in the regression analysis. Table 3 presents the multicollinearity check results, showing that across all specifications, the VIF values range from approx. 1.03 to 1.45, far below the commonly accepted threshold of 5. The results confirm that all models are free from multicollinearity, and it won't distort regression estimates.

Table 3. Multicollinearity check results

Variable	Base Model VIF	Set 1 GVIF	Set 2 GVIF
Female	1.030992	1.253639	1.276325
Age	1.300308	1.368229	1.368661
Years of Schooling	1.168218	1.184647	1.186011
Economically Active	1.322291	1.341583	1.345231
Lives Alone	1.186344	1.190268	1.193839
Household Earnings	1.430283	1.446435	1.454429
Urban	1.044043	1.052814	1.053927
Smokes		1.354754	1.356672
Alcohol Consumption		1.283467	1.289345
Eating Fruits and Vegetables		1.046211	1.053793
Eating Processed Food		1.233507	1.241908
Mental Health			1.051691

Source: author's calculations, World Health Organization 2020

Robust standard errors

Robust standard errors were computed with the help of a commonly used HC1 estimator. Why do we use robust SEs? To correct for potential non-constant variance (heteroskedasticity) in the residuals which could lead to biased standard errors, incorrect p-values and confidence intervals. As a result, we might think a variable is statistically significant when it actually isn't, and vice versa. Robust SEs were applied to all OLS and Logit models.

Marginal effects

In LPM, the coefficients are directly interpretable as marginal effects. However, in non-linear models like Logit, average marginal effects (AMEs) must be computed separately using *margins()* function. AMEs are average partial derivatives of the predicted probability with respect to each predictor. They show how much the predicted probability of obesity changes on average, given a one-unit change in a predictor (or a category change for factors). AMEs are easier to interpret than raw logistic coefficients (which are in log-odds). For example, suppose we calculate the average marginal effects for urban (urban residency) and obtain – 0.07. This means that living in an urban area reduces the probability of obesity by about 7 percentage points, holding other factors constant. This, in turn, implies rural areas might have higher obesity rates.

Goodness-of-Fit measures

For LPM models, adjusted R-squared was readily available. For Logit models, McFadden's pseudo R-squared was computed separately. It is analogous to R-squared and evaluates how much better your model fits the data relative to a null model (intercept-only model with no predictors which serves a baseline for comparison). McFadden's interpretation is also straightforward – if let's say pseudo R-squared of our base Logit model equals 0.065, it means that about 6.5% of the variation in obesity status is explained by the independent variables. This is a relatively low explanatory power. At the same time, we must note that it is common that binary dependent variable models (like Logit) have lower R-squared values than continuous dependent variable models. Additionally, such a complex health outcome as obesity expectedly has lower R-squared values due to the multitude of unobserved factors (e.g., genetic predispositions, environmental factors, etc.).

Chapter 5

ESTIMATION RESULTS

In Table 4, regression results are presented for each model specification (base, extended with health behaviors, and extended with mental health status). To save space only significant coefficients are kept, but whole output is available from appendices. Default standard errors in parenthesis were replaced with robust ones. For the Logit models, average marginal effects (AMEs) are computed and displayed, which is perfect for side-by-side comparisons.

Table 4. Estimation results (significant coefficients)

	LPM Base	**Logit Base (AMEs)**	**LPM Set1**	**Logit Set1 (AMEs)**	**LPM Set2**	**Logit Set2 (AMEs)**
Female	0.105*** (0.015)	0.1106*** (0.0159)	0.080*** (0.017)	0.0845*** (0.0178)	0.078*** (0.017)	0.0825*** (0.0178)
Age	0.007*** (0.001)	0.0078*** (0.0006)	0.007*** (0.001)	0.0073*** (0.0007)	0.007*** (0.001)	0.0073*** (0.0007)
Economically Active	0.026 (0.017)	0.0396** (0.0177)	0.034** (0.017)	0.0460** (0.0176)	0.035** (0.017)	0.0475** (0.0178)
Urban	-0.048*** (0.016)	-0.0505** (0.0156)	-0.050*** (0.017)	-0.0515** (0.0156)	-0.053*** (0.017)	-0.0518** (0.0156)
Smokes			-0.071** (0.019)	-0.0749*** (0.0209)	-0.071** (0.019)	-0.0745*** (0.0209)
Eating Processed Food – Always			reference	reference	reference	reference

	LPM Base	Logit Base (AMEs)	LPM Set1	Logit Set1 (AMEs)	LPM Set2	Logit Set2 (AMEs)
Eating Processed Food – Rarely			-0.073** (0.045)	-0.0773 (0.0503)	-0.072* (0.044)	-0.0773 (0.0503)
Observations	3354	3354	3354	3354	3354	3354
R^2	0.075		0.083		0.083	
Adjusted R^2	0.073		0.079		0.079	
Residual Std. Error	0.431		0.43		0.43	
F Statistic	38.828***		21.514***		18.916***	
Pseudo R^2		0.0687		0.0751		0.0754

Source: author's calculations, World Health Organization 2020

Talking about the overall explanatory power, R square for LPM and McFadden's R square for Logit models are modest, but typical for cross-sectional analysis of health behavior. The base model explains 6.9% (Logit) to 7.3% (LPM) of the variance in obesity outcomes. The inclusion of health behavior variables in Set 1 led to a small increase – 7.5 percentage points for Logit and 7.9 for LPM. The addition of mental health status in Set 2 did not contribute anyhow to the model's accuracy.

Along with the modest R square, the F statistic is high and jointly significant across all LPM models suggesting that there is a strong relationship between the response and the set of predictors. It ranges from almost 19 to almost 40 and signifies that our effort is meaningful and we're not just fitting noise.

What's important though is that both linear and non-liner models tell the consistent story – in terms of signs, significance etc. It means that there are reliable patterns in the data. Behavioral and demographic characteristics like urban residence, gender, age, and smoking behavior etc. appear to be strong correlates of obesity in the Ukrainian context.

Let's now walk through the key significant predictors focusing on how their effect evolves across all models.

Urban

Urban residency is our key variable of interest. In all models, the results show a clear, consistent, and statistically significant negative gradient – urban residency is associated with a lower probability of obesity. Basically, that's what we've been talking about in the hypothesis.

In the Base LPM, the coefficient is approximately – 0.048, indicating that individuals in urban areas are 4.8 percentage points less likely to be obese compared to rural residents, holding all else constant. After adjusting for health behaviors (Set 1), and then for mental health (Set 2), the effect is even stronger – 5.0% and 5.3% respectively. The average marginal effects (AMEs) in the Logit models confirm this pattern, though they are slightly less significant ($p<0.05$).

We'll try to decompose in the concluding sections why urban dwellers are more protected – among obvious reasons, they may have better access to health information, recreational facilities, and healthier food options.

Female

Women have a significantly higher probability of being obese compared to men – in LPM Base, 10.5 percentage points higher, in Logit Base even more – 11.06 pp higher, all else equal. As health behaviors and mental health are added in Set 1 and Set 2, the effect remains significant but slightly weaker – it hovers around 7.8% to 8.5%. But still being female shows very strong effect, stronger than any other statistically significant predictor of obesity (urban, age, economically active, smoke). The persistent gender gap may be provoked by post-partum weight retention, physical activity patterns, and by many other factors which are worth separate discussion. All in all, these results also mirror finding from several European studies (e.g., Diamantis et al. 2022, Stival et al. 2022) and national research in Poland (Stos et al. 2022).

Age

Age is strongly significant but with a small effect. Specifically, a 1-year increase in age increases the probability of being obese by approximately 0.7 percentage points. This is consistent across all models and reflects typical aging effect – slower metabolism, lower physical activity, cumulative lifestyle risks and others. Also, age output is in line with numerous prior studies (Traczyk et al. 2023, Wang et al. 2021).

Economically active

In LPM Base, economically active was not statistically significant, but after the inclusion of health behavior variables and in the Logit AMEs the marginal effects are always statistically significant at 5% level. Economically active individuals roughly have a 4 percentage points higher probability of obesity compared to non-active ones. What's interesting is that this result runs counter to some Western literature that identifies an inverse relationship between income and obesity, but at the same time it aligns with evidence from transitional economies where rising affluence is associated with dietary shifts toward processed, calorie-dense food (Okunogbe et al. 2022). Indeed, if let's say in the case of age the interpretation is simple, here it's more challenging to figure out what's going on. Another theory – along with eating processed food – is that being employed might correlate with more sedentary lifestyles, stress-related eating and less-time for exercise.

Smoke

Smoking is a negative and statistically significant predictor of obesity in all models where it is included – Set 1 and Set 2. In Logit AMEs, smoking is associated with 7.5 percentage points reduction in obesity probability. In LPM, the effect is similar, around 7%, albeit with slightly lower significance. This is likely due to appetite-suppressing effects of nicotine or substitution of food intake with smoking, though the health consequences of smoking far outweigh any apparent benefits in weight reduction.

Eating processed food

The variable "eating processed food" is a significant negative predictor, but only in the LPM Set 1 model ($p<0.05$) and to some extent in LPM Set 2 ($p<0.1$). The coefficient everywhere is about 0.07, implying that respondents who consume processed food rarely are 7 percentage points less likely to be obese compared to those who consume it always. This finding aligns with extensive global evidence linking ultra-processed foods to obesity. Also, it echoes to the abovementioned justification why economically active individuals are more likely to be obese – due to dietary shifts toward processed food in transitional economies (Okunogbe et al. 2022).

Non-significant predictors

"Living alone", "years of schooling", and "household earnings" did not show a statistically significant effect on obesity in the base specification and continued to lack significance across other model specifications as well. Even though smoking status is found to be significant, as well as sometimes eating process food, other behavioral variables – alcohol consumption and fruit/vegetable intake – do not reach statistical significance. Despite theoretical importance, mental health also does not exhibit statistically significant associations in the final models. What's probably important though is that the direction of the effects (especially for healthy eating) was still intuitive.

Chapter 6

CONCLUSIONS AND POLICY RECOMMENDATIONS

The hypothesis of this paper was that rural dwellers are more obese in Ukraine. Indeed, regression results indicate that rural residency is positively and significantly associated with obesity, even after controlling for income, education, age, and behaviors such as smoking, alcohol and processed food consumption. The conclusion goes in line with key patterns observed in both national and European literature (Stos et al. 2022, Zatonska et al. 2021). Now, the question is what can be done about it from a policy perspective?

Awareness and education campaigns

Efficient treatment of obesity should start with the acknowledgement that it is indeed a problem. So, *awareness and education campaigns* should be conducted and finetuned for the rural specifics. We can think about this adjustment from different perspectives. First, who delivers the message? National public health campaigns may not reach or resonate with rural populations. For tailored public health messaging it is important to include local authorities as well as trusted community leaders (teachers, doctors, religious leaders). Another perspective is the channel of communication. Internet campaigns may be less effective in areas with digital gaps. Public events and community workshops usually draw more attention in rural settings so they can be employed to promote nutritional literacy.

Preventive care

The best treatment for the disease is its *prevention*. In that matter, it is important to pay enough attention to rural schools – teachers should deliver the right message. They should assist in forming healthy habits. School food standards should be corresponding with an accent on healthy food whenever possible.

Family doctors in rural areas also shouldn't neglect the problem. They should be trained to track obesity issues and integrate nutrition counselling into routine visits. In urban setting there much more possibilities – nowadays even separate personalized nutrition industry is emerging (like the ZOE app), which allows to read body characteristics and provides tailored body advice. It is doable in urban context because of health apps familiarity, internet access and higher smartphone penetration. But in rural setting at this point only family doctors can cover that role.

Also let's not forget about built environment and infrastructure. Local authorities should create recreation facilities like playgrounds, bike lanes, walking paths, community centers offering physical sessions. One can argue that rural dwellers anyway work hard so what's the point of investing in the physical activity infrastructure – still active work often has seasonal character, and even if not, still in essence it is different from moderate well-designed recreational physical activities in the gym.

Access to treatment

Suppose that people are well-informed on and did their best to prevent the problem, still it is there – then rural dwellers should have access to the treatment. That is a huge inequality compared to urban setting. Even in the STEPS survey context, fieldwork interviewers emphasized that cholesterol and blood glucose tests were rare prior to the interview for many respondents (WHO STEPS 2019).

If stationary treatment is not an option, we can think about mobile clinics which can deliver care to hard-to-reach communities. Another option is telehealth – it can resolve the issue with outreach to remote areas but still some specialized centers with corresponding technology should be available.

Above all, rural obesity should not be treated as something simple to fight. The problem as suggest regression results is complex. Not just the type of residency turned out to be significant but also other factors which means that obesity is a multifaceted phenomenon and the first step in fighting it should be acknowledgement of its complexity. Specialized programs should receive enough funding and be conducted by professionals, ideally under the collaboration of central government, local councils, NGOs etc.

This discussion section is subjective and some of the suggestions might not be valid under some specific circumstances. Additionally, we need to recognize that the research itself has a few limitations.

Research limitations

It is based on cross-sectional data from the STEPS survey, which captures information at a single point in time (2019). It means that we can talk about some associations between let's say obesity and rural residence but we're not able to insist on causality. For that reason, longitudinal data should better be used.

Another related limitation is the current external validity as the research was conducted prior to russian full-scale invasion in 2022. Now it is questionable whether we can generalize our finding to other time periods after 2022.

The key variable of interest – urban-rural classification – is oversimplified to some extent as it is binary. Probably more granular geographic classification might yield richer insights. Regarding other variables, there is a rich set of them included in the model, but still we cannot assure against omitted variable bias as important factors like genetics or cultural beliefs simply cannot be captured. Also, some variables are self-reported (like alcohol or diet for example) so there is a risk of underreporting or social desirability bias, especially around sensitive health behaviors.

Even though there are some risks associated, the topic of obesity remains an important area of research. As a next step of this study with a slightly different emphasis we can talk about childhood obesity. There is a WHO European Childhood Obesity Surveillance Initiative called COSI. Five rounds have already been conducted, and Ukraine is a part of 6th round (2023) of COSI (Breda et al. 2021). This is certainly something worth paying attention to in the future when results are published. However, this is the topic for a separate study.

WORKS CITED

Ajayi, I. O., C. Adebamowo, H. O. Adami, S. Dalal, M. B. Diamond, F. Bajunirwe, D. Guwatudde, M. Njelekela, J. Nankya-Mutyoba, F. S. Chiwanga, J. Volmink, R. Kalyesubula, C. Laurence, T. G. Reid, D. Dockery, D. Hemenway, D. Spiegelman, and M. D. Holmes. 2016. "Urban-Rural and Geographic Differences in Overweight and Obesity in Four Sub-Saharan African Adult Populations: A Multi-Country Cross-Sectional Study." *BMC Public Health* 16, no. 1 (October): 1126. https://doi.org/10.1186/s12889-016-3789-z.

Brinduse, L. A., I. Eclemea, A. E. Neculau, B. A. Paunescu, E. C. Bratu, and M. A. Cucu. 2024. "Rural Versus Urban Healthcare Through the Lens of Health Behaviors and Access to Primary Care: A Post-Hoc Analysis of the Romanian Health Evaluation Survey." *BMC Health Services Research* 24, no. 1 (November): 1341. https://doi.org/10.1186/s12913-024-11861-9.

Diamantis, D. V., K. Karatzi, P. Kantaras, S. Liatis, V. Iotova, Y. Bazdraska, T. Tankova, G. Cardon, K. Wikström, I. Rurik, E. Antal, A. M. Ayala-Marín, N. G. Legarre, K. Makrilakis, and Y. Manios. 2022. "Prevalence and Socioeconomic Correlates of Adult Obesity in Europe: The Feel4Diabetes Study." *International Journal of Environmental Research and Public Health* 19, no. 19 (October): 12572. https://doi.org/10.3390/ijerph191912572.

Dixon, B. N., E. Piervil, A. Eastman, and K. M. Ross. 2019. "A Social-Ecological Review of the Rural Versus Urban Obesity Disparity." *Health Behavior and Policy Review* 6, no. 4 (July): 378–394.

Gareth, James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning: With Applications in R*. 1st ed. New York: Springer Texts in Statistics.

Gorodnichenko, Yuriy, and Klara Sabirianova Peter. 2004. "Returns to Schooling in Russia and Ukraine: A Semiparametric Approach to Cross-Country Comparative Analysis." SSRN, September. https://doi.org/10.2139/ssrn.598229.

Neuman, M., I. Kawachi, S. Gortmaker, and S. V. Subramanian. 2013. "Urban-Rural Differences in BMI in Low- and Middle-Income Countries: The Role of Socioeconomic Status." *The American Journal of Clinical Nutrition* 97, no. 2 (February): 428–436. https://doi.org/10.3945/ajcn.112.045997.

NCD Risk Factor Collaboration (NCD-RisC). 2019. "Rising Rural Body-Mass Index is the Main Driver of the Global Obesity Epidemic in Adults." *Nature* 569 (May): 260–264. https://doi.org/10.1038/s41586-019-1171-x.

Okunogbe, A., R. Nugent, G. Spencer, J. Powis, J. Ralston, and J. Wilding. 2022. "Economic Impacts of Overweight and Obesity: Current and Future Estimates for 161 Countries." *BMJ Global Health* 7 (April): e009773. https://doi.org/10.1136/bmjgh-2022-009773.

Specchia, M. L., M. A. Veneziano, C. Cadeddu, A. M. Ferriero, A. Mancuso, C. Ianuale, P. Parente, S. Capri, and W. Ricciardi. 2015. "Economic Impact of Adult Obesity on Health Systems: A Systematic Review." *European Journal of Public Health* 25, no. 2 (April): 255–262. https://doi.org/10.1093/eurpub/cku170.

Stival, C., A. Lugo, A. Odone, P. A. van den Brandt, E. Fernandez, O. Tigova, J. B. Soriano, M. José López, S. Scaglioni, and S. Gallus. TackSHS Project Investigators. 2022. "Prevalence and Correlates of Overweight and Obesity in 12 European Countries in 2017–2018." *Obesity Facts* 15, no. 5: 655–665. https://doi.org/10.1159/000525792.

Stoś, K., E. Rychlik, A. Woźniak, M. Ołtarzewski, M. Jankowski, M. Gujski, and G. Juszczyk. 2022. "Prevalence and Sociodemographic Factors Associated with Overweight and Obesity Among Adults in Poland: A 2019/2020 Nationwide Cross-Sectional Survey." *International Journal of Environmental Research and Public Health* 19, no. 3 (January): 1502. https://doi.org/10.3390/ijerph19031502.

Thapa, R., C. Dahl, W. P. Aung, and E. Bjertness. 2021. "Urban-Rural Differences in Overweight and Obesity Among 25–64-Year-Old Myanmar Residents: A Cross-Sectional, Nationwide Survey." *BMJ Open* 11, no. 3 (March): e042561. https://doi.org/10.1136/bmjopen-2020-042561.

Traczyk, I., A. Kucharska, B. I. Sińska, M. Panczyk, L. Wronka, F. Raciborski, D. Szostak-Węgierek, and B. Samoliński. 2023. "Every Second Adult Inhabitant of Poland (Aged 18–64) is Overweight—Results of Representative Cross-Sectional Studies Conducted in 2017–2020." *Annals of Agricultural and Environmental Medicine* 30, no. 2 (June): 322–330. https://doi.org/10.26444/aaem/165913.

Vandevijvere, S., R. De Pauw, S. Djojosoeparto, V. Gorasso, L. Guariguata, A. L. Løvhaug, M. Mialon, I. Van Dam, and P. von Philipsborn. 2023. "Upstream Determinants of Overweight and Obesity in Europe." *Current Obesity Reports* 12, no. 4 (December): 417–428. https://doi.org/10.1007/s13679-023-00524-1.

Wang, L., B. Zhou, Z. Zhao, L. Yang, M. Zhang, Y. Jiang, Y. Li, M. Zhou, L. Wang, Z. Huang, X. Zhang, L. Zhao, D. Yu, C. Li, M. Ezzati, Z. Chen, J. Wu, G. Ding, and X. Li. 2021. "Body-Mass Index and Obesity in Urban and Rural China: Findings From Consecutive Nationally Representative Surveys During

2004–18." *The Lancet* 398, no. 10294 (July): 53–63. https://doi.org/10.1016/S0140-6736(21)00798-4.

Wen, M., J. X. Fan, L. Kowaleski-Jones, and N. Wan. 2018. "Rural-Urban Disparities in Obesity Prevalence Among Working-Age Adults in the United States: Exploring the Mechanisms." *American Journal of Health Promotion* 32, no. 2 (February): 400–408. https://doi.org/10.1177/0890117116689488.

Wooldridge, Jeffrey M. 2019. *Introductory Econometrics: A Modern Approach*. 7th ed. Boston, MA: Cengage Learning.

World Obesity Federation. 2022. *The Economic Impact of Overweight & Obesity in 2020 and 2060. 2nd Edition with Estimates for 161 Countries*. https://data.worldobesity.org/publications/WOF-Economic-Impacts-2-V2.pdf.

World Obesity Federation. 2023. *World Obesity Atlas 2023*. https://www.worldobesityday.org/assets/downloads/World_Obesity_Atlas_2023_Report.pdf.

World Health Organization. 2022. *WHO European Regional Obesity Report 2022*. https://iris.who.int/bitstream/handle/10665/353747/9789289057738-eng.pdf.

World Health Organization. 2024. *Obesity and Overweight*. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight.

World Health Organization. Regional Office for Europe. 2020. *Risk Factors for Noncommunicable Diseases in Ukraine in 2019*. https://cdn.who.int/media/docs/librariesprovider2/country-sites/ukraine/risk-factors-ncd-ukraine-summary-eng.pdf.

Zatońska, K., P. Psikus, A. Basiak-Rasała, Z. Stępnicka, D. Gaweł-Dąbrowska, M. Wołyniec, J. Gibka, A. Szuba, and K. Połtyn-Zaradna. 2021. "Obesity and Chosen Non-Communicable Diseases in PURE Poland Cohort Study." *International Journal of Environmental Research and Public Health* 18, no. 5 (March): 2701. https://doi.org/10.3390/ijerph18052701.

Zweifel, Peter, Friedrich Breyer, and Mathias Kifmann. 2009. *Health Economics*. 2nd ed. Heidelberg: Springer-Verlag.

APPENDIX A

OVERVIEW OF THE KEY PAPERS MENTIONED IN THE LITERATURE REVIEW

Table 5. Summary table with an overview of the key papers mentioned in the literature review

	Study Title, Year, and Authors	**Data Source and Methodology**	**Key Findings and Relevance to Research**
1	Wen M., Fan J.X., Kowaleski-Jones L., Wan N. (2018). "Rural-Urban Disparities in Obesity Prevalence Among Working Age Adults in the United States"	Multilevel Logistic Regression Data Source: 2003–2008 NHANES (National Health and Nutrition Examination Survey)	Rural residents are 36% more likely to be obese than urban counterparts; the gap drops by 94% after controlling for education, income, and built environment features, highlighting socioeconomic status (SES) as the driver of rural-urban obesity disparities.
2	Dixon B.N., Piervil E., Eastman A., Ross K.M. (2019). "A Social-Ecological Review of the Rural versus Urban Obesity Disparity"	Conceptual and Policy Analysis	Rural-urban obesity is analyzed across individual, interpersonal, community, and societal levels. Rural communities consistently fare worse with poorer diets, less physical activity, and lower SES.
3	Wang L., Zhou B., Zhao Z., et al. (2021). "Body-mass index and obesity in urban and rural China."	Conceptual and Policy Analysis Data Source: Six consecutive health surveys conducted between 2004 and 2018	Urban obesity prevails in China, though it slowed significantly after 2010 policy interventions. Thus, the study demonstrates a reversal of the urban-rural obesity dynamics observed in the U.S.
4	Neuman M., Kawachi I., Gortmaker S., Subramanian S.V. (2013). "Urban-rural differences in BMI in low- and middle-income countries"	Linear and Ordered Multinomial Analysis Data Source: 1991–2010 data on 678,000 women across 38 countries	In LMICs, urban areas exhibit higher BMI, but SES weakens the association between BMI and urban residence.

	Study Title, Year, and Authors	Data Source and Methodology	Key Findings and Relevance to Research
5	Stival C., Lugo A., Odone A., et al. (2022). "Prevalence and Correlates of Overweight and Obesity in 12 European Countries in 2017-2018"	Multilevel Logistic Random Effects Analysis Data Source: TackSHS survey conducted in 2017–2018 across 12 European countries	53% of European adults are overweight, and 13% are obese. Eastern Europe shows the highest obesity rates, driven by low economic growth and SES barriers.
6	Diamantis D.V., Karatzi K., Kantaras P., et al. (2022). "Prevalence and Socioeconomic Correlates of Adult Obesity in Europe: The Feel4Diabetes Study"	Conceptual and Policy Analysis Data Source: Feel4Diabetes survey conducted in 6 European countries: Bulgaria, Hungary, Finland, Belgium, Greece, and Spain	The authors introduced SEBS (Socioeconomic Burden Score), correlating higher unemployment, lower education, and income insecurity with obesity. This association applies across all examined countries.
7	Vandevijvere S., De Pauw R., et al. (2023). "Upstream Determinants of Overweight and Obesity in Europe"	Conceptual and Policy Analysis	Vandevijvere et al. introduced the notion of upstream determinants (e.g., food availability and the built environment), which operate at a higher level compared to individual factors (e.g., genetics, diet, etc.). Upstream determinants are the most impactful for obesity reduction.
8	Zatońska K., Psikus P., et al. (2021). "Obesity and Chosen Non-Communicable Diseases in PURE Poland Cohort Study"	Logistic Regression Data Source: PURE (Prospective Urban and Rural Epidemiology)	Both urban and rural populations were carefully included, as one of the objectives of the study was to address health inequalities. Rural residence significantly increases obesity odds (OR = 1.79).

	Study Title, Year, and Authors	Data Source and Methodology	Key Findings and Relevance to Research
9	Traczyk I., Kucharska A., et al. (2023). "Every second adult inhabitant of Poland (aged 18-64) is overweight"	Logistic Regression Data Source: 2017–2020 cross-sectional surveys	The study assessed the prevalence of not only general but also abdominal obesity and overweight. Women were more prone to abdominal obesity compared to men.
10	Stoś K., Rychlik E., et al. (2022). "Prevalence and Sociodemographic Factors Associated with Overweight and Obesity in Poland"	Multivariate Logistic Regression Data Source: Nationwide survey (2019/2020)	Out of 11 analysed factors, only 5 were significant predictors of obesity: males, individuals with chronic diseases, occupationally active individuals, rural residents, and older participants.
11	WHO STEPS Survey (2020). "Risk factors for noncommunicable diseases in Ukraine in 2019"		Provides foundational obesity data for Ukraine and enables future trend tracking.
12	World Obesity Federation (2022). "Economic Impact of Overweight and Obesity in 2020 and 2060"		Projects significant economic losses from obesity for Ukraine by 2060.

APPENDIX B

ESTIMATION RESULTS

Table 6. Estimation results (whole output)

	LPM Base	Logit Base (AME)	LPM Set1	Logit Set1 (AME)	LPM Set2	Logit Set2 (AME)
Female	0.105*** (0.015)	0.1106*** (0.0159)	0.080*** (0.017)	0.0845*** (0.0178)	0.078*** (0.017)	0.0825*** (0.0178)
Age	0.007*** (0.001)	0.0078*** (0.0006)	0.007*** (0.001)	0.0073*** (0.0007)	0.007*** (0.001)	0.0073*** (0.0007)
Years of Schooling	-0.004 (0.004)	-0.0042 (0.0037)	-0.006 (0.004)	-0.0052 (0.0037)	-0.006 (0.004)	-0.0054 (0.0037)
Economically Active	0.026 (0.017)	0.0396** (0.0177)	0.034** (0.017)	0.0460** (0.0176)	0.035** (0.017)	0.0475** (0.0178)
Lives Alone	0.023 (0.019)	0.0150 (0.0178)	0.025 (0.019)	0.0172 (0.0177)	0.024 (0.019)	0.0185 (0.0198)
Household Earnings	0.004 (0.016)	0.0068 (0.0168)	0.010 (0.016)	0.0117 (0.0168)	0.011 (0.017)	0.0131 (0.0168)
Urban	-0.048*** (0.016)	-0.0505** (0.0156)	-0.050*** (0.017)	-0.0515** (0.0156)	-0.053*** (0.017)	-0.0518** (0.0156)
Smokes			-0.071** (0.019)	-0.0749*** (0.0209)	-0.071** (0.019)	-0.0745*** (0.0209)
Alcohol Consumption			-0.006 (0.010)	-0.0010 (0.0116)	-0.006 (0.010)	-0.0037 (0.0110)
Eating Fruits and Vegetables			-0.018 (0.015)	-0.0178 (0.0154)	-0.016 (0.016)	-0.0161 (0.0154)

	LPM Base	Logit Base (AME)	LPM Set1	Logit Set1 (AME)	LPM Set2	Logit Set2 (AME)
Eating Processed Food – Always			reference	reference	reference	reference
Eating Processed Food – Often			-0.056 (0.045)	-0.0654 (0.0516)	-0.054 (0.045)	-0.0646 (0.0515)
Eating Processed Food – Sometimes			-0.063 (0.045)	-0.0683 (0.0507)	-0.062 (0.043)	-0.0682 (0.0507)
Eating Processed Food – Rarely			-0.073** (0.045)	-0.0773 (0.0503)	-0.072* (0.044)	-0.0773 (0.0503)
Eating Processed Food – Never			-0.003 (0.045)	-0.0162 (0.0534)	-0.002 (0.044)	-0.0184 (0.0534)
Mental Health – Good					reference	reference
Mental Health – Moderate					0.021 (0.018)	0.0203 (0.0177)
Mental Health – Poor					0.012 (0.025)	0.0132 (0.0244)
Observations	3354	3354	3354	3354	3354	3354
R^2	0.075		0.083		0.083	
Adjusted R^2	0.073		0.079		0.079	
Residual Std. Error	0.431		0.43		0.43	
F Statistic	38.828***		21.514***		18.916***	
Pseudo R^2		0.0687		0.0751		0.0754

Source: author's calculations, World Health Organization 2020