UNIVERSITA' DI UNIGE
SCUOLA DI SCIENZE SOCIALI
DIPARTIMENTO DI ECONOMIA

Tesi di laurea magistrale in

Economics and Data

Science

Body Mass Index, Long COVID and Sugar Tax: An Empirical Analysis in the European Context

Relatore: Leporatti Lucia

Candidato: Sciutto Alessia

Anno accademico 2025/2026

Table of Contents

Abstract

This thesis investigates the relationship between body mass index (BMI) and the probability of developing Long COVID in a European context. The motivation stems from the need to understand how pre-existing metabolic conditions influence post-pandemic health outcomes, moving beyond the analysis of acute COVID-19.

The empirical analysis is based on longitudinal data from the Survey of Health, Ageing and Retirement in Europe (SHARE). A logistic regression model is used to estimate the association between BMI and the likelihood of persistent symptoms, controlling for demographic, clinical, and socioeconomic factors. The results show a positive and statistically significant relationship, indicating that higher BMI levels are associated with an increased probability of developing Long COVID.

In addition, the study considers the broader policy environment by examining the role of sugar tax measures as part of preventive strategies targeting obesity. While not directly causal, these policies are interpreted as structural tools that may influence population-level metabolic risk.

Overall, the findings highlight the importance of addressing obesity not only as a chronic condition, but also as a key factor in strengthening resilience to future health shock.

Chapter I: Introduction

1.1 Long Covid: A New Challenge for European Health Systems

The emergence of the SARS-CoV-2 virus on the global health scene in early 2020 represented an unprecedented disruptive event for contemporary societies (WHO, 2021). What initially appeared as a localized health crisis rapidly evolved into a systemic shock affecting not only public health but also the economic and institutional organization of European countries. Governments found themselves facing a rapid and unpredictable viral spread, adopting extraordinary containment measures, and reorganizing healthcare delivery and service provision in a very short time to respond to the hospital emergency.

The initial phase of the pandemic highlighted structural weaknesses in several European healthcare systems (OECD, 2024). In particular, the sudden pressure on hospital capacity and available resources revealed the limitations of organizational models geared primarily towards short-term efficiency. Healthcare systems proved unprepared to manage such a crisis. The shortage of essential medical equipment and the difficulty of coordinating between different levels of care highlighted how the resilience of the system was lower than expected under normal conditions.

As the pandemic progressed, it also became apparent that COVID-19 infection did not affect the population uniformly. The severity of the disease and clinical outcomes were strongly influenced by pre-existing individual characteristics such as age, gender, and general health (Wilk et al., 2024). Environmental characteristics were also found to be decisive in the development of the disease. At this stage, the scientific community's attention has focused primarily on reducing mortality and managing the acute phase of infection, thanks in part to the development and distribution of vaccines.

However, as acute mortality was gradually brought under control, it became clear that resolution of the infection did not always coincide with a full return to previous health. Many patients continued to report persistent symptoms even months after the initial infection (Nalbandian et al., 2021), despite testing negative in diagnostic tests. This discrepancy between virological cure and actual recovery marked the transition from the pandemic emergency to a new phase of the health crisis. Although severe acute infection increases the risk, long COVID can also develop after mild or asymptomatic cases. The diagnosis of long COVID is primarily clinical, unlike acute COVID-19, which is identified through molecular testing.

In this context, the concept of Long COVID, or Post-Acute Sequelae of SARS-CoV-2 (PASC), has been established to describe symptoms persisting at least three months after infection (WHO, 2021). Long COVID is not a single disease but a complex, multisystemic syndrome. The symptoms reported in the literature are numerous and variable, and include fatigue, respiratory difficulties, and cognitive impairment (Nalbandian et al., 2021).

From the perspective of healthcare systems, the spread of Long COVID introduces a significant change in healthcare demand (OECD, 2024). Emergency care is now accompanied by the need to monitor patients with persistent limitations over time. These can be related to both physical and cognitive functioning. These conditions are often difficult to diagnose using traditional diagnostic tools, but they have a significant impact on the quality of life of those affected.

Recent data provide an initial quantification of the extent of Long COVID in the European population. According to an international survey coordinated by the Organisation for Economic Co-operation and Development, approximately 7.2% of the population in primary care reported experiencing Long COVID symptoms. Approximately 5.1% reported persistent symptoms at the time of the survey (OECD, 2024).

Although these estimates vary depending on the definition of Long COVID adopted and the detection methodologies used, they offer an initial picture of the extent of the post-acute condition in the European region. These data highlight the significant prevalence of persistent symptoms after infection, but also the heterogeneity of experiences related to Long COVID. They therefore underscore the need for more standardized and consistent surveillance systems across countries.

For European healthcare systems, this development represents a new and largely unexplored challenge. The lack of established knowledge about the mechanisms of onset and duration of symptoms makes it difficult to plan medium- and long-term healthcare responses. In this uncertain scenario, it is essential to understand which individual factors are associated with a higher probability of developing post-acute sequelae (Wilk et al., 2024). Analyzing the determinants of Long COVID is therefore an essential step in guiding future health strategies.

This “low intensity but high-volume” pressure complicates planning processes and makes it difficult for institutions to clearly delimit the boundary between recovery and chronic illness. As a result, Long COVID emerges not only as a medical condition, but as a structural stress test for the organizational capacity of post-pandemic health systems.

1.2 Obesity as a Pre-pandemic Global Epidemic

While the world focused on the COVID-19 pandemic, another health crisis continued to grow silently: obesity. Already defined as 'globesity', this condition is no longer considered a problem linked exclusively to individual lifestyle choices. Today it is recognized as a chronic and progressive disease characterized by excessive accumulation of body fat that can compromise health (Blüher, 2019). It is no longer seen merely as an aesthetic issue, but as a complex clinical condition that interacts with many other chronic diseases and reduces the resilience of the population.

Obesity is commonly measured using the Body Mass Index (BMI): a simple but effective indicator that relates weight to height. BMI is calculated by dividing weight in kilograms by the square of height in meters (kg/m^2). The classification most used by the WHO distinguishes three main categories: overweight (BMI $\geq$ 25), obesity grade I (BMI 30–34.9), obesity grade II (BMI 35–39.9) and severe obesity grade III (BMI $\geq$ 40) (WHO, 2000).

BMI does not directly measure body fat but provides an index for the general population. This allows the risk associated with various metabolic conditions to be estimated. Clinical studies have shown that as BMI increases, the risk of developing diabetes, high blood pressure, cardiovascular disease and certain forms of cancer increases proportionally (Blüher, 2019).

BMI also allows us to distinguish between overweight and obesity, identifying individuals at risk and vulnerable populations. However, it is important to note that it does not fully capture the distribution of body fat. Visceral fat, located around the internal organs, plays a more critical role in determining metabolic risk than subcutaneous fat

(Blüher, 2019). For this reason, obesity is now considered a progressive disease: its impact on health depends not only on the amount of fat accumulated, but also on its distribution and the duration of exposure.

Obesity stems from genetic, behavioral, environmental, and socio-economic factors (Ng et al., 2014). For example, high-calorie diets, excessive sugar consumption, and appetite dysregulation contribute to the development of obesity. The environment is also one of the causes of obesity, along with a country's food policies: aggressive marketing or unequal access to healthy foods can be among the causes. These factors interact to create a systemic epidemic that cannot be solved by individual choices alone.

People who are obese have a higher risk of type 2 diabetes, high blood pressure, cardiovascular disease, and dyslipidemia. It can lead to certain types of cancer (colon, breast, endometrial), certain musculoskeletal disorders and a reduced immune response to infections (WHO, 2000).

Until the mid-20th century obesity was rare. In rich countries, it was associated with affluence; in poor countries, malnutrition was prevalent. In the 1970s, its prevalence began to rise rapidly. The globalization of food markets, the industrialization of food processes and urbanization have radically changed diets and lifestyles. Ultra-processed, high-calorie, low-cost foods have become widespread. Daily physical activity has decreased due to sedentary work and passive transport.

The prevalence of obesity in adults has grown from 5% in the 1970s to over 15% in 2014 (Ng et al., 2014). In the United States, it exceeds 30%, while in Europe it averages over 20%. Mediterranean countries such as Italy, Spain and Greece show slightly lower percentages, but these are growing rapidly. In low- and middle-income countries, obesity is growing alongside malnutrition, creating a 'double burden' of disease.

Obesity has been a predictor of hospitalization during the COVID-19 pandemic, and severe complications (Popkin et al., 2020). Adipose tissue expresses ACE2 receptors, which can facilitate virus entry and increase its persistence in tissues. This contributes to Long COVID. Obesity acts as a risk multiplier, not just as a comorbidity.

Before the COVID-19 pandemic, obesity was already a silent health crisis, characterized by high prevalence, complex causes, significant clinical risks, and profound social impacts. Its definition as a progressive disease allows us to understand its progression over time and the risk of complications. Understanding the history, causes and spread of obesity is essential for developing effective prevention strategies and analyzing interactions with new emerging diseases such as Long COVID.

1.3 Problem Statement and Research Objectives

The COVID-19 pandemic represented a structural shock for European healthcare systems, revealing not only short-term pressures related to hospital capacity and mortality, but also deeper and more persistent vulnerabilities. As the emergency phase of the pandemic gradually receded, increasing attention has been directed toward Long COVID, a condition characterized by the persistence of symptoms well beyond the acute infection. This emerging syndrome highlights how pre-existing individual characteristics—such as chronic diseases, metabolic conditions, and broader health status—play a crucial role in shaping post-pandemic outcomes.

Within this framework, obesity represents a particularly relevant dimension of vulnerability. Even before the pandemic, European countries were experiencing a steady increase in overweight and obesity prevalence, often described as a metabolic epidemic affecting a large share of the adult population. This trend has important implications not only for individual health, but also for the long-term sustainability of healthcare systems. The COVID-19 crisis has further emphasized how metabolic risk factors may influence both the severity of infection and the resilience of individuals to adverse health shocks.

While existing evidence has established a clear association between obesity and severe acute COVID-19 outcomes, including hospitalization and mortality (Popkin et al., 2020), less attention has been devoted to the relationship between pre-pandemic metabolic conditions and the development of persistent post-infection symptoms (Nalbandian et al., 2021). This gap in the literature suggests the need to move beyond the analysis of acute outcomes and to investigate how pre-existing vulnerabilities may shape longer-term health trajectories.

This thesis approaches Long COVID as the result of an interaction between viral exposure and pre-existing individual conditions. In this context, body mass index (BMI) is used as a standardized indicator of accumulated metabolic risk, widely available in large-scale European datasets. At the same time, the analysis recognizes that the relationship between BMI and health outcomes may differ across population groups, and between men and women, reflecting both biological and epidemiological heterogeneity.

The empirical motivation of the study draws on recent contributions that use longitudinal European data to investigate the link between pre-pandemic BMI and Long COVID risk. Rather than constituting a direct replication, this thesis takes such evidence as a starting point and develops a broader analytical framework. The aim is to assess whether the association between metabolic risk and persistent symptoms remains robust when additional dimensions of individual vulnerability are considered.

In particular, the analysis extends the standard approach in four main directions. First, it adopts a categorical representation of BMI that allows the analysis to capture non-linear differences across clinically relevant weight groups, without imposing a strictly linear relationship between body mass and Long COVID risk. Second, it embeds the empirical analysis within a health economics perspective, where health is interpreted as a form of capital that affects both individual well-being and economic productivity (Grossman, 1972). In this setting, persistent symptoms may translate into longer-term economic consequences through reduced functional capacity and labor market participation. Third, the study incorporates a broader set of pre-pandemic socio-economic and behavioral characteristics, including education, income, employment status, smoking, and physical inactivity, in order to capture multiple dimensions of vulnerability that may jointly influence health outcomes.

Fourth, the analysis considers the potential role of the policy environment by introducing a country-level classification based on the presence of sugar-sweetened beverage (SSB) taxes. This extension allows the study to explore whether structural fiscal policies are associated with differences in metabolic risk and long COVID outcomes, and whether such a relationship may operate indirectly through BMI as a mediating factor.

The empirical analysis is conducted using longitudinal data from the Survey of Health, Ageing and Retirement in Europe (SHARE), combining pre-pandemic information with data collected during the COVID-19 period. Multivariate logistic regression models are employed to estimate the probability of experiencing Long COVID as a function of pre-pandemic characteristics, controlling for demographic factors, baseline health conditions, and indicators of disease severity. Interaction effects are included to account for heterogeneous patterns across population groups.

The broader objective of the thesis is to move beyond a simple descriptive association and to provide an interpretable characterization of risk. In particular, the analysis seeks to understand whether individuals with higher pre-pandemic BMI face a systematically higher probability of developing persistent symptoms, whether this relationship varies across different segments of the population, and whether it is influenced by differences in the policy environment. By addressing these questions, the study aims to contribute to the understanding of how pre-existing metabolic and socio-economic vulnerabilities, together with structural policy factors, shape post-pandemic health outcomes.

Chapter II: The Economic Burden of Obesity

2.1 Methodology of Cost-of-Illness in Health Economics

Health economics is a specific field of public and applied economics that analyses resource allocation in a sector characterized by structural constraints, information imperfections, and significant redistributive implications. Unlike other markets, the health sector cannot be modelled using standard competitive schemes without major theoretical adjustments. What makes health a different asset from ordinary consumer goods is the uncertainty surrounding illness, the asymmetry of information between patients and healthcare providers, the heterogeneity of healthcare needs, and the presence of positive and negative externalities.

Kenneth Arrow highlighted how uncertainty and imperfect information generate systematic market failures and justify public intervention in healthcare financing and regulation (Arrow, 1963). Michael Grossman's health capital model, instead, treats health as a productive stock affecting income generation and individual productivity throughout life (Grossman, 1972). From this perspective, illness reduces human capital, with effects beyond the clinical sphere, influencing labor market participation and overall economic performance.

The concept of cost is central in economic decision-making and varies depending on the type of analysis. In cost-benefit analysis (CBA), costs are compared with expected benefits, while in cost-effectiveness analyses (CEA) they are compared with intervention effectiveness. In healthcare, comparisons with health-adjusted quality of life indicators, such as QALYs, or other utility measures like cost-utility analyses (CUA), are crucial for guiding allocative efficiency in resource-scarce systems (Drummond et al., 2015). COI quantifies resources spent on disease-related care and the associated opportunity costs.

To operationalize the burden of disease (BOD), it is essential to refer to the Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) 2019. This study evaluates 87 risk factors across 204 countries and territories. This methodology shifts the focus from mortality rates to the Disability-Adjusted Life Years (DALYs).

One DALY represents the loss of one year of full health. It is the sum of two components. Years of Life Lost (YLLs) measure premature mortality by calculating the difference between actual age at death and standard life expectancy, highlighting lost economic potential. Years Lived with Disability (YLDs) quantify time spent in less-than-full health, weighted by condition severity (GBD 2019 Risk Factors Collaborators, 2020).

To provide a quantitative overview of the global health transition, Table 2.1 illustrates the shift from environmental to metabolic risks and the rapid rise in the burden of BMI relative to other risk factors.

Tab. 2.1: Global DALYs and Ranking of Leading Risk Factors (1990–2019)

Risk Factor	**DALYs 2019 (%)**	**Rank 1990**	**Rank 2019**	**% Change in DALYs (1990–2019)**
High Systolic Blood Pressure	9.3%	7	1	+53.1%
Smoking (Tobacco)	7.9%	5	2	+24.3%
High Fasting Plasma Glucose	6.8%	11	3	+122.9%
High Body-Mass Index	6.3%	16	5	+138.2%
Household Air Pollution	3.6%	4	10	-56.1%
Unsafe Water Source	2.6%	6	13	-59.3%

Source: Elaborated from GBD 2019 Risk Factors Collaborators, 2020: pp. 1236-1237.

The data from the GBD 2019 study confirms a rapid growth in the burden of BMI compared to other risks. This growth is significantly steeper than that of smoking, which increased by only 24.3% in the same period. In contrast, traditional environmental risks such as household air pollution and unsafe water have declined by 56.1% and 59.3% respectively. High BMI moved from the 16th position in 1990 to the 5th position in 2019.

The increase in the ranking of high BMI reflects broader changes in the determinants of population health. It signals the consolidation of the epidemiological transition in high and middle-income countries, where metabolic conditions progressively replace environmental and infectious threats. Unlike risks related to sanitation, which can be mitigated through infrastructural investment, metabolic exposure is embedded in long-term consumption patterns and daily behavioral routines. From a health capital perspective (Grossman, 1972), this implies that accumulated lifestyle choices increasingly shape future productivity and medical expenditure. The persistence of high BMI across development levels therefore indicates a structural transformation in the determinants of population health.

Beyond its direct contribution to morbidity, an elevated BMI serves as a fundamental "risk multiplier" within the metabolic nexus. It acts as a primary driver for a cascade of secondary conditions, most notably high systolic blood pressure (SBP) and elevated fasting plasma glucose (FPG). The GBD framework accounts for this through mediation matrices, recognizing that the impact of BMI is often filtered through these interrelated physiological pathways. This interconnectedness implies that strategic interventions targeting adiposity could yield a "compounded health dividend" by simultaneously alleviating the burden of hypertension and type 2 diabetes.

The manifestation of the BMI-related burden exhibits significant sex-specific variations. While men generally account for a higher volume of Years of Life Lost (YLLs) due to premature fatal outcomes, the female population sustains a disproportionate burden of Years Lived with Disability (YLDs). This phenomenon, often referred to as "survival bias," suggests that women endure longer periods of functional impairment and multi-morbidity. Furthermore, the risk profile for women undergoes a critical acceleration during the post-menopausal transition, where metabolic vulnerability to blood pressure and glucose fluctuations increases sharply.

In high-SDI regions such as Western Europe, metabolic risks have largely superseded infectious and environmental threats. Italy, included as a specific subnational assessment in GBD 2019, reflects this regional trend where high BMI, SBP, and dietary risks represent the dominant triumvirate of health loss. Despite the presence of advanced healthcare systems, the persistent rise in BMI across all SDI quintiles highlights structural limitations in current prevention frameworks.

The Cost-of-Illness methodology provides a comprehensive framework for quantifying the economic dimensions of disease, integrating prevalence, mortality, morbidity, and quality of life (Withrow & Alter, 2011).

Constructing a robust COI study requires defining the analytical perspective. While a healthcare system perspective focuses on public or insurance resources, a social perspective incorporates the economic impacts on labor and families, offering a more holistic estimate. For a chronic condition like obesity, the choice of time horizon is essential; while prevalence-based models estimate the burden within a fixed period, incidence-based models track costs over a patient's lifetime to account for the accumulation of expenses due to comorbidities (Blüher, 2019).

COI estimation can be executed through top-down or bottom-up methodologies. Top-down approaches start from aggregate expenditures and attribute costs via Population Attributable Fractions (PAF), a method that relies heavily on accurate epidemiological data (Withrow & Alter, 2011). Conversely, bottom-up approaches utilize individual-level data to estimate average costs per patient, providing greater granularity for subgroup analysis and reducing aggregation distortion.

Interaction and non-linearity are crucial for accurate estimation. Chronic health events often produce synergistic rather than purely additive cost effects. In the case of obesity, the relationship between BMI and expenditure is notably convex; moving from overweight to class I obesity may result in a modest cost increase, whereas class III obesity often triggers an exponential rise in healthcare demand. This necessitates the use of quadratic terms or interaction effects to avoid underestimating the total economic impact.

In this research, the COI approach is more appropriate than CBA or CEA. Only after quantifying economic burden can efficiency assessments of prevention or treatment policies be made.

Although Cost-of-Illness (COI) studies provide an essential framework for quantifying the economic burden of disease, their estimates must be interpreted with caution. The apparent precision of aggregate figures often conceals structural limitations related to data availability, attribution methods, and modelling assumptions.

A first limitation concerns causal attribution. Obesity rarely operates as an isolated determinant. Many chronic conditions emerge from the interaction between genetic predisposition, behavioral factors, and socio-environmental influences. Assigning a specific monetary share of treatment costs to elevated BMI therefore requires assumptions about epidemiological risk that may vary across datasets. Even when Population Attributable Fractions are applied, estimates depend on the quality and comparability of underlying risk coefficients.

A second issue relates to comorbidity. Patients with obesity frequently present multiple simultaneous conditions. In practice, healthcare expenditure is not neatly divisible by diagnosis. Administrative data often record primary causes of hospitalization, while secondary metabolic factors remain implicit. This may lead either to underestimation or to double counting, depending on the accounting structure adopted.

Third, COI studies are typically static representations of a dynamic process. Most prevalence-based models calculate annual expenditure without fully capturing long-term behavioral adaptation, medical innovation, or policy change. As treatment protocols evolve and pharmaceutical options expand, cost structures may shift independently of prevalence trends.

A further limitation concerns cross-country comparability. Healthcare systems differ in financing mechanisms, pricing regulations, and access pathways. Identical clinical conditions may generate substantially different expenditure levels depending on institutional context. International comparisons must therefore be interpreted as structural approximations rather than precise equivalences.

Finally, intangible costs remain only partially captured. While direct and productivity losses can be expressed in monetary terms, reductions in well-being, autonomy, and social participation resist straightforward quantification. As a result, total economic burden may be systematically underestimated when analysis is confined to measurable market transactions.

Recognizing these methodological boundaries does not weaken the relevance of COI analysis. Rather, it clarifies its role: not as an exact accounting exercise, but as an instrument for approximating the scale and structure of economic exposure generated by chronic conditions such as obesity.

2.2 Economic Costs of Obesity

Obesity generates a multidimensional economic burden that extends beyond clinical outcomes. While Chapter 2.1 examined the methodological foundations of cost-of-illness analysis, this section focuses specifically on the concrete economic implications of obesity and distinguishes three analytically separate domains: direct medical costs, indirect productivity losses, and intangible and societal effects. Each component represents a distinct pathway through which excess body weight produces measurable economic consequences.

2.2.1 Direct Costs

Direct medical costs of obesity refer to healthcare expenditures arising from the clinical consequences associated with excess body weight. These expenditures emerge when elevated BMI increases the likelihood of developing chronic conditions that require diagnosis, monitoring, and therapeutic intervention.

In practical terms, healthcare spending linked to obesity is connected to diseases epidemiologically associated with high BMI. Global risk assessments indicate that elevated BMI contributes substantially to cardiovascular diseases, type 2 diabetes, certain cancers, and musculoskeletal disorders (GBD 2019 Risk Factors Collaborators, 2020). The economic burden therefore derives from the medical management of these conditions.

The structure of direct medical costs associated with obesity can be analytically decomposed into distinct categories of healthcare utilization.

Hospital care constitutes a major component. Individuals with obesity face increased risks of acute cardiovascular events, respiratory complications, and metabolic decompensation, all of which require inpatient treatment. Excess body mass may also complicate surgical procedures and postoperative recovery, increasing resource use per episode of care.

Pharmaceutical expenditure represents another substantial cost driver. The management of obesity-related conditions frequently requires long-term treatment for hypertension, dyslipidemia, and glycemic control. These therapeutic regimens often extend over several years and contribute significantly to overall healthcare spending.

Outpatient services and specialist consultations further contribute to direct costs. Regular monitoring of cardiometabolic indicators, diagnostic imaging, and specialist follow-up are integral components of chronic disease management.

Finally, structured weight-management programs and bariatric procedures generate targeted but high per-patient expenditures. Although applied to a limited subset of patients, these interventions are relevant in cases of severe obesity and involve considerable healthcare resources (WHO, 2000; Withrow & Alter, 2011).

Obesity does not only increase the frequency of medical intervention, but also modifies its intensity. Surgical procedures in patients with severe obesity often require specialized equipment, adapted anesthetic protocols, and longer post-operative monitoring. These elements increase the marginal cost of individual treatment episodes, even when the underlying diagnosis remains unchanged.

Each of these components represents a specific channel through which obesity increases demand for healthcare services.

The economic consequences of obesity become more pronounced in later adulthood. After the age of 50, the higher prevalence of obesity-related chronic conditions is associated with increased per-capita healthcare utilization.

Evidence from longitudinal ageing surveys indicates that elevated BMI is associated with greater vulnerability to cardiometabolic and functional impairments in older populations (Wilk et al., 2024). These health conditions are linked to more frequent medical visits and higher hospitalization rates.

In countries experiencing demographic ageing, the interaction between obesity prevalence and population structure contributes to sustained growth in aggregate medical expenditure. The fiscal relevance of obesity must therefore be considered in light of demographic trends, as ageing cohorts with higher BMI levels generate greater healthcare demand.

The clinical implications of excess adiposity intensify at higher BMI levels, leading to increasingly complex patterns of medical management. Severe obesity is associated with greater systemic inflammation, metabolic instability, and functional limitation (Blüher, 2019). These mechanisms raise the probability of advanced disease states requiring specialized care.

As a result, patients in higher BMI categories often require more intensive monitoring, additional specialist interventions, and more complex therapeutic protocols. The concentration of severe obesity cases within a population increases treatment complexity and overall resource intensity within healthcare systems.

Global analyses confirm that high BMI is among the fastest-growing risk factors in terms of overall health burden worldwide (GBD 2019 Risk Factors Collaborators, 2020). The expansion of severe obesity categories contributes to sustained pressure on healthcare systems through higher average treatment requirements.

2.2.2 Indirect Costs

Indirect costs of obesity refer to economic losses occurring outside the healthcare budget that are linked to reduced functional capacity associated with excess body weight. These costs reflect diminished economic output and reduced participation in productive activities.

Obesity is associated with impaired mobility, increased fatigue, and heightened susceptibility to chronic disease (Blüher, 2019). These conditions may constrain work capacity, particularly in occupations requiring sustained physical or cognitive performance. Indirect costs therefore translate health limitations into measurable economic effects at the labor market level.

A primary channel through which obesity generates indirect costs is reduced workplace productivity. Individuals with obesity are more likely to experience health-related absences due to complications such as musculoskeletal disorders and cardiovascular symptoms.

In addition to absenteeism, diminished performance during working hours may occur when chronic conditions limit endurance or concentration. Although less visible than formal absence, this reduction in effective productivity contributes cumulatively to economic loss.

The global increase in obesity prevalence over recent decades (Ng et al., 2014) implies that these productivity effects are not marginal but structurally expanding within working-age populations. As exposure to metabolic risk grows, aggregate labor market impacts become more significant.

These productivity constraints may precede permanent withdrawal from the workforce, especially among individuals who develop obesity-related complications during midlife.

In later stages of working life, obesity-related morbidity may increase transitions into disability status or early retirement. Longitudinal evidence suggests that elevated pre-existing BMI is associated with greater vulnerability to persistent health impairment (Wilk et al., 2024). When chronic conditions reduce functional autonomy, sustained participation in employment becomes more difficult.

In ageing societies, this dynamic has macroeconomic implications. Higher prevalence of obesity among older workers may reduce effective labor supply, increase dependency ratios, and place additional pressure on pension systems.

Obesity-related indirect costs also become visible during systemic health shocks. The COVID-19 pandemic demonstrated how pre-existing metabolic vulnerability can exacerbate economic disruption.

Individuals with obesity experienced higher risks of severe COVID-19 outcomes and prolonged recovery (Popkin et al., 2020). Post-acute sequelae have been documented as sources of extended functional limitation (Nalbandian et al., 2021; WHO, 2021). Prolonged symptoms may reduce labor participation and work capacity over extended periods.

OECD analyses indicate that long COVID has generated measurable economic impacts across member countries, including reductions in labor market participation (OECD, 2024). When obesity increases susceptibility to severe infection or extended recovery, it contributes indirectly to broader macroeconomic effects.

This interaction illustrates how obesity may amplify productivity losses during periods of systemic stress.

2.2.3 Intangible and Social Costs

Beyond measurable healthcare expenditure and productivity losses, obesity imposes a substantial burden on individual well-being. Excess adiposity is associated with reduced physical functioning, chronic pain, fatigue, and limitations in daily activities. These impairments affect quality of life even in the absence of acute medical events.

The Global Burden of Disease Study identifies high BMI as a leading contributor to years lived with disability (GBD 2019 Risk Factors Collaborators, 2020). This dimension reflects prolonged health impairment rather than mortality alone. For many individuals, obesity represents a long-term condition that constrains autonomy and everyday functioning.

Unlike direct and indirect costs, these welfare losses are not fully captured by market transactions. Nevertheless, they represent a significant reduction in individual and social well-being.

Individuals with obesity frequently experience stigma, discrimination, and social exclusion. Such dynamics can adversely affect mental health, leading to anxiety, depressive symptoms, and diminished self-esteem.

Blüher (2019) emphasizes the multifactorial biological and environmental determinants of obesity, highlighting that excess body weight cannot be reduced to individual choice alone. Despite this complexity, social narratives often reinforce individual responsibility, contributing to marginalization.

Psychological distress and social isolation may further influence health behaviors and engagement with preventive services. Although difficult to express in monetary terms, these mechanisms intensify the broader societal burden associated with obesity.

When high BMI prevalence becomes widespread across cohorts, long-term demographic and economic implications emerge. Chronic disease prevalence rises, healthcare systems face sustained demand, and younger generations may enter adulthood with elevated baseline metabolic risk.

These structural dynamics indicate that obesity is a population-level phenomenon with persistent intergenerational consequences. The accumulation of metabolic vulnerability across successive cohorts may alter future patterns of morbidity and dependency within ageing societies.

Recent global health emergencies have revealed an additional societal dimension. Individuals with elevated BMI were disproportionately affected by severe COVID-19 outcomes (Popkin et al., 2020). Post-acute conditions recognized by the WHO (2021) demonstrate how pre-existing metabolic risk may intensify long-term health impairment during crises.

This interaction underscores how obesity increases population vulnerability to systemic shocks, reinforcing existing health inequalities.

2.3 Evidence from European Cost-of-illness Studies

Empirical studies conducted across European countries provide consistent estimates of the economic burden associated with obesity. Although national healthcare systems differ in structure and financing, results converge on a set of common patterns regarding direct medical expenditure and productivity losses.

Germany provides one of the most comprehensive national estimates. Effertz et al. (2016) calculated total annual obesity-related costs of approximately €63 billion. Direct medical costs account for around €29–30 billion. These expenditures are mainly associated with the treatment of type 2 diabetes, cardiovascular diseases, and musculoskeletal disorders. Indirect costs represent roughly €33–34 billion, slightly exceeding direct costs. Productivity losses arise primarily from sickness absence, disability, and early retirement. The German case illustrates a burden structure in which indirect costs represent a large share of total economic impact.

In France, obesity-related healthcare expenditure has been estimated at approximately €8–10 billion per year, corresponding to around 4–5% of total national health spending (Emery et al., 2017). Costs increase significantly across BMI categories. Individuals in severe obesity classes incur annual medical expenditure that is 30–40% higher than overweight individuals. French estimates focus predominantly on direct medical costs borne by the public insurance system. Indirect components are less extensively quantified in national analyses. The French case therefore reflects a burden concentrated primarily within the healthcare system.

In the United Kingdom, OECD estimates indicate that obesity-related conditions account for roughly 5% of total NHS spending (OECD, 2019). This corresponds to several billion pounds annually. When productivity losses are included, long-term projections suggest that obesity may generate economic costs equivalent to 2–3% of GDP over time. Ageing plays a central role in these projections. As the share of individuals aged over 65 increases, demand for cardiometabolic treatment rises proportionally. The UK presents a dual pattern: sustained direct healthcare expenditure combined with measurable macroeconomic productivity effects.

In Italy, Turchetti et al. (2012) estimated total annual obesity-related costs at approximately €8–9 billion. Direct medical expenditure represents the majority share, mainly linked to cardiovascular disease and diabetes treatment within the "Servizio Sanitario Nazionale". Indirect costs stem from reduced productivity and premature withdrawal from employment. Although overall prevalence remains slightly lower than in some Northern European countries, demographic ageing increases projected fiscal pressure in the long term. The Italian profile is therefore characterized by a predominance of direct healthcare costs amplified by population ageing.

OECD evidence indicates that obesity in Portugal reduces life expectancy by nearly three years on average (OECD, 2019). The combined economic impact of healthcare expenditure and productivity losses is estimated at approximately 1–2% of GDP. Indirect effects are particularly visible among older workers, where obesity contributes to lower employment rates and higher disability incidence. In Portugal, productivity-related losses represent a relevant component relative to overall healthcare budget size.

Across OECD member states, obesity is estimated to reduce life expectancy by approximately 2.7 years on average (OECD, 2019). This average reduction conceals relevant sex-specific patterns. While men tend to experience higher premature mortality, women often live longer with chronic metabolic conditions. As a result, part of the economic burden shifts from early mortality to prolonged morbidity, increasing the relative weight of long-term care and disability-related expenditure within ageing populations.

Healthcare expenditure attributable to obesity typically ranges between 3% and 8% of total national health spending. When indirect costs are included, total economic burden may reach up to 3% of GDP in high-prevalence countries.

Countries with higher baseline obesity rates experience steeper projected growth in healthcare expenditure. Lower labor participation among individuals with obesity further amplifies macroeconomic impact. Overall, countries with ageing populations tend to experience increasing direct medical expenditure, while systems with labor market fragility show more pronounced indirect losses.

These differences (Table 2.2) indicate that the economic burden of obesity depends not only on epidemiological prevalence, but also on demographic composition and institutional structure.

Table 2.2 Comparative Estimates of Direct and Indirect Costs of Obesity in Selected European Countries

Author(s)	**Year**	**Country**	**Method of Estimation**	**Main Cost Components**	**Estimated Costs**
Effertz et al.	2016	Germany	Prevalence-based Cost-of-Illness; top-down + productivity modelling	Direct medical costs; indirect costs (sickness absence, disability, early retirement)	€63 billion annually (≈2% GDP)
Emery et al	2017	France	Prevalence-based COI; administrative health insurance data	Direct healthcare expenditure by BMI class	€8–10 billion annually (≈4–5% health expenditure)
OECD	2019	UK	Dynamic simulation model	NHS expenditure; labor market participation effects	5% NHS expenditure; ≈2–3% GDP
Turchetti et al.	2012	Italy	Cost-of-Illness approach; national health expenditure data	Direct medical costs; indirect productivity losses	€8–9 billion annually
Withrow & Alter	2011	High-income countries	Systematic review of COI studies	Direct healthcare costs attributable to obesity	3–8% of total

		(including Europe)			healthcare expenditure
OECD	2019	OECD countries	Microsimulation modelling; long-term projections	Healthcare costs; productivity losses; GDP impact	3.3% GDP annually (projected)

The comparison presented in Table 2.2 highlights the interaction between epidemiological exposure and welfare architecture. In countries characterized by ageing populations and sustained labor market participation among older cohorts, indirect costs linked to disability and early retirement represent a substantial share of total burden. This suggests that obesity influences not only healthcare expenditure, but also effective labor supply and social protection systems. Methodological variation across studies further advises caution in direct numerical comparison, as differences in estimation strategy may partially explain observed cost shares. Nevertheless, the overall pattern confirms that the economic implications of obesity extend beyond the healthcare sector and interact with broader macroeconomic structures.

Beyond aggregate fiscal impact, obesity-related costs display a pronounced socioeconomic gradient. Prevalence patterns are not evenly distributed across populations. Lower educational attainment, income constraints, and unequal access to healthy food environments are associated with higher obesity rates (Ng et al., 2014).

This distributional asymmetry has economic consequences. In publicly financed healthcare systems, higher prevalence among disadvantaged groups implies a redistribution of medical expenditure across the tax base. While access to care remains universal, exposure to metabolic risk is socially patterned. The burden is therefore both epidemiological and distributive.

Labor market effects also exhibit inequality. Individuals in physically demanding occupations may experience earlier functional limitation when affected by obesity-related conditions. In contrast, higher-income groups are more likely to hold positions compatible with reduced physical capacity or flexible working arrangements. Productivity losses may thus disproportionately affect lower socioeconomic strata.

Gender intersects with socioeconomic status in shaping vulnerability. Women in lower-income groups often face combined constraints related to caregiving responsibilities, occupational segmentation, and health access. When obesity interacts with these structural factors, the resulting economic exposure may be compounded over the life course.

Regional disparities further reinforce this gradient. Urban planning, food pricing, and availability of recreational infrastructure differ across territories. These environmental determinants influence long-term metabolic risk and, indirectly, the distribution of future healthcare demand.

Considering these distributional patterns broadens the interpretation of obesity-related costs. The burden is not only a matter of aggregate fiscal sustainability, but also of social equity. Policies aimed at reducing obesity prevalence therefore operate simultaneously as health interventions and as instruments of redistribution.

These structural and distributional cost patterns provide the economic rationale for examining preventive fiscal interventions designed to reduce long-term metabolic risk.

Chapter III: Preventive Fiscal Policies: The Case of the Sugar Tax

3.1 Pigouvian Taxes in Nutrition: Correcting Externalities and Behavioral Market Failures

3.1.1 Externalities and the Pigouvian Framework

Public intervention in food markets can be grounded in standard welfare economics. Under competitive conditions, prices are assumed to reflect the full costs and benefits of consumption. When information is perfect, and no external effects are present, decentralized choices lead to efficient outcomes. The health sector does not operate under these assumptions.

Arrow (1963) showed that uncertainty and asymmetric information are structural features of health markets. Individuals cannot fully anticipate future health states, nor can they accurately evaluate the long-term consequences of present consumption decisions. In systems where healthcare is largely financed through taxation or social insurance, individual behaviors may generate costs that are redistributed across the population.

The consumption of sugar-sweetened beverages provides a clear illustration. The price paid by the consumer reflects production and distribution costs, but it does not incorporate the expected increase in public healthcare expenditure associated with obesity-related conditions.

In welfare terms, this divergence can be expressed as:

$$MSC = MPC + MEC$$

where:

- MSC (Marginal Social Cost) indicates the marginal social cost; the total cost to society resulting from an additional unit of consumption;
- MPC (Marginal Private Cost) represents the marginal cost incurred directly by the consumer or producer;
- MEC (Marginal External Cost) is the marginal external cost; the additional damage imposed on third parties and not incorporated in the market price.

If decisions are based solely on MPC, the equilibrium quantity exceeds the socially optimal level. The presence of MEC implies overconsumption relative to the welfare maximizing allocation.

Within the classical Pigouvian framework, the corrective tax is defined as:

t = MEC(Q^*)

where:

- t is the Pigouvian tax per unit of goods;
- MEC(Q^*) is the marginal external cost calculated at the socially optimal quantity (Q^*);
- Q^* represents the quantity that maximizes collective welfare, the point at which the marginal benefit of consumption equals the marginal social cost.

In this formulation, the optimal tax exactly offsets the external component embedded in individual consumption decisions, aligning private incentives with social costs.

3.1.2 Behavioral Market Failures and Internalities

While the classical Pigouvian framework justifies taxation based on external costs imposed on third parties, more recent contributions in behavioral welfare economics suggest that overconsumption may also arise from distortions internal to the individual. The problem is not limited to the divergence between private and social costs; it may also reflect a mismatch between short-run choices and long-run preferences.

Models of time inconsistent behavior formalize this mechanism. Gruber and Kőszegi (2001) and O'Donoghue and Rabin (2006) show that individuals may display present bias, systematically giving excessive weight to immediate gratification relative to future consequences. Under hyperbolic discounting, preferences can be represented as:

$$U = U_0 + \beta \sum_{t=1}^{T} \delta^t U_t$$

where:

- U represents the individual's overall utility;
- U_0 is the immediate utility derived from present consumption;
- U_t indicates utility (or disutility) in future periods;
- δ is the traditional discount factor, between 0 and 1, which expresses preference for the present.

When $\beta < 1$, the individual places excessive weight on immediate benefits compared to future costs. Decisions that appear optimal in the present may therefore conflict with the individual's own long-term welfare objectives.

In the context of dietary choices, this mechanism is particularly relevant. The utility derived from consuming sugar-sweetened beverages is immediate and salient, while the associated health consequences emerge gradually, such as weight gain, metabolic dysregulation, and increased risk of chronic disease. If individuals systematically underweight these delayed costs, observed consumption may exceed the level consistent with their own long-term preferences. The resulting welfare loss is referred to as an internality.

O'Donoghue and Rabin (2006) demonstrate that corrective taxation can improve welfare even when external costs are limited, provided that the tax helps individuals better align short-term decisions with long-term goals. Similarly, Gruber and Kőszegi (2001) show that sin taxes may function as commitment devices, reducing the distortions generated by time-inconsistent behavior.

The introduction of internalities therefore expands the normative foundation of Pigouvian taxation. Fiscal intervention in nutrition policy is not justified solely as a means of internalizing social costs, but also as a response to systematic behavioral biases that lead to persistent over consumption of health damaging goods.

Corrective taxation in nutrition policy is frequently criticized for its potential regressivity. Taxes on sugar-sweetened beverages are often described as burdensome for low-income households, since these groups devote a larger share of their income to such products. However, financial incidence alone does not provide a complete welfare assessment.

Allcott, Lockwood and Taubinsky (2019) propose a framework in which the optimal level of a sin tax depends on both external costs and internalities, as well as on differences in behavioral responses across income groups. In this setting, a tax that appears regressive in monetary terms may still increase welfare if it reduces systematic decision-making distortions. When present bias is stronger among certain consumers, the corrective effect of taxation may be larger for those individuals.

Price responsiveness is central to this evaluation. Consider a beverage priced at 1 euro per unit. A tax of 0.20 euros increases the price by 20%. If the price elasticity of demand is -1.0, consumption declines by approximately 20%. If elasticity is -0.3, consumption falls by only 6%. The same tax therefore produces different behavioral adjustments. Consumers who respond more strongly to price changes reduce intake more substantially and are more likely to experience long-term health improvements. Those who adjust less face a higher relative financial burden without a comparable behavioral response.

For this reason, welfare analysis cannot be confined to short-term expenditure shares. The relevant perspective must include long-term health outcomes and productivity effects. A corrective tax is designed to address a specific distortion associated with external or internal costs. A revenue tax is primarily intended to raise fiscal resources. Conflating these objectives obscures the purpose of the intervention.

When internalities and externalities are considered jointly, the normative question becomes more precise. The key issue is whether the tax reduces inefficient consumption in a way that improves aggregate welfare while limiting unintended redistributive consequences. Optimal sin tax design therefore requires balancing efficiency, behavioral correction, and equity within a coherent policy framework.

3.1.3 Health Capital and Preventive Fiscal Policies

The economic rationale for corrective taxation becomes clearer when health is viewed as a form of capital. In the model developed by Grossman (1972), health is not only a source of utility but also a productive asset. Individuals invest in health to extend healthy time and maintain productivity. When health deteriorates, both utility and earning capacity decline.

Obesity can be interpreted within this framework as a gradual depreciation of health capital. It does not produce immediate collapse, but it increases the probability of chronic disease and functional limitations over time.

The cost of illness approach discussed in Chapter 2 measures the economic consequences of this process once disease has materialized. It quantifies hospitalizations, pharmaceutical treatments, productivity losses and long-term care. This perspective is necessary, but it is reactive. It evaluates the cost of managing an existing condition rather than the cost of preventing its emergence.

Fiscal instruments operate at an earlier stage. By altering relative prices, taxation affects consumption patterns before metabolic risk accumulates. The objective is not to treat obesity, but to influence the determinants that shape its distribution within the population. In this sense, corrective taxation functions as a form of ex-ante regulation of health risk.

The potential advantage of this approach lies in its cumulative effect. Small reductions in average sugar intake can translate into gradual shifts in body mass index at the population level. Over time, even modest changes in the distribution of BMI may reduce the incidence of obesity-related diseases and the associated economic burden. This preventive logic differs fundamentally from ex-post medical intervention.

The argument is reinforced when health systems are publicly financed. In such settings, preventing the deterioration of health capital contributes to fiscal sustainability. A reduction in long-term chronic disease incidence implies lower future expenditure and potentially higher productivity. This dynamic is often described as a double dividend. The first dividend arises from improved health outcomes. The second emerges through reduced pressure on public budgets.

From this perspective, nutrition taxation can be interpreted as a structural instrument aimed at influencing the long-term trajectory of health capital rather than merely financing the treatment of established disease.

Health shocks do not affect populations uniformly. The COVID 19 pandemic showed that preexisting metabolic conditions influenced both the severity of infection and the persistence of symptoms. Obesity emerged as a consistent risk factor for complications and prolonged recovery (Popkin et al., 2020; Wilk et al., 2024).

Body mass index can be interpreted as an indicator of accumulated metabolic exposure. It reflects long term consumption patterns and lifestyle choices that evolve within specific economic environments.

When metabolic risk is unevenly distributed, systemic shocks may produce asymmetric outcomes. Elevated BMI is associated with higher probabilities of severe disease and post-acute conditions. In this context, preventive fiscal measures acquire structural relevance. By influencing incentives before risk fully materializes, policy may alter baseline vulnerability.

The connection between taxation and post infectious outcomes is indirect but coherent. Fiscal incentives shape consumption. Consumption patterns affect metabolic risk. Metabolic risk conditions health resilience. Within this framework, nutrition taxation can be understood as a strategy aimed at moderating structural vulnerability rather than responding only to realized health expenditure.

3.2 Comparative Analysis of Sugar-Sweetened Beverage (SSB) Taxes in Europe

Fiscal intervention in food and beverage markets has progressively moved from revenue-oriented excise duties to instruments explicitly framed as public health policies. Within this evolution, sugar-sweetened beverage taxation has become one of the most visible examples of health-related fiscal reform.

Excise duties on soft drinks are not new. Denmark introduced a tax on non-alcoholic beverages in the 1930s, primarily for fiscal purposes. For decades, this levy functioned as a standard consumption tax without an explicit preventive objective. In 2011 Denmark also implemented a tax on foods high in saturated fat, commonly referred to as the "fat tax," which was repealed in 2012 after concerns about administrative complexity and cross-border shopping. The Danish soft drink excise itself was abolished in 2014. These experiences illustrate that early beverage taxes were not originally designed as structured responses to obesity or diet-related disease.

A clearer public health orientation emerged in the 2010s. In 2014 Mexico introduced a nationwide excise tax of one peso per liter on sugar-sweetened beverages under the Special Tax on Production and Services. The reform was justified by rising obesity and diabetes prevalence. Empirical evaluations documented significant reductions in purchases during the first years of implementation (Colchero et al., 2016). This experience marked a turning point in the international policy debate and contributed to the global diffusion of similar measures.

International organizations subsequently endorsed fiscal approaches. The World Health Organization (2016) recommended taxation of sugar-sweetened beverages as part of strategies to prevent non-communicable diseases, while the OECD (2019) identified beverage excise duties as instruments with potential long-term health and fiscal benefits. By the late 2010s, multiple jurisdictions across Latin America, North America, Asia, and Europe had adopted some form of sugar tax.

Although policy objectives converged around reducing excessive sugar intake, tax structures differed substantially. Some governments applied flat volumetric excise duties per liter of beverage. Others implemented tiered systems linked to sugar concentration. These structural differences shaped incentives for reformulation, price transmission, and consumer response, generating heterogeneous outcomes across countries.

The following sections examine selected European cases within this broader international trajectory. Each country is analyzed according to policy design, transmission mechanisms, and documented empirical effects, allowing for a structured comparison of different fiscal models within a common regional context.

3.2.1 United Kingdom: Soft Drinks Industry Levy

The United Kingdom introduced the Soft Drinks Industry Levy in April 2018. The measure was announced in 2016 and designed as a tiered excise tax applied to producers and importers rather than directly to consumers. The levy applies to sugar-sweetened beverages according to sugar concentration thresholds. Drinks containing between 5 and 8 grams of sugar per 100 milliliters are subject to a lower rate, while those exceeding 8 grams per 100 milliliters face a higher rate (HM Treasury, 2016). Beverages below 5 grams per 100 milliliters are not taxed.

This tiered structure was intended to create an incentive for reformulation rather than simply increasing prices. By linking the tax liability to sugar content, the policy encouraged manufacturers to reduce sugar concentration to move products below taxable thresholds. The mechanism therefore operated on the supply side as well as through consumer prices.

Empirical evidence indicates that the reform led to substantial changes in product composition. Scarborough et al. (2020) document a reduction of approximately 29% in the average sugar content of soft drinks available on the UK market between 2015 and 2019, following the announcement and implementation of the levy. A significant share of products was reformulated to avoid the higher tax band. The overall volume of sugar sold in beverages declined, even though total sales volume remained relatively stable. This suggests that reformulation played a central role in the observed outcomes.

Evidence on consumer response shows moderate price pass-through, with variation across product categories (Scarborough et al, 2020). The tax burden was not uniformly shifted to final prices. In several cases, producers absorbed part of the levy or adjusted product characteristics instead of fully increasing retail prices. As a result, the policy generated both price and composition effects.

From an economic perspective, the UK model illustrates how tax design shapes behavioral responses. A tiered structure reduces incentives for simple substitution toward slightly cheaper sugary alternatives and instead targets sugar concentration directly. Compared with flat volumetric taxes, the Soft Drinks Industry Levy created stronger reformulation incentives while maintaining flexibility for producers. At the same time, the policy required clear thresholds and monitoring capacity, increasing administrative complexity.

Overall, the UK experience demonstrates that fiscal intervention can influence both supply and demand channels. The magnitude of reformulation observed after 2018 distinguishes the British model from earlier excise designs and has made it a reference case in the European policy debate.

3.2.2 Belgium: Fixed-volume sugar-sweetened beverage tax

Belgium introduced a tax on sugar-sweetened beverages in 2016 as part of a broader fiscal reform package aimed at improving public health outcomes. The measure takes the form of a fixed volumetric excise applied per liter of beverage. The tax does not increase progressively with sugar concentration beyond the taxable threshold. All beverages within the taxable category are subject to the same rate per volume (WHO, 2016; OECD, 2019).

This design implies that the fiscal burden is largely independent of marginal differences in sugar content above the minimum threshold. Unlike tiered systems, the Belgian model does not create strong incentives for producers to reformulate products to move across tax brackets. The primary transmission channel therefore operates through price adjustments rather than structural changes in product composition.

Available evidence suggests moderate price pass-through and limited reformulation effects compared with more differentiated tax structures (OECD, 2019). Because the tax liability does not vary with sugar concentration once the product is taxable, producers face weaker incentives to substantially reduce sugar content. As a result, the policy mainly affects consumer prices rather than directly reshaping product formulation.

Economically, the Belgian case illustrates the trade-off between administrative simplicity and behavioral targeting. A uniform volumetric excise is easier to implement and monitor, and it provides predictable revenue. However, it offers weaker supply-side incentives relative to tiered models that directly penalize higher sugar concentrations. The Belgian experience therefore underscores how tax structure conditions the channel through which adjustment occurs in the beverage market.

3.2.3 Hungary: Tax on sugar-containing foods and beverages

Hungary introduced the Public Health Product Tax in 2011, becoming one of the first

European countries to implement a health-oriented excise targeting sugar-containing foods and beverages. Unlike policies limited to soft drinks, the Hungarian measure applied to a broader range of products, including sugary beverages, confectionery, energy drinks, and certain processed foods with high sugar or salt content (WHO, 2016). The objective was explicitly linked to improving dietary habits and reducing the prevalence of diet-related diseases.

The tax was structured as a product-based excise, with rates defined according to predefined nutrient thresholds. Products exceeding specified sugar or salt content levels became taxable. This approach created incentives not only to reduce consumption through price increases but also to adjust product composition in order to avoid taxation.

Empirical evidence suggests that the reform generated measurable behavioral responses. Bíró (2015) reports a decline in the consumption of taxed products following implementation, accompanied by evidence of product reformulation in certain categories. The effects were not uniform across all product groups, reflecting differences in market structure and consumer substitution patterns. While price increases played a role, changes in supply and product characteristics were also observed.

From an economic standpoint, the Hungarian model illustrates a more comprehensive application of fiscal instruments to unhealthy food categories. By extending taxation beyond beverages, the policy aimed to influence a wider set of dietary choices. At the same time, the broader scope increased administrative complexity and required clear nutrient-based definitions. The Hungarian experience highlights how the breadth of the tax base can affect both behavioral impact and implementation challenges.

3.2.4 France: Modulated sugar tax model

France introduced a tax on sugar-sweetened beverages in 2012, becoming one of the first European countries to implement a health-oriented excise on soft drinks. Initially structured as a flat volumetric tax applied per liter of beverage, the measure was later revised to incorporate modulation according to sugar concentration. This adjustment aimed to strengthen incentives for reformulation while maintaining a broad tax base (WHO, 2016).

Under the original design, the tax applied uniformly to sugar-sweetened beverages regardless of precise sugar content. This structure primarily affected retail prices, with limited differentiation across products. Subsequent modifications introduced a more graduated system, linking the tax burden more closely to sugar levels and thereby increasing the incentive to reduce concentration.

Empirical evidence indicates that the initial flat tax generated measurable price increases and moderate reductions in purchases. Using French micro-level data, Berardi et al. (2016) find that the tax was largely passed through to consumer prices and associated with a decline in purchases of taxed beverages. The magnitude of the effect, however, reflected the relatively modest tax rate and the absence of strong differentiation across sugar categories in the early phase of the policy.

In economic terms, the French experience illustrates the evolution from a simple revenue-oriented excise toward a more targeted public health instrument. The shift from a uniform volumetric tax to a modulated structure reflects recognition that tax design influences behavioral adjustment. Compared with fully tiered systems, the French model occupies an intermediate position: broader than simple flat taxes yet less sharply segmented than the British levy. The case highlights how incremental policy reform can alter incentive structures over time without fundamentally redesigning the fiscal instrument.

3.2.5 Portugal: Tax integrated with preventive policies

Portugal introduced a tax on sugar-sweetened beverages in 2017 within the framework of its Integrated Strategy for the Promotion of Healthy Eating. The measure formed part of a broader set of preventive policies aimed at addressing obesity and diet-related non-communicable diseases. Unlike purely fiscal reforms, the Portuguese approach was explicitly embedded in a national public health strategy (WHO, 2016).

The tax was designed as a tiered volumetric excise applied per liter of beverage, with rates differentiated according to sugar concentration. By introducing distinct brackets, the policy sought to create incentives for reformulation while also affecting final prices. The structure therefore combined price effects with supply-side adjustment mechanisms.

Available evidence indicates reductions in the sugar content of beverages following implementation, alongside moderate price transmission (OECD, 2019). Reformulation played a relevant role, as producers adjusted product composition to move into lower tax categories. The impact on overall sales volumes appears more limited than in cases characterized by higher tax differentials, suggesting that design parameters influence the magnitude of behavioral change.

From a policy perspective, the Portuguese model illustrates how fiscal instruments can be integrated into a broader preventive framework. The policy did not operate in isolation but was accompanied by nutritional guidelines and public awareness initiatives. This coordination may enhance policy coherence, although isolating the independent effect of taxation becomes more complex. The case demonstrates that tax design and institutional embedding jointly shape observed outcomes.

3.2.6 Italy: Case of incomplete implementation

Italy approved a sugar-sweetened beverage tax within the 2020 Budget Law, formally introducing an excise on sweetened drinks. The measure was designed as a volumetric tax applied per hectoliter of beverage, with a differentiated rate for finished products and concentrates. The stated objective was to discourage excessive sugar consumption and contribute to public health goals.

Despite legislative approval, implementation was repeatedly postponed. Successive budget revisions delayed entry into force, citing economic conditions, administrative readiness, and concerns raised by industry representatives. As a result, the tax has not been fully operational. Official technical reports accompanying the Budget Law estimated potential revenue and projected moderate reductions in consumption, but these projections remain ex ante assessments rather than observed outcomes.

The Italian proposal did not adopt a strongly tiered structure linked to sugar concentration. The design therefore resembled a standard volumetric excise rather than a differentiated model explicitly targeting reformulation. In the absence of progressive brackets, the expected adjustment mechanism would have relied primarily on price transmission rather than substantial changes in product composition.

From an economic perspective, the Italian case illustrates the gap that can emerge between policy design and policy implementation. The measure was formally justified on health grounds and aligned with recommendations from international organizations, yet political and economic constraints prevented its execution. This incomplete implementation limits the availability of empirical evidence but offers insight into the institutional challenges surrounding fiscal interventions in food markets. The Italian experience highlights how stakeholder opposition, sectoral interests, and timing considerations can influence the feasibility of preventive taxation.

3.2.7 Comparative Evidence and Quantitative Outcomes

The comparative analysis of the European cases highlights substantial differences in the intensity and adjustment mechanisms generated by sugar-sweetened beverage taxes, with quantitative outcomes directly linked to tax design. The United Kingdom represents the case where the intervention most deeply reshaped market supply. The Soft Drinks Industry Levy, structured around thresholds of 5g and 8g of sugar per 100 ml, led to a reduction of 33.8 percentage points in the share of beverages above the 5g threshold relative to the counterfactual; by 2019, only 15.4% of "intervention" products remained above that threshold. In addition, an average reduction of 2.13 grams of sugar per 100 ml was estimated for affected products, with price pass-through of approximately 31% in the higher tier (Scarborough et al., 2020). Crucially, the decline in sugar content occurred primarily through reformulation rather than contraction in total sales volume. Adjustment therefore operated mainly through supply-side transformation rather than demand compression.

In France, the initial effect was more consistent with a price-demand mechanism. For regular cola beverages, prices increased by approximately 5% in 2012 and 3.1% in 2013, accompanied by reductions in demand of about 3.3% and 3.4% respectively (Berardi et al., 2016). Here, the tax primarily affected the extensive margin of consumption, with moderate implicit price elasticity and without evidence of systematic reformulation in the early phase. Compared with the United Kingdom, the reduction in sugar consumption in France stemmed mainly from lower purchased quantities rather than structural changes in product composition.

Hungary displays stronger quantitative demand effects than France, but with a different adjustment pattern from the UK. For cola beverages, price increases of 3.4% in 2011, 1.2% in 2012, and 0.7% in 2013 were observed, while demand fell by 2.7% in 2011, 7.5% in 2012, and 6% in 2013 (Bíró, 2015). Evidence also indicates substitution toward non-branded or untaxed products. Compared with France, volume contraction was more pronounced in certain years; however, unlike the UK, there was no comparable structural reformulation effect. The dominant channel remained reduced consumption, partially offset by substitution behavior.

Belgium presents more limited quantitative outcomes, consistent with its uniform volumetric excise structure and absence of strong sugar-content differentiation. Evidence indicates partial pass-through to prices and moderate reductions in purchased quantities, with limited signs of systematic reformulation (Vandevijvere et al.,2020; OECD, 2019). Compared with France and Hungary, the contraction in volumes appears weaker; compared with the United Kingdom, the absence of sharply differentiated sugar thresholds reduced marginal incentives to reformulate. The Belgian case illustrates how a flat tax generates more linear and less transformative effects.

Portugal occupies an intermediate position but shows quantitatively relevant results. The 2017 tax is associated with an annual reduction of approximately 6.6 million litres in sales (around −21% in volume) and an average reduction in energy density of 3.1 kcal per 100 ml. Epidemiological modelling further estimates the prevention of approximately 40–78 new obesity cases per year during the period considered, with stronger relative effects among adolescents (Goiana-da-Silva et al., 2020). Unlike Belgium and partially France, the Portuguese experience combines volume reduction and reformulation. However, compared with the UK, the structural supply adjustment appears less sharply concentrated around threshold incentives and more evenly distributed across both channels.

Overall, three distinct economic configurations emerge. The British model generates the most substantial impact on product composition and total sugar introduced into the market. The French and Hungarian models primarily reduce demand through price effects, with measurable but moderate elasticities. Belgium produces comparatively limited quantitative changes, reflecting its simpler design. Portugal combines both supply and demand channels with measurable projected health gains. From an allocative efficiency perspective, tiered systems linked directly to sugar concentration are associated with deeper qualitative transformation of the beverage market, whereas uniform volumetric excises operate mainly through quantity adjustment.

Italy is excluded from the quantitative comparison because, in the absence of effective and stable implementation, there is no observable post-policy period allowing credible estimation of pass-through, sales changes, or reformulation responses comparable to the other cases.

Table 3.1: Comparative quantitative evidence on European sugar-sweetened beverage taxes

Country	**Year of implementation**	**Tax structure**	**Price increase %**	**Demand reduction %**	**Reformulation evidence**
UK	2018	Tiered (5g / 8g per 100ml)	~31% pass-through (high tier)	Limited contraction in total volume	Strong, widespread reformulation; −33.8 pp above 5g threshold
France	2012	Volumetric (initially flat)	+5% (2012); +3.1% (2013)	−3.3% (2012); −3.4% (2013)	Limited initial reformulation
Hungary	2011	Limited initial reformulation	+3.4% (2011)	−2.7% (2011); −7.5% (2012); −6% (2013)	Evidence of substitution toward untaxed/non-branded products

Belgium	2016	Uniform volumetric excise	Partial pass-through	Moderate reduction	Weak reformulation incentives
Portugal	2017	Tiered (sugar concentration-based)	Moderate pass-through	−21% volume (~6.6M liters/year)	Reformulation combined with volume reduction

Sources: Quantitative estimates are drawn from Scarborough et al. (2020), Berardi et al. (2016), Bíró (2015), Vandevijvere et al., and Goiana-da-Silva et al. (2020).

The comparative analysis shows that the effects of sugar-sweetened beverage taxes vary according to tax design and institutional context. The evidence indicates that it is not merely the introduction of a tax that determines its impact, but the way in which it structures incentives for producers and consumers. In some cases, adjustment occurs mainly through prices and quantities; in others, it involves changes in product composition and supply structure.

This heterogeneity makes it necessary to move beyond descriptive evidence and toward economic evaluation. The next section therefore examines the criteria through which fiscal interventions can be assessed in terms of efficiency, sustainability, and overall welfare implications.

3.3 Cost-effectiveness, Structural Prevention and Post-pandemic Resilience

Before the COVID-19 pandemic, the economic evaluation of nutrition-related fiscal policies focused mainly on chronic non-communicable diseases and related healthcare expenditure. Preventive interventions were assessed according to their ability to reduce obesity prevalence, diabetes incidence, cardiovascular events, and long-term treatment costs (OECD, 2019; GBD 2019 Risk Factors Collaborators, 2020). Within this framework, public intervention was justified by market failures and imperfect information (Arrow, 1963; Grossman, 1972), and cost-effectiveness analysis concentrated on measurable health gains.

Under this conventional approach, the Incremental Cost-Effectiveness Ratio (ICER) represents a central decision tool. In preventive nutrition policy, the incremental effect has typically been defined in terms of reduced obesity cases, improvements in metabolic indicators, or gains in health-adjusted life years (Drummond et al., 2015; OECD, 2019).

Empirical evidence from economic evaluations provides a more concrete assessment of these mechanisms. A systematic review by Liu et al. (2022), covering 15 economic evaluation studies across different countries, shows that sugar taxation is consistently found to be cost-effective and, in many cases, cost-saving. The results indicate that reductions in sugar consumption can generate substantial health gains, reaching up to 3.4 million quality-adjusted life years over a lifetime horizon, alongside significant reductions in healthcare expenditure. Estimated avoided healthcare costs range from around $87 million to over $113 billion depending on the setting and time horizon. Importantly, avoided healthcare costs and tax revenues are found to exceed implementation costs, implying negative net costs from both governmental and societal perspectives. In some cases, each dollar invested in sugar taxation is estimated to generate savings of around $60, highlighting its potential as an economically efficient preventive intervention.

The pandemic has altered this perspective. Evidence shows that pre-existing metabolic conditions, including elevated BMI, increased both the severity of acute infection and the probability of persistent symptoms (Popkin et al., 2020; Wilk et al., 2024). BMI therefore emerges not only as a determinant of chronic disease, but as a factor shaping vulnerability to systemic shocks.

This shift changes the interpretation of what counts as a “health effect” in economic evaluation. If higher BMI amplifies long-term impairment following infection, then the incremental benefit of preventive policy extends beyond cardiometabolic outcomes. It includes reduced exposure to prolonged functional limitation and productivity losses linked to post-acute conditions (OECD, 2024). In this sense, the objective of intervention expands from limiting disease-specific costs to strengthening structural resilience.

The temporal dimension is central in this assessment. Sugar taxation does not generate immediate health improvements. Consumption patterns adjust gradually. Product reformulation takes time. Body weight responds slowly to sustained dietary changes. BMI reflects cumulative exposure to behavioral and environmental factors (Blüher, 2019; Ng et al., 2014). For this reason, short-term evaluations may underestimate long-run effects.

The economic relevance of this dynamic becomes clearer when considering the relationship between BMI and healthcare costs. Expenditure does not increase proportionally across the BMI distribution. At higher obesity levels, comorbidities become more frequent and treatment more complex (Withrow & Alter, 2011; GBD 2019 Risk Factors Collaborators, 2020). Small reductions among individuals in higher-risk categories may therefore generate larger economic benefits than equivalent reductions around the average.

Preventive taxation can influence not only mean sugar intake but also the distribution of metabolic risk. Even modest changes in product composition or prices, documented in several European contexts (Scarborough et al., 2020; Berardi et al., 2016; Bíró, 2015), may gradually reduce the probability that individuals move into more severe BMI classes. Over time, this affects both morbidity and expenditure trajectories.

Tax design also matters. Tiered systems linked to sugar concentration create incentives for reformulation. Producers may reduce sugar content to move products below taxable thresholds. In such cases, part of the health effect occurs through supply-side adjustment. Uniform volumetric taxes rely more heavily on price responses and consumer elasticity. The same nominal rate can therefore produce different long-term outcomes depending on how incentives are structured.

Efficiency cannot be assessed solely through average consumption changes. The location of the effect within the population is relevant. If policy influences individuals already close to obesity thresholds, the long-term return may be greater than aggregate averages suggest. Heterogeneity across demographic groups reinforces this point. Metabolic risk and its consequences differ by sex and biological characteristics (Blüher, 2019; Wilk et al., 2024). Identical changes in BMI may not translate into identical health outcomes for men and women.

Preventive fiscal policy also has value under uncertainty. Health systems operate in environments where future shocks are unpredictable. The COVID-19 crisis has shown how pre-existing vulnerabilities can amplify economic disruption (OECD, 2024). Before 2020, Long COVID was not included in policy projections. Yet metabolic conditions were associated with higher risk of persistent impairment (Popkin et al., 2020; Wilk et al., 2024).

Reducing metabolic risk today can therefore be interpreted as a form of protection against uncertain future losses. By moderating the accumulation of vulnerability, policy may limit the scale of health and productivity effects when shocks occur. This dimension does not replace traditional cost-effectiveness metrics, but it broadens their interpretation.

Overall, evaluating the cost-effectiveness of sugar taxation requires a perspective that goes beyond immediate price effects and short-term savings. It involves considering long-term exposure, distributional patterns of BMI, and the interaction between metabolic risk and systemic crises.

From a policy perspective, understanding whether metabolic risk factors influence post-viral health outcomes is relevant for the economic evaluation of preventive interventions. If higher BMI increases the probability of persistent health impairment, policies aimed at reducing metabolic risk may generate benefits that extend beyond traditional chronic disease prevention.

Within this framework, analyzing the relationship between pre-pandemic BMI and later health outcomes becomes particularly relevant. The empirical analysis developed in the following chapters examines whether elevated BMI increased the probability of Long COVID in a European longitudinal setting. By quantifying this association, the study provides evidence on one of the mechanisms through which accumulated metabolic risk may translate into persistent health and economic consequences.

Chapter IV: Data and Methodology

This chapter addresses the research question introduced in Section 1.3, namely whether pre-pandemic body mass index (BMI) is associated with the probability of developing Long COVID and whether this relationship varies across individuals. It presents the data and empirical strategy used to analyze the relationship between pre-pandemic BMI and the risk of developing Long COVID.

The study investigates whether higher BMI is associated with an increased probability of persistent symptoms following COVID-19 infection, and how this relationship is influenced by demographic characteristics, pre-existing health conditions, and socio-economic factors.

The empirical objective is to estimate this association using individual-level longitudinal data. To this end, logistic regression models are employed, incorporating a set of control variables.

The following sections describe the data, the construction of the analytical sample, the definition of variables, and the econometric specification.

4.1 The Survey of Health, Ageing and Retirement in Europe (SHARE): Waves 6, 7, 8, and Corona Surveys.

This study uses data from the Survey of Health, Ageing and Retirement in Europe (SHARE), a multidisciplinary and cross-national research infrastructure designed to collect harmonized micro-level information on individuals aged 50 and over across European countries (Börsch-Supan et al., 2013). SHARE was developed to support the study of population ageing by providing internationally comparable data on health conditions, socio-economic circumstances, and demographic characteristics among middle-aged and older individuals. Because of its longitudinal structure and its broad thematic coverage, SHARE represents an appropriate data source for analyzing how pre-existing individual characteristics are associated with later health outcomes.

The survey covers several dimensions of respondents' lives and includes detailed information on physical and mental health, diagnosed chronic diseases, health behaviors, educational attainment, employment status, retirement transitions, household income and wealth, as well as family relations and social networks. This multidimensional structure is one of the main strengths of SHARE, since it allows researchers to combine indicators of health and socio-economic vulnerability within a single harmonized framework. This richness is particularly relevant for the present analysis, which extends beyond BMI to include a broader set of pre-pandemic characteristics.

Since its first implementation in 2004, SHARE has become one of the main sources of harmonized microdata for research on ageing in Europe (Börsch-Supan et al., 2013). The project is coordinated by an international research consortium and follows common methodological guidelines across participating countries and survey waves. Questionnaire design, translation procedures, fieldwork protocols, and data processing are centrally harmonized to improve cross-country comparability and reduce the risk that differences across national samples reflect methodological discrepancies rather than substantive variation in respondents' characteristics (SHARE Release Guide, 2022).

A defining characteristic of SHARE is its longitudinal panel structure. The same respondents are followed across successive waves, making it possible to observe the evolution of health conditions, socio-economic status, and demographic circumstances over time (Börsch-Supan et al., 2013). Unlike cross-sectional surveys, panel data allow the researcher to establish a clearer temporal ordering between explanatory factors and later outcomes. This feature is central to the empirical strategy adopted in this thesis, since the main objective is to relate pre-pandemic individual characteristics to health outcomes observed during and after the COVID-19 period. More specifically, the longitudinal structure of SHARE allows information collected before the outbreak of the pandemic to be linked to respondents' subsequent experience of COVID-19 infection and persistent post-infection symptoms.

The survey was conceived as a harmonized European infrastructure able to collect comparable microdata across countries characterized by different welfare systems, demographic structures, and public health environments (Börsch-Supan et al., 2013). This feature makes SHARE particularly useful not only for studying ageing and health in general, but also for investigating how risks associated with COVID-19 and Long COVID may emerge within a heterogeneous European population. The availability of harmonized information across countries is therefore an essential element of the empirical basis of this thesis.

Under ordinary conditions, SHARE data collection is mainly based on computer-assisted personal interviews (CAPI), in which trained interviewers administer structured questionnaires directly to respondents in face-to-face settings. This mode of data collection makes it possible to implement complex questionnaire routing, consistency checks, and standardized interviewing procedures, thereby supporting the overall quality of the data (SHARE Release Guide, 2022). However, the outbreak of the COVID-19 pandemic required important adjustments to these procedures. Restrictions on mobility and social interaction made traditional face-to-face interviews difficult or impossible in several countries, and the SHARE project therefore introduced alternative methods based on computer-assisted telephone interviews (CATI) to continue data collection during the pandemic (Börsch-Supan, 2020).

Within this context, SHARE developed two specific pandemic survey modules, commonly referred to as the SHARE Corona Surveys. These surveys were designed to monitor respondents' health conditions and living circumstances during the pandemic and collected information on COVID-19 infection, symptoms, hospitalization, access to healthcare, economic difficulties, and changes in daily life associated with the public health emergency (SHARE Corona Survey Release Guide, 2021). Unlike the standard pre-pandemic waves, the Corona Surveys were conducted entirely by telephone, but they retained a crucial feature of the SHARE infrastructure: they were administered to individuals who had already participated in previous SHARE waves. This continuity preserves the panel nature of the data and makes it possible to link pandemic-period outcomes directly to information collected before the onset of COVID-19.

The empirical analysis developed in this thesis combines information from the main SHARE pre-pandemic waves with the Corona Survey modules. It draws on Wave 6, Wave 7, and Wave 8 of the standard SHARE survey, together with Corona Survey 1 and Corona Survey 2**.** Corona Survey 2 provides the core information used to identify respondents who experienced COVID-19 infection and to construct the outcome of interest, namely Long COVID, defined through the presence of persistent symptoms after infection. The same survey also provides information on hospitalization related to COVID-19 and on the calibrated sampling weights used in the empirical analysis. Corona Survey 1 is instead used to complement the reconstruction of respondents' health status by providing updated information on selected chronic conditions during the earlier pandemic stage. The integration of these pandemic modules with the standard pre-pandemic SHARE waves makes it possible to construct a longitudinal dataset in which explanatory variables precede the observed COVID-related outcomes.

More specifically, the pre-pandemic information used in this thesis is drawn from multiple SHARE modules and is combined according to a hierarchical logic. Body mass index, which represents the main explanatory variable inherited from the study by Wilk et al. (2024), is primarily taken from Wave 8 and, when unavailable, from Wave 7, while Wave 6 is used as an additional fallback source to recover observations that would otherwise be lost. A similar strategy is used for chronic conditions, which are reconstructed by combining information from Corona Survey 1 with the health modules of Waves 8, 7, and 6. This approach allows the identification of pre-existing conditions such as heart disease, hypertension, diabetes, and lung disease before the measurement of Long COVID outcomes. In addition, the present study extends the information set used in the baseline framework by incorporating socio-economic and behavioral variables observed before the pandemic, including education, household income, employment status, smoking history, and physical inactivity. These variables are drawn from the relevant SHARE modules in Waves 7 and 8 and are integrated into the final analytical dataset to build a broader pre-pandemic profile of each respondent.

In this respect, the study is inspired by the empirical approach proposed by Wilk et al. (2024), but it should not be interpreted as a strict replication. Rather, that contribution provides the starting point for the analysis developed here. The present thesis preserves the main underlying idea of relating pre-pandemic BMI to the probability of Long COVID among older Europeans, but it deliberately expands the scope of the original framework in two directions. First, it extends the temporal coverage of the pre-pandemic information by including Wave 6 whenever more recent data from Waves 7 or 8 are not available for key variables. Second, and more importantly, it broadens the explanatory structure of the model by adding socio-economic and behavioral dimensions that are not central in the original specification. The aim of this extension is to move beyond a narrower health-based interpretation and to assess whether the risk of Long COVID is also associated with broader forms of vulnerability that were already observable before the pandemic.

This choice is consistent with the wider analytical objective of the thesis. Long COVID is unlikely to depend exclusively on biomedical factors, and a richer empirical framework may help capture the role of social and behavioral inequalities that shape both baseline health and resilience to adverse shocks. For this reason, the use of SHARE is particularly advantageous: its longitudinal and multidimensional structure makes it possible not only to observe respondents before and during the pandemic, but also to combine information on health status with indicators of socio-economic position and health behavior within a single harmonized European dataset. The inclusion of additional controls such as education, income, employment status, smoking, and physical inactivity is therefore not a merely technical modification of the original paper, but part of a broader attempt to enlarge the interpretative scope of the analysis.

Overall, the integration of Waves 6, 7, and 8 with the two Corona Surveys provides a coherent longitudinal framework for studying Long COVID among Europeans aged 50 and over. Earlier waves supply the pre-pandemic demographic, health, socio-economic, and behavioral characteristics of respondents, while the Corona Surveys capture COVID-19 infection and the persistence of symptoms during the pandemic period. The resulting dataset makes it possible to analyze whether body mass index and other pre-existing vulnerabilities are associated with the probability of experiencing persistent post-COVID symptoms in later life. The availability of harmonized panel data across several countries and waves therefore provides a solid empirical basis for the investigation developed in this thesis (Börsch-Supan et al., 2013; SHARE Release Guide, 2022; SHARE Corona Survey Release Guide, 2021).

4.2 Sample Selection: Inclusion Criteria for Middle-aged and Older Adults.

The sample selection procedure follows a sequence of restrictions designed to ensure the consistency of the analytical framework and the availability of the key variables required for the empirical model.

The starting point is the full integrated SHARE dataset, which includes 105,392 individual observations The initial reduction in the number of observations does not reflect only an age-based restriction, but rather the process of linking different SHARE modules. The analysis combines information from the Corona Survey respondent file and the COVID-specific module, and only individuals with valid information in both sources are retained, resulting in a subsample of 47,793 respondents.

The analysis then focuses on individuals who experienced COVID-19. This implies that the analysis is conditional on infection and does not aim to model the probability of contracting COVID-19. This definition allows the identification of infection episodes based on both confirmed tests and self-reported symptoms, accounting for both diagnosed and self-reported cases, which is particularly relevant in the early phases of the pandemic when testing availability was limited. Applying this criterion yields a sample of 4,021 individuals aged 50 and above who experienced COVID-19.

The construction of the dependent variable requires valid information on persistent symptoms following infection. Among individuals who experienced COVID-19, a total of 4,008 respondents reports non-missing information on post-COVID symptoms and can therefore be used to define the long COVID indicator. Restricting the sample to this group ensures that the outcome variable is based on reliable and interpretable information, excluding observations with incomplete or non-informative responses.

The final analytic sample is obtained by further excluding individuals with missing information on the key variables included in the empirical specification. In particular, observations are required to have non-missing values for long COVID status, pre-pandemic BMI, hospitalization due to COVID-19, and age. These restrictions ensure internal consistency of the dataset and allow the empirical model to be estimated on a well-defined and comparable sample. After applying all selection criteria, the final analytic sample consists of 3,806 individuals. This sample defines the maximum core analytic sample based on the key variables required for the baseline specifications; models including additional socioeconomic and behavioral covariates may be estimated on smaller subsamples due to item non-response.

Overall, the sample selection strategy balances the need for data completeness with the objective of maintaining a sufficiently large and representative sample of older individuals who experienced COVID-19. While successive restrictions reduce the number of observations, they are necessary to ensure the validity of the dependent variable and the coherence of the empirical specification.

The procedure described above is summarized in Table 4.1.

Table 4.1 Sample Selection Procedure

Step	**Sample restriction**	**Number of observations**
Initial SHARE dataset	Full integrated dataset	105,392
Linked Corona Survey sample	Individuals aged ≥ 50 and with valid information in both respondent and COVID files	47,793
COVID sample	Individuals aged ≥ 50 who experienced COVID-19 (positive test or symptoms)	4,021
Valid long COVID information	Non-missing persistent symptoms	4,008
Final analytic sample	Non-missing long COVID, BMI, hospitalization, age	3,806

Source: Author's calculations based on SHARE data.

These restrictions are consistent with standard practices in applied micro econometric analysis using survey data, where ensuring comparability and internal consistency of the estimation sample is essential.

4.3 Definition of Variables:

4.3.1 Dependent Variable: Presence of Long COVID Symptoms.

The dependent variable used in the empirical analysis is a binary indicator capturing the presence of persistent symptoms following COVID-19 infection. The variable, labelled *long_covid*, is constructed using data from the SHARE Corona Survey, which collects detailed information on post-infection health outcomes among individuals aged 50 and over across European countries.

The measure is based on a set of survey questions on lingering symptoms experienced after the acute phase of COVID-19. Specifically, respondents are asked whether they suffered from a range of persistent conditions, including fatigue, respiratory problems, loss of taste or smell, headaches, musculoskeletal pain, chest or abdominal pain, gastrointestinal symptoms, confusion, and other long-term effects. In the dataset, these correspond to the indicators *cac120_1* through *cac120_8*, together with *cac120_97*, which captures additional reported symptoms.

The variable *long_covid* is defined only for respondents who experienced COVID-19, as identified by the variable *had_covid*. This indicator equals one for individuals who either reported a positive COVID-19 test or declared having had symptoms attributed to COVID-19, and zero otherwise. Restricting the outcome to this group ensures that the analysis focuses on post-infection symptom persistence rather than unrelated health conditions in the general population.

Conditional on having had COVID-19, *long_covid* is coded as equal to one if the respondent reports at least one persistent symptom among the items listed above. It is coded as zero if the respondent explicitly indicates the absence of any persistent symptoms, corresponding to the response "no symptoms at all" (*cac120_98* = 1). Observations with incomplete or non-informative responses to the symptom questions are treated as missing values and excluded from the regression samples at later stages of the analysis.

In the working sample, 4,021 individuals are classified as having experienced COVID-19. Among these, 2,885 report at least one persistent symptom and are therefore classified as long COVID cases, while 1,123 report no persistent symptoms. A small number of observations (13 individuals) lack sufficient information to determine symptom persistence and are coded as missing on the dependent variable.

Modelling Long COVID as a binary outcome is consistent with the empirical strategy commonly adopted in the literature and allows the analysis to focus on the probability of experiencing persistent symptoms following infection within a logistic regression framework.

4.3.2 Primary Exposure: Pre-pandemic BMI

The main explanatory variable in the empirical analysis is pre-pandemic body mass index (BMI), used as an indicator of individual metabolic risk prior to COVID-19 infection. In SHARE, BMI is defined as weight in kilograms divided by the square of height in meters (kg/m^2) and represents a standard measure of excess body weight widely employed in epidemiological and health economics research (Blüher, 2019).

The variable *bmi_pre* is constructed using the most recent pre-pandemic information available. BMI measured in Wave 8 is used whenever available; when missing, BMI from Wave 7 is adopted; if both are unavailable, BMI from Wave 6 is used as a further fallback. This hierarchical approach ensures that BMI is measured before infection while maximizing sample size. An additional variable, *bmi_time*, records the timing of the BMI observation and is included among the control variables to account for potential differences in measurement timing.

To improve data quality, the BMI distribution is restricted to a plausible range by truncating values below 18 to 18 and values above 45 to 45. In the working dataset, *bmi_pre* is observed for 100,339 individuals, with a mean of 27.07 (SD = 4.63), while 5,053 observations have missing BMI information.

This specification differs from Wilk et al. (2024), who model BMI as a continuous variable, and allows for a more flexible identification of non-linear risk patterns across clinically relevant weight categories. This modelling choice prioritizes interpretability and allows for a more direct comparison across clinically relevant weight groups. Pre-pandemic BMI is divided into three groups: normal weight (BMI 18–25), overweight (BMI >25–30), and obesity (BMI >30–45). The resulting variable, *bmi_cat*, constitutes the main exposure used in the empirical analysis.

This modelling choice allows for a more direct and interpretable comparison across clinically meaningful weight groups and avoids imposing a specific functional form on the relationship between BMI and long COVID risk. It facilitates the identification of potential differences between overweight and obesity, which may be obscured under a continuous specification.

In the final analytic sample of 3,806 individuals, 29.40% are classified as normal weight, 40.73% as overweight, and 29.87% as obese. Most BMI observations originate from Wave 8 (65.03%), followed by Wave 7 (31.45%) and earlier pre-pandemic information (3.52%). By construction, BMI is observed for all individuals included in the regression samples.

4.3.3 Covariates: Control Variables.

In addition to the primary exposure variable, the empirical specification includes a set of control variables capturing demographic characteristics, COVID-19 severity, the timing of the pre-pandemic BMI measure, pre-existing health conditions, and geographical heterogeneity. The inclusion of these covariates is intended to account for observable sources of variation that may simultaneously affect both the likelihood of experiencing persistent post-infection symptoms and individual health vulnerability prior to the pandemic. Controlling for these factors allows the analysis to isolate more accurately the association between pre-pandemic metabolic risk and Long COVID outcomes.

The model first controls for sex and age, two fundamental determinants of both COVID-19 outcomes and long-term health conditions in older populations. Sex is included as a binary indicator equal to one for female respondents. Age is measured in years and enters the specification through three categorical indicators distinguishing individuals aged 50–64, 65–74, and 75 years or older, with the youngest group used as the reference category. This categorization captures non-linear age gradients in infection severity, immune response, and comorbidity profiles that are well documented in ageing populations. In the analytic sample, the mean age is 68.46 years, with 36.44% of individuals aged 50–64, 40.54% aged 65–74, and 23.02% aged 75 or above, while women account for 59.06% of respondents.

In addition to the main exposure variable, the empirical model includes an indicator capturing the timing of the pre-pandemic BMI measurement. Because BMI information is drawn from different SHARE waves, the variable *bmi_time* distinguishes whether the BMI observation originates from Wave 8, Wave 7, or an earlier pre-pandemic source. This variable is constructed by assigning priority to Wave 8, then Wave 7, and finally Wave 6 when more recent information is unavailable. Including this indicator in the regression model allows the analysis to account for potential differences related to the reference year of the anthropometric measurement. In the analytic sample, 65.03% of BMI observations come from Wave 8, 31.45% from Wave 7, and 3.52% from earlier waves.

To capture heterogeneity in the clinical severity of infection, the specification includes a binary indicator for COVID-19-related hospitalization. The variable *hospitalized_covid* is derived from the Corona Survey and equals one for individuals who report having been hospitalized due to COVID-19. It is coded as zero for respondents who report not having been hospitalized, and for cases where hospitalization information is missing but the survey indicates that no hospital treatment occurred. This construction ensures consistency in distinguishing between severe and non-severe infection episodes. Hospitalization serves as a proxy for acute disease severity, which may independently increase the probability of persistent symptoms and therefore confound the relationship between pre-pandemic health status and Long COVID risk. In the analytic sample, 12.74% of respondents were hospitalized due to COVID-19.

The empirical specification further controls for a set of pre-existing health conditions that may influence both susceptibility to severe infection and the persistence of post-COVID symptoms. In particular, the model includes binary indicators for heart disease, hypertension, diabetes, and chronic lung disease. These variables are constructed by combining information from the Corona Survey and from pre-pandemic SHARE waves (Waves 6, 7, and 8), using the most recent available valid observation for everyone. For each condition, the indicator is coded as one if the respondent reports the presence of the condition in any available source, and zero if all available sources consistently indicate its absence. In the analytic sample, 22.81% of individuals report pre-pandemic heart disease, 53.00% hypertension, 16.42% diabetes, and 8.49% chronic lung disease.

Finally, the model accounts for geographical heterogeneity through a categorical variable identifying four European macro-regions: Western, Northern, Eastern, and Southern Europe. This variable is based on grouping countries according to standard European geographical classifications commonly used in SHARE-based analyses and is included in the specification to capture broad, time-invariant differences in healthcare systems, pandemic management strategies, testing intensity, and reporting practices. Compared to a full set of country fixed effects, this approach reduces dimensionality while preserving meaningful cross-country variation. In the analytic sample, 19.86% of respondents are in Western Europe, 20.18% in Northern Europe, 30.22% in Eastern Europe, and 29.74% in Southern Europe.

Taken together, these covariates account for the main demographic, clinical, temporal, and geographical sources of heterogeneity in the older European population. Their inclusion ensures that the estimated relationship between pre-pandemic BMI and the probability of experiencing Long COVID symptoms is not driven by differences in baseline characteristics, disease severity, or institutional context.

This analysis also incorporates a set of additional pre-pandemic socioeconomic and behavioral variables that are not included in Wilk et al. (2024). The inclusion of these variables represents an extension of the original empirical framework and aims to provide a more comprehensive characterization of individual vulnerability. By accounting for socioeconomic status, labor market position, and health-related behaviors prior to the pandemic, the model can capture additional sources of heterogeneity that may influence both exposure to health risks and the persistence of post-COVID symptoms. All variables are constructed using information collected before the onset of the COVID-19 pandemic, ensuring full temporal consistency with the pre-pandemic BMI measure.

Educational attainment is measured using the ISCED 1997 classification available in SHARE. The variable *education* is constructed by prioritizing information from Wave 8 and using Wave 7 as a fallback when more recent data are unavailable. It is then grouped into three categories: low education (ISCED levels 0–2), medium education (levels 3–4), and high education (levels 5–6). This categorization captures broad differences in human capital and health literacy. In the analytic sample, 28.03% of individuals are classified as low educated, 47.29% as medium, and 21.36% as highly educated, with 3.31% missing.

Household income is included as a proxy for economic resources and material living conditions. The variable *income* is defined using total household income from Wave 8, with Wave 7 used as a fallback when necessary. To account for the highly skewed distribution of income, the variable is transformed into categories. Individuals reporting zero income are classified separately, while positive income observations are divided into terciles based on their distribution. This results in four groups: no/zero income, low income, middle income, and high income. In the analytic sample, 27.82% of respondents report no or zero income, while the remaining observations are distributed across the three incomes terciles: missing income information accounts for 3.07% of the sample.

Labor market status is captured using the SHARE employment variable describing current job situation. This variable is derived from Wave 8 information, with Wave 7 used as a fallback when necessary, and is recoded into four mutually exclusive categories: retired, employed, unemployed, and other inactive, the latter including individuals not in the labor force for reasons other than retirement, such as homemakers, disabled individuals, and other non-employed statuses. In the analytic sample, 55.41% of individuals are retired, 29.27% are employed, 1.89% are unemployed, and 8.30% are classified as other inactive, with 5.12% missing.

Two behavioral variables are included to capture pre-pandemic health-related behaviors. The first, *smoked_ever*, indicates whether the respondent has ever smoked daily and is constructed using Wave 8 information with fallback to Wave 7. The second, *physically_inactive*, captures whether the respondent reports low levels of physical activity prior to the pandemic, again using Wave 8 data with fallback to Wave 7. Both variables are defined as binary indicators. In the analytic sample, 27.04% of individuals report having ever smoked daily, while 42.77% report never having done so. Similarly, 6.57% are classified as physically inactive, while 63.24% are not. Both variables exhibit a substantial share of missing observations (approximately 30.2%), reflecting incomplete information across survey waves.

The inclusion of these additional covariates represents a key extension of the baseline empirical framework. A relevant feature of these variables concerns the presence of missing information, particularly for behavioral indicators such as smoking history and physical inactivity. This reflects incomplete responses across survey waves and the fact that these variables are not always observed for all individuals in the pre-pandemic period. By accounting for socioeconomic status and health-related behaviors, the analysis captures broader dimensions of individual vulnerability that are not considered in the reference study. At the same time, the presence of missing information for some variables leads to a reduction in the estimation sample when they are jointly included, highlighting the trade-off between model richness, data availability and sample size.

For clarity, Table 4.2 summarizes the main variables used in the empirical analysis, along with their definitions and data sources.

Table 4.2 Variable Definitions

Variable	Description	Source
Long Covid	Binary indicator equal to 1 if the respondent reports at least one persistent symptom after COVID-19 infection	SHARE Corona Survey (Wave 9)
BMI	Pre-pandemic body mass index categorized into normal weight, overweight, and obesity	SHARE Waves 6–8
Female	Binary indicator equal to 1 for female respondents	SHARE
Age	Indicators for age groups (50–64, 65–74, 75+)	SHARE

Hospitalization	Indicator equal to 1 if the respondent was hospitalized due to COVID-19	SHARE Corona Survey
Heart disease	Indicator for pre-pandemic cardiovascular conditions	SHARE Waves 6–8 + Corona Survey 1
Hypertension	Indicator for diagnosed hypertension	SHARE Waves 6–8 + Corona Survey 1
Diabetes	Indicator for diagnosed diabetes	SHARE Waves 6–8 + Corona Survey 1
Lung disease	Indicator for chronic respiratory conditions	SHARE Waves 6–8 + Corona Survey 1
BMI timing	Indicator for the wave from which BMI is obtained	SHARE
Region	European macro-area classification (West, North, East, South)	SHARE
Education level	Categorical variable based on ISCED classification (low, medium, high)	SHARE Waves 7–8
Income category	Household income categorized into zero and terciles of positive income	SHARE Waves 7–8
Employment status	Categorical variable: retired, employed, unemployed, other inactive	SHARE Waves 7–8
Smoking	Indicator for having ever smoked daily (pre-pandemic)	SHARE Waves 7–8
Physical inactivity	Indicator for low physical activity (pre-pandemic)	SHARE Waves 7–8

4.4 Empirical Strategy: Logistic Regression Framework

The empirical analysis adopts a logistic regression framework to examine the relationship between pre-pandemic individual characteristics and the probability of experiencing long COVID. The analysis is conducted on individuals who experienced COVID-19 infection, and therefore estimates the probability of developing long COVID conditional on infection. This modelling approach is appropriate given the binary nature of the dependent variable and is standard in applied micro-econometric analysis of discrete outcomes. The logistic specification ensures that predicted probabilities remain bounded between zero and one and allows for flexible relationships between covariates and the outcome of interest.

Formally, the probability that individual experiences long COVID is specified as:

(P ($LONG\_Covid_i$= 1) = F (β_0+ $\beta_1 BMI_i$ + $X_i\gamma$))

where (F(.)) denotes the logistic cumulative distribution function, (BMI_i) represents pre-pandemic body mass index categorized into normal weight, overweight, and obesity, and (X_i) is a vector of control variables capturing demographic characteristics, baseline health conditions, COVID-19 severity, and socio-economic and behavioral factors observed prior to the pandemic.

The empirical strategy is inspired by Wilk et al. (2024), but it does not constitute a strict replication. While the reference study models BMI as a continuous variable and focuses primarily on biomedical determinants, the present analysis adopts a categorical specification of BMI and extends the empirical framework by incorporating a broader set of pre-pandemic socio-economic and behavioral characteristics. This extension reflects the hypothesis that the risk of developing long COVID may depend not only on metabolic conditions, but also on wider dimensions of individual vulnerability.

The estimation strategy follows a progressive modelling approach. The analysis starts from a parsimonious specification including BMI and basic demographic characteristics, and then gradually incorporates additional sets of covariates. This approach allows the assessment of how the estimated association between pre-pandemic BMI and long

COVID evolves when controlling for different dimensions of individual heterogeneity, including health status, disease severity, and socio-economic and behavioral factors.

Estimated coefficients are interpreted in terms of changes in the log-odds of experiencing long COVID. To facilitate interpretation, the analysis also reports predicted probabilities and marginal comparisons derived from the estimated models, allowing the main relationships to be expressed in more intuitive terms.

All models are estimated using standard maximum likelihood logistic regression. The models are estimated without the use of sampling weights. This choice reflects the focus of the analysis on estimating conditional associations within the analytic sample rather than producing population-representative estimates. Accordingly, the results should be interpreted as sample-based conditional associations rather than population-weighted estimates for the broader European population.

In conclusion, this empirical framework provides a coherent basis for evaluating the association between pre-pandemic metabolic risk and long COVID, while allowing for a flexible and progressive assessment of the role of different sources of individual heterogeneity. The results of the empirical analysis are discussed in the following chapter.

Chapter V: Empirical Results

5.1 Descriptive Statistics of the European Sample.

Descriptive statistics are presented for individuals who experienced COVID-19, providing an initial overview of the distribution of Long COVID across key individual, socioeconomic, and regional characteristics. The analysis focuses on patterns in the occurrence of persistent symptoms in relation to pre-pandemic conditions, with particular attention to differences across body mass index categories, demographic factors, and indicators of disease severity.

Within the COVID-19 sample, a substantial share of individuals reports persistent symptoms following infection. Approximately 71.7% of respondents experienced at least one long-term symptom, while about 27.9% reported no persistent effects. The proportion of missing information is negligible.

The distribution of the number of symptoms further highlights the extent of post-infection health issues. On average, individuals report slightly more than two persistent symptoms, with a median of two and an interquartile range from zero to four. While a non-negligible group reports no lingering effects, a significant fraction of individuals experiences multiple concurrent symptoms, with values reaching up to eight or more in the upper tail of the distribution.

Overall, these descriptive patterns indicate that long COVID represents a widespread condition within the infected population, characterized not only by high prevalence but also by considerable heterogeneity in symptom intensity.

A clear gradient emerges when examining the prevalence of Long COVID across pre-pandemic BMI categories. The share of individuals reporting at least one persistent symptom increases steadily with body mass index, rising from 68.4% among individuals with normal weight to 72.3% among those who are overweight, and reaching 76.4% among those classified as obese.

These differences are not only substantial but also potentially relevant from an economic perspective, suggesting a positive association between higher BMI and the prevalence of Long COVID. The pattern is consistent across both tabulations and mean comparisons, with the estimated probability of Long COVID increasing monotonically across BMI categories.

The descriptive evidence points to a clear and increasing relationship between pre-pandemic body mass index and the occurrence of persistent symptoms, suggesting a positive association between higher BMI and the prevalence of Long COVID.

Fig. 5.1: Long COVID prevalence by BMI category

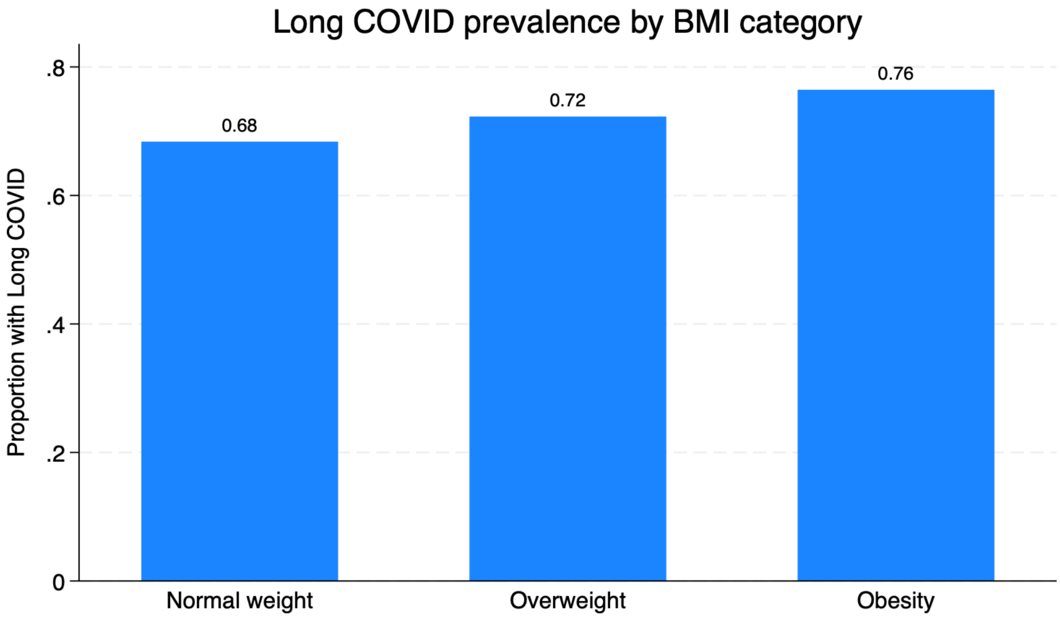
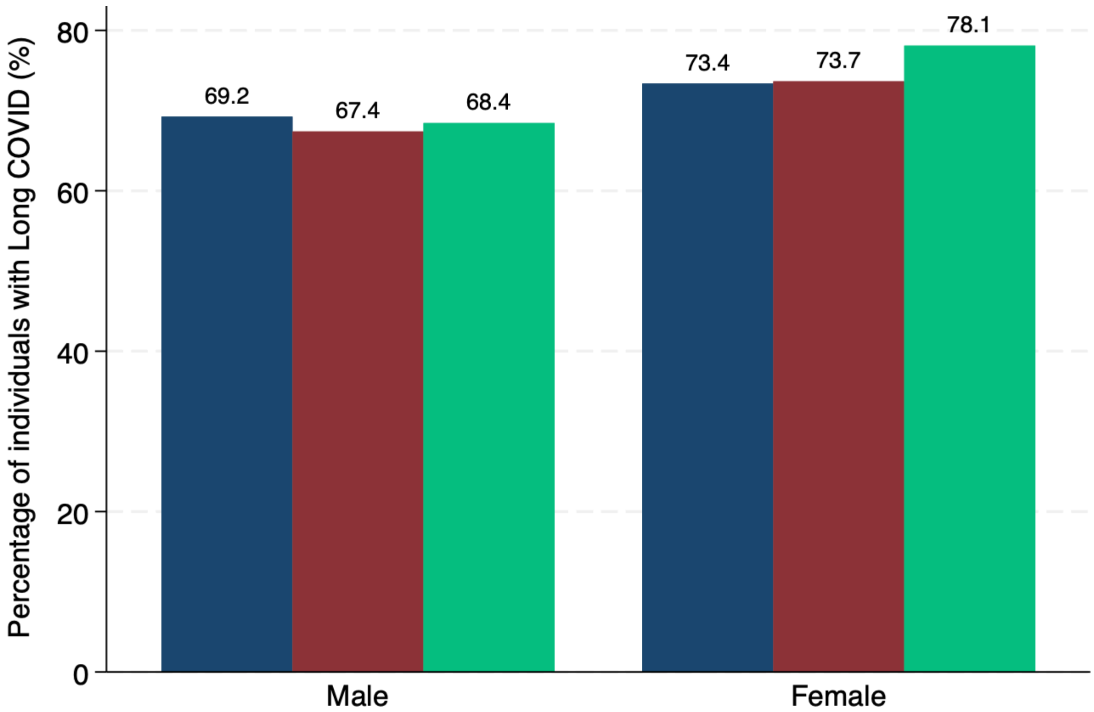


Figure 5.1 provides a direct descriptive illustration of the main relationship investigated in this thesis, showing that the prevalence of Long COVID increases across pre-pandemic BMI categories and reaches its highest level among individuals with obesity.

Differences in the prevalence of Long COVID also emerge across key individual characteristics. A clear gender gap is observed, with women reporting a higher probability of persistent symptoms compared to men. Approximately 74.5% of female respondents experience Long COVID, compared to 68.3% among males, indicating a substantial disparity across sexes. This pattern is illustrated in Figure 5.2 divided into age groups (with blue indicating the youngest group, red the intermediate group, and green individuals aged 75 and over).

Fig. 5.2: Long COVID by gender and age group



Notes:3 groups: age 50–64 (blue), age 65–74 (red), age75+ (green)

By contrast, differences across age groups appear more limited. The prevalence of Long COVID remains relatively stable across categories, ranging from about 71.0% among individuals aged 65–74 to approximately 74.0% among those aged 75 and over, suggesting only modest variation along the age dimension.

A much stronger pattern is observed with respect to disease severity, proxied by COVID-19 hospitalization. Individuals who were hospitalized exhibit a markedly higher prevalence of Long COVID, with nearly 88.6% reporting persistent symptoms, compared to about 69.6% among those who were not hospitalized. This substantial gap highlights the strong association between severe acute infection and the likelihood of long-term health consequences.

Some differences in the prevalence of Long COVID are also observed across socioeconomic characteristics, although the patterns are less clear-cut compared to those associated with BMI or disease severity.

With respect to education, individuals with higher educational attainment exhibit a lower probability of reporting persistent symptoms. Approximately 64.8% of highly educated respondents experience Long COVID, compared to about 74–75% among those with low or medium levels of education, suggesting a potential gradient along the education dimension.

A similar pattern emerges across income groups. Individuals in the highest income category display a substantially lower prevalence of Long COVID (around 57.2%), while those in lower income categories report higher rates, reaching up to 80.0% among low-income respondents. This pattern suggests that socioeconomic disadvantage may be correlated with a higher likelihood of persistent symptoms.

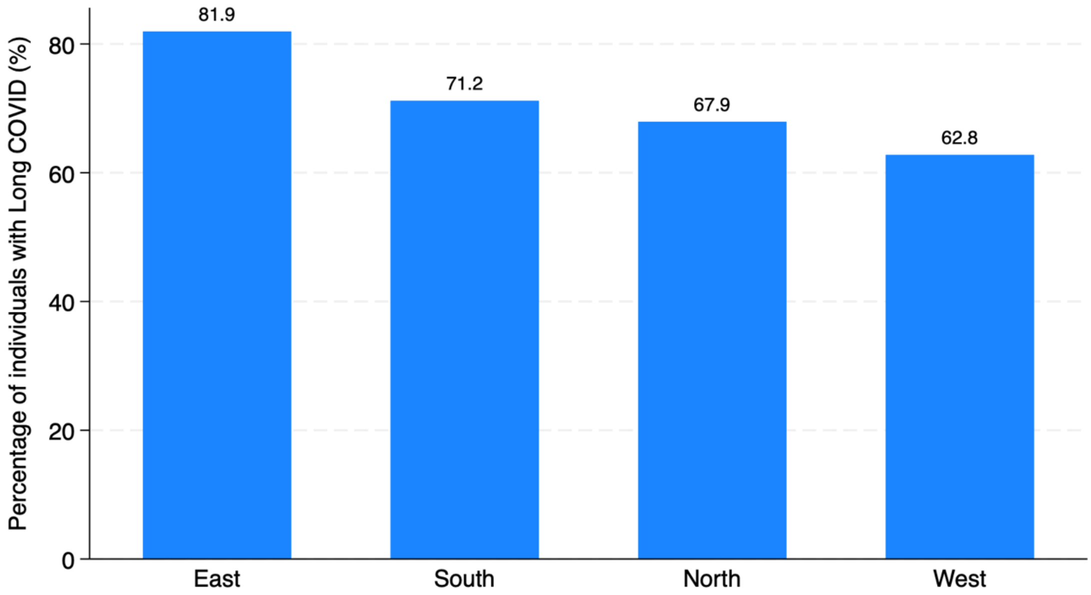
By contrast, differences across employment status appear more limited, with relatively similar prevalence rates observed across groups.

Overall, while some socioeconomic gradients are visible, these patterns are less pronounced and more heterogeneous than those observed for body mass index and hospitalization.

Differences in the prevalence of Long COVID are also evident across European regions. A clear gradient emerges, with the highest prevalence observed in Eastern Europe, where approximately 82.0% of individuals report persistent symptoms. This is followed by Southern Europe (around 71.1%) and Northern Europe (67.8%), while the lowest prevalence is recorded in Western Europe (62.8%).

Results illustrated in Figure 5.3, which shows a clear gradient in Long COVID prevalence across European regions.

Figure 5.3: Long COVID prevalence across European regions



These patterns suggest substantial geographic heterogeneity in the occurrence of Long COVID across Europe. Additional descriptive evidence at the country level indicates considerable variation within regions, pointing to the presence of both between- and within-region differences.

The regional distribution highlights that contextual factors may play a role in shaping post-COVID health outcomes, although these differences will be further investigated in a multivariate framework in the following sections.

5.2 Logistic Regression Estimates

This section presents the results of logistic regression models estimating the association between pre-pandemic body mass index (BMI) and the probability of experiencing Long COVID. The specifications are progressively extended to assess the robustness of the estimated relationships across demographic, health, and socioeconomic dimensions.

Table 5.4 reports the odds ratios and standard errors from the logistic regression models estimating the association between pre-pandemic BMI and the probability of Long COVID.

Tab 5.4: Logistic regression estimates of Long COVID (odds ratios and standard errors)

Variables	Model 1	Model 2	Model 3	Model 4
Overweight	1.25** 0.11	1.33*** 0.14	1.25** 0.13	1.19 0.13
Obesity	1.55*** 0.14	1.51*** 0.18	1.37*** 0.16	1.29** 0.16
Female	1.41*** 0.10	1.51*** 0.13	1.36*** 0.13	1.44*** 0.14
Age 65-74	0.95 0.07	0.85 0.08	0.81 0.11	0.77 0.10
Age 75+	1.11 0.10	0.83 0.10	0.93 0.15	0.74 0.13
Hospitalized	-	3.37*** 0.60	-	3.44*** 0.63
Heart disease	-	1.41** 0.16	-	1.32** 0.15
Hypertension	-	1.26** 0.12	-	1.14 0.11
Diabetes	-	0.97 0.13	-	0.89 0.12
Lung disease	-	1.15 0.19	-	1.20 0.20
Higher education	-	-	0.73**	0.78

			0.09	0.10
Higher Income	-	-	0.47 0.21	0.47 0.21
Employment status	-	-	0.87 0.12	0.87 0.13
Eastern Europe	-	-	1.46** 0.24	1.42** 0.23
Pseudo R^2	low	high	high	0.073

Tab 5.4b: Logistic regression estimates of Long COVID per country (odds ratios and standard errors)

COUNTRY	**MODEL 4**
Germany	0.62 0.24
Sweden	0.46** 0.17
Netherlands	0.38* 0.19
Spain	0.34*** 0.12
Italy	0.82 0.28
France	0.79 0.27

Denmark	0.97 0.39
Greece	1.48 0.57
Switzerland	0.89 0.30
Belgium	1.26 0.42
Czech Republic	1.84* 0.61
Poland	3.93*** 1.40
Luxembourg	1.96* 0.77
Hungary	2.60* 1.40
Slovenia	1.67 0.54
Estonia	2.19** 0.76
Croatia	2.76** 1.10
Lithuania	1.32

	<u>0.49</u>
Bulgaria	2.94** <u>1.29</u>
Finland	0.33*** <u>0.13</u>
Latvia	2.16 <u>1.07</u>
Malta	0.52 <u>0.23</u>
Romania	0.73 <u>0.28</u>
Slovakia	1.52 <u>0.59</u>

*Notes: Odds ratios reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; "n.s." indicates non-statistically significant coefficients.*

The main patterns emerging from Table 5.4a and Tab5.4b are discussed below.

The baseline specification (Model 1) includes BMI categories and basic demographic characteristics, namely gender and age groups. In this specification, BMI is positively and significantly associated with the likelihood of Long COVID. Relative to normal-weight individuals, those classified as overweight exhibit higher odds of persistent symptoms (OR = 1.25, $p < 0.05$), while obesity is associated with a higher odd of experiencing Long COVID (OR = 1.55, $p < 0.01$). This pattern indicates a clear gradient in which higher BMI levels correspond to a greater probability of long-term post-infection conditions.

Gender differences emerge as a consistent and statistically significant factor. Women display substantially higher odds of Long COVID compared to men ($OR = 1.41$, $p < 0.01$), suggesting the presence of systematic differences in vulnerability or symptom reporting across sexes.

By contrast, age does not appear to be a significant predictor in this specification. Neither the 65–74 nor the 75+ age groups differ significantly from the reference category, indicating that, conditional on infection, the probability of Long COVID is not strongly stratified by age in the baseline model.

Overall, the model is jointly significant ($p < 0.01$), although the relatively low pseudo-R^2 suggests that BMI and basic demographic characteristics alone capture only a limited share of the variation in Long COVID outcomes. This motivates the inclusion of additional covariates in subsequent specifications.

To account for clinical factors that may influence the development of persistent symptoms, the baseline specification is extended by including indicators of COVID-19 severity and pre-existing health conditions. In particular, Model 2 adds hospitalization status and a set of chronic diseases, allowing for a more comprehensive assessment of whether the association between BMI and Long COVID reflects underlying health vulnerabilities.

The inclusion of health-related controls does not attenuate the association between BMI and Long COVID; on the contrary, the estimated effects remain stable and statistically significant. Overweight individuals exhibit higher odds of persistent symptoms ($OR = 1.33$, $p < 0.01$), while obesity is associated with a substantial increase in risk ($OR = 1.51$, $p < 0.01$). This stability suggests that BMI appears to capture a distinct dimension of vulnerability that is not fully accounted for by the included controls.

The most pronounced effect emerges for hospitalization, which is strongly associated with Long COVID. Individuals who were hospitalized display odds of persistent symptoms more than three times higher than those who were not ($OR = 3.37$, $p < 0.01$), highlighting the strong association between disease severity and long-term outcomes.

Among pre-existing conditions, heart disease and hypertension are positively and significantly associated with Long COVID (OR = 1.41 and OR = 1.26, respectively), indicating that cardiovascular risk factors contribute to prolonged recovery. In contrast, diabetes and lung disease do not show statistically significant effects in this specification.

Gender differences remain robust, with women exhibiting a significantly higher probability of Long COVID (OR = 1.51, $p < 0.01$), while age continues to be non-significant. Overall, the increase in pseudo-R^2 compared to the baseline model indicates an improved explanatory power, confirming the relevance of clinical factors in understanding the persistence of symptoms.

To further examine whether the observed relationship between BMI and Long COVID is influenced by socioeconomic conditions, the specification is extended by introducing regional and socioeconomic controls. Model 3 includes education, income, employment status, and broad geographic regions, while excluding clinical variables, allowing for a focused assessment of socioeconomic gradients.

The estimated association between BMI and Long COVID remains positive and statistically significant. Overweight individuals show higher odds of persistent symptoms (OR = 1.25, $p < 0.05$), and obesity is associated with an even stronger increase in risk (OR = 1.37, $p < 0.01$). The persistence of these effects indicates that socioeconomic differences do not account for the BMI gradient observed in previous specifications.

Gender differences remain stable, with women exhibiting higher odds of Long COVID (OR = 1.36, $p < 0.01$), while no significant differences emerge across age groups. These results reinforce the limited role of basic demographic factors in explaining heterogeneity in long-term outcomes.

Among socioeconomic variables, evidence of a systematic gradient is limited. Higher education is associated with a lower probability of Long COVID (OR = 0.73, $p < 0.05$), suggesting a potential protective effect, while income and employment status do not display statistically significant associations. This pattern indicates that, conditional on infection, socioeconomic status plays a relatively minor role compared to individual health characteristics.

Regional differences emerge more clearly. Individuals residing in Eastern Europe exhibit significantly higher odds of Long COVID (OR = 1.46, $p < 0.05$), pointing to potential heterogeneity in health systems, reporting behavior, or underlying population health.

Overall, the increase in explanatory power relative to the baseline model suggests that socioeconomic and contextual factors contribute to understanding Long COVID, although their role remains secondary compared to clinical determinants.

The final specification (Model 4) integrates the full set of covariates, combining demographic characteristics, clinical factors, socioeconomic variables, behavioral indicators, and regional controls. This comprehensive model provides the most complete assessment of the relationship between BMI and Long COVID, minimizing potential omitted variable bias and allowing for a more precise assessment of conditional associations.

In this specification, the association between BMI and Long COVID remains present but becomes more selective. The coefficient for overweight is no longer statistically significant (OR = 1.19, $p = 0.107$), whereas obesity retains a positive and statistically significant effect (OR = 1.29, $p < 0.05$). This pattern suggests that the association between BMI and Long COVID appears to be concentrated at higher levels, suggesting that the association becomes more concentrated in the obesity category.

The effect of gender remains stable and precisely estimated. Women exhibit significantly higher odds of Long COVID (OR = 1.44, $p < 0.01$), confirming that the gender gap persists even after accounting for a wide range of confounding factors.

Clinical severity continues to be the dominant predictor. Hospitalization is associated with a more than threefold increase in the likelihood of persistent symptoms (OR = 3.44, $p < 0.01$), reinforcing the interpretation that the long-term burden of COVID-19 is strongly linked to the intensity of the initial infection.

Among pre-existing conditions, heart disease remains positively associated with Long COVID (OR = 1.32, $p < 0.05$), while the effect of hypertension becomes statistically insignificant once the full set of controls is included. Diabetes and lung disease continue to show no detectable association, indicating that not all comorbidities contribute equally to long-term outcomes.

Socioeconomic and behavioral variables display limited explanatory power in the full specification. Higher education and income are associated with lower odds of Long COVID, although these effects do not reach conventional levels of statistical significance. Employment status, smoking behavior, and physical inactivity are likewise not significant predictors, suggesting that their influence is largely mediated by health status and disease severity.

Regional heterogeneity remains evident. Individuals in Eastern Europe display higher odds of Long COVID (OR = 1.42, $p < 0.05$), pointing to persistent structural differences that are not fully captured by individual-level characteristics.

Model 4 shows a substantial improvement in explanatory power (pseudo-R^2 = 0.073), indicating that the combined inclusion of clinical and socioeconomic factors enhances the ability to account for variation in Long COVID outcomes. Taken together, this specification provides the most comprehensive specification for interpreting the results and serves as the reference model for subsequent analysis.

Across all specifications, a consistent positive association emerges between pre-pandemic BMI and the probability of experiencing Long COVID. While both overweight and obesity are significant in simpler models, the effect becomes more concentrated in the obesity category once the full set of controls is introduced, indicating a non-uniform gradient across BMI levels.

At the same time, several patterns remain stable. The coefficient for female is persistently positive and statistically significant across all models, pointing to a robust gender gap. Age does not display a systematic relationship with Long COVID in any specification. Clinical severity, proxied by hospitalization, represents the strongest determinant, with effect sizes consistently exceeding a threefold increase in risk. Among comorbidities, cardiovascular conditions show a moderate but significant association.

Socioeconomic and behavioral variables contribute only marginally to explaining variation in Long COVID outcomes once clinical factors are accounted for, while regional differences, particularly in Eastern Europe, remain statistically significant.

The results indicate that the association between BMI and Long COVID is robust to the inclusion of a wide set of controls, while varying in intensity across BMI categories. The relationship operates within a broader structure shaped primarily by clinical severity and demographic factors. These findings provide a solid empirical basis for the heterogeneity analysis presented in the following section.

5.3 Heterogeneity Analysis

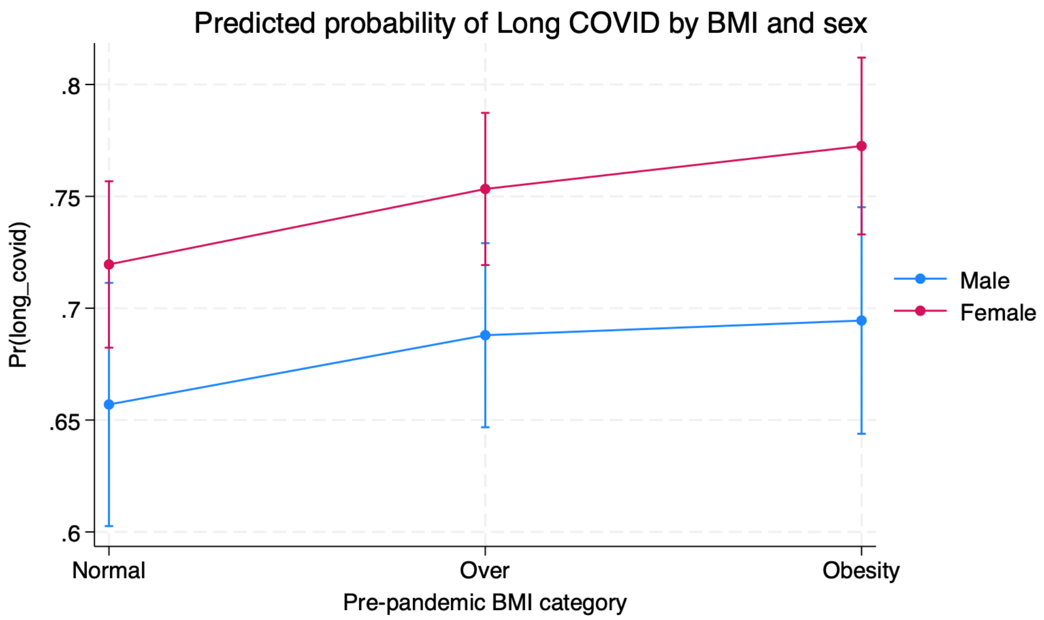
This section explores potential heterogeneity in the relationship between pre-pandemic BMI and the probability of experiencing Long COVID. Building on the results of the previous section, the analysis focuses on whether the estimated association varies across key dimensions that emerged as particularly relevant in the main specification. BMI is specified in categorical form in this analysis. This allows for comparisons across discrete weight groups but does not capture continuous non-linear patterns within the BMI distribution.

In particular, heterogeneity is examined with respect to gender and COVID-19 severity, proxied by hospitalization status. To this end, interaction terms between BMI categories and selected variables are introduced within the full model specification. In addition, predicted probabilities are computed using marginal effects in order to provide a more intuitive interpretation of the estimated relationships across groups.

To assess whether the association between BMI and Long COVID differs across sexes, interaction terms between BMI categories and gender are introduced within the full model specification. The estimated interaction coefficients are not statistically significant. Neither the interaction for overweight ($OR = 1.03$, $p = 0.876$) nor that for obesity ($OR = 1.12$, $p = 0.635$) shows evidence of differential effects. This is further confirmed by the joint test of interaction terms, which fails to reject the null hypothesis of no heterogeneity ($\chi^2 = 0.24$, $p = 0.889$).

Although the interaction terms are not statistically significant, predicted probabilities help illustrate the relationship.

Fig. 5.5: Predicted probability of Long COVID by BMI category and sex



As shown in Figure 5.5, the probability of experiencing Long COVID increases with BMI for both males and females. For normal-weight individuals, the estimated probability is approximately 65.7% for men and 71.9% for women. This difference persists across BMI categories, reaching 68.8% versus 75.3% among overweight individuals and 69.5% versus 77.3% among those classified as obese.

The graphical evidence indicates that the increase in risk across BMI categories follows a similar pattern for both sexes. The lines remain approximately parallel, with no clear divergence as BMI increases. This suggests that BMI affects the likelihood of persistent symptoms in a comparable manner for males and females, while gender differences operate primarily through a higher baseline probability among women.

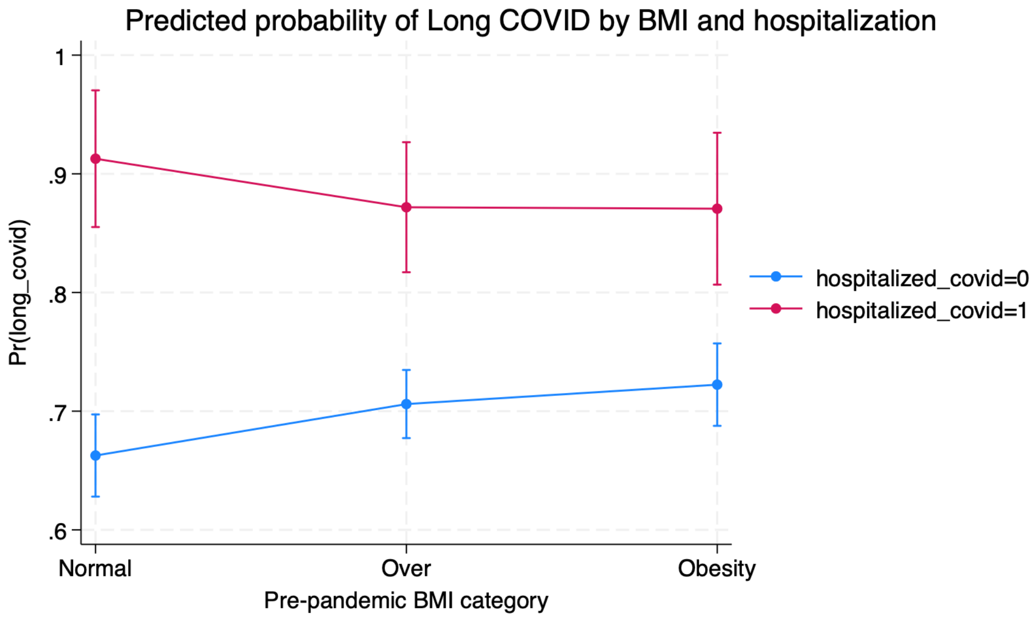
To further investigate the presence of potential heterogeneity, additional checks focus on the direction and magnitude of predicted probabilities across groups, even in the absence of statistical significance. The analysis does not provide statistical evidence of heterogeneity in the BMI–Long COVID relationship across gender. Differences between sexes appear to operate primarily through level effects rather than differences in the slope of the relationship.

This finding differs from Wilk et al. (2024), who identify pronounced sex-specific non-linearities in the relationship between BMI and Long COVID. This difference likely reflects the different functional form imposed on BMI. While their analysis treats BMI as a continuous variable and includes quadratic terms, the present study relies on categorical BMI indicators. This specification captures average differences across weight groups but may smooth out non-linear patterns within categories.

In order to evaluate whether the association between BMI and Long COVID varies with the severity of the acute infection, an interaction between BMI categories and COVID-19 hospitalization is introduced in the full model specification. The estimated interaction terms are negative but not statistically significant. In particular, the interaction for overweight is not significant (OR = 0.52, $p = 0.162$), and the same holds for obesity (OR = 0.47, $p = 0.130$). The joint test confirms the absence of statistically significant heterogeneity in the BMI–Long COVID relationship across hospitalization status ($\chi^2 = 2.61$, $p = 0.272$).

Despite the lack of significant interaction effects, predicted probabilities reveal a marked difference in levels between hospitalized and non-hospitalized individuals.

Figure 5.6: Predicted probability of Long COVID by BMI category and hospitalization status



As shown in Figure 5.6, among respondents who were not hospitalized, the estimated probability of Long COVID rises from 66.3% for normal-weight individuals to 70.6% for those who are overweight and 72.2% for those with obesity. Among hospitalized individuals, the corresponding predicted probabilities are substantially higher, at 91.3%, 87.2%, and 87.1%, respectively.

The graphical evidence therefore indicates that hospitalization shifts the overall level of risk upward across all BMI categories but does not substantially alter the shape of the relationship between BMI and Long COVID. In other words, COVID-19 severity appears to operate primarily as a level effect rather than as a modifier of the BMI gradient. While BMI remains associated with a higher probability of persistent symptoms among non-hospitalized individuals, hospitalized respondents display uniformly elevated predicted probabilities regardless of weight category.

In conclusion, this result reinforces the interpretation that hospitalization primarily affects the baseline risk of Long COVID, rather than altering the role of pre-pandemic BMI.

Building on the previous results, additional heterogeneity analyses were conducted to assess whether the BMI–Long COVID relationship differs across other dimensions of individual vulnerability. In particular, the analysis considers whether the association varies by baseline health conditions and behavioral characteristics, addressing the question of whether pre-existing vulnerability amplifies or modifies the effect of BMI on long-term outcomes.

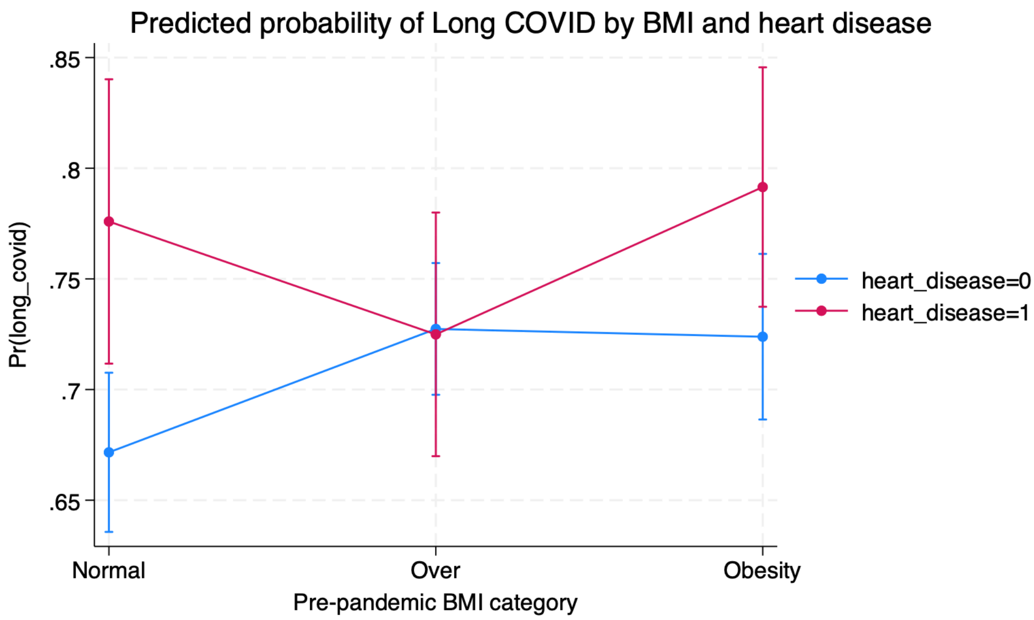
The results do not provide evidence of statistically significant heterogeneity along these dimensions. The interaction between BMI categories and age groups is not significant ($\chi^2 = 1.71$, $p = 0.789$), indicating that the association between BMI and Long COVID does not vary systematically across older age categories. Similarly, no significant interaction emerges with physical inactivity ($\chi^2 = 0.83$, $p = 0.659$), suggesting that lifestyle-related vulnerability does not alter the BMI gradient.

To further investigate baseline health vulnerability, an interaction between BMI and the presence of any chronic condition is introduced. The joint test again fails to reject the null hypothesis of no heterogeneity ($\chi^2 = 1.06$, $p = 0.589$), indicating that the relationship between BMI and Long COVID does not differ significantly between individuals with and without pre-existing conditions.

These findings suggest that the effect of BMI on the probability of experiencing Long COVID is relatively stable across different dimensions of individual vulnerability. The absence of significant interaction effects indicates that BMI appears to be associated with Long COVID in a relatively stable way across individuals.

As an additional robustness check, heterogeneity was also explored with respect to pre-existing cardiovascular conditions. The interaction terms provide limited and only partial evidence of differential effects. In particular, the interaction between overweight and heart disease is statistically significant (OR = 0.56, $p < 0.05$), while the interaction for obesity is not significant. The joint test of interaction terms is only marginally significant ($\chi^2 = 5.05$, $p = 0.080$), suggesting weak overall evidence of heterogeneity. The statistical significance is marginal, and therefore this result should be interpreted with caution.

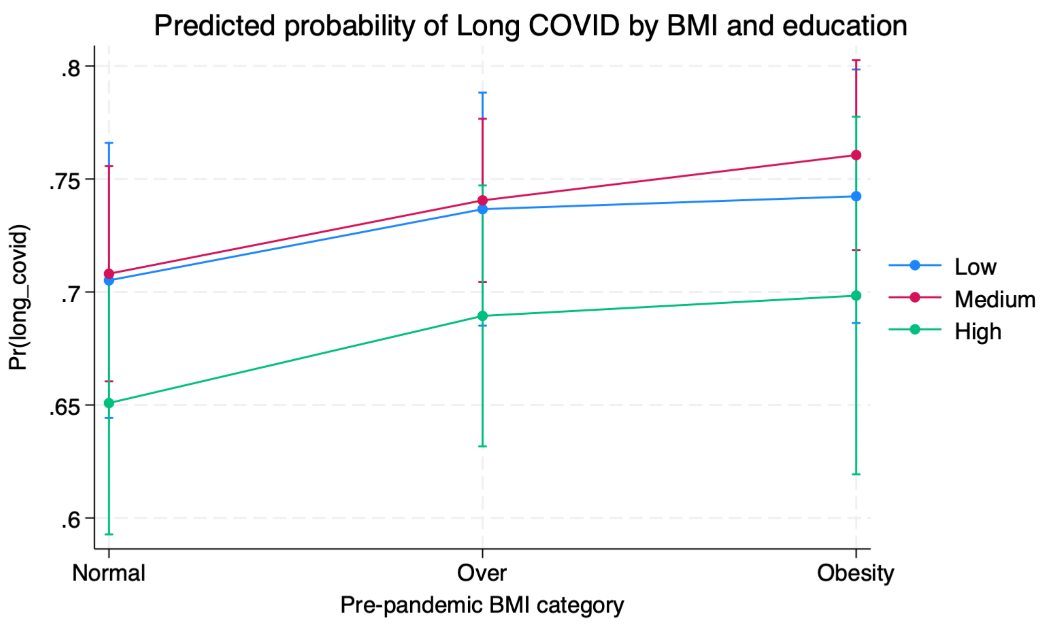
Figure 5.7: Predicted probability of Long COVID by BMI category and heart disease status



Among individuals without heart disease, the probability of Long COVID increases with BMI, whereas for those with heart disease the relationship appears less monotonic, with similar probabilities across BMI categories. This suggests that the role of BMI may be less pronounced among individuals with pre-existing cardiovascular conditions, although the evidence remains limited.

To further assess whether the BMI–Long COVID relationship varies across socioeconomic strata, an interaction between BMI categories and educational attainment is introduced in Figure 5.4. This specification examines whether differences in knowledge, health behaviors, or access to resources modify the association between body weight and long-term outcomes.

Figure 5.8 – Predicted Probability of Long COVID by BMI and Education



The joint test of the interaction terms does not indicate statistically significant heterogeneity ($\chi^2 = 0.15$, $p = 0.997$), suggesting that the effect of BMI on Long COVID is largely invariant across education levels. The absence of significant interaction coefficients confirms that the BMI gradient observed in previous models is not driven by differences in educational attainment.

Predicted probabilities, however, reveal a consistent level difference across groups. Individuals with higher education display systematically lower probabilities of Long COVID across all BMI categories, while those with low and medium education exhibit higher values (Figure 5.8). For instance, among individuals with normal weight, the predicted probability is approximately 0.65 for the highly educated, compared to about 0.71 for lower education groups. A similar ordering persists for overweight and obese individuals.

The graphical evidence supports this interpretation: the curves corresponding to different education levels remain approximately parallel across BMI categories, indicating that education shifts the baseline probability of Long COVID without altering the slope of the BMI gradient.

This suggests that education is associated with differences in the overall likelihood of persistent symptoms but does not significantly modify the marginal effect of BMI. The relationship between body weight and Long COVID therefore appears structurally stable across socioeconomic strata defined by educational attainment.

Taken together, the results provide limited evidence that the effect of BMI on Long COVID varies substantially across subgroups. Most covariates influence the baseline probability of experiencing persistent symptoms, but do not significantly alter the association between BMI and Long COVID.

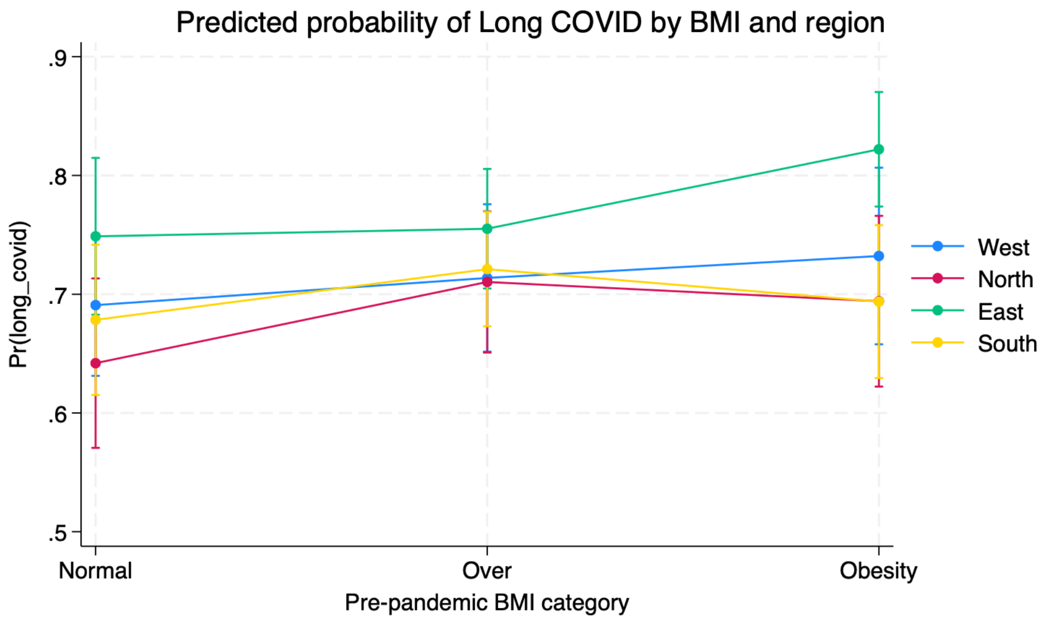
These findings suggest that, within the categorical specification adopted in this study, BMI is associated with Long COVID in a relatively stable manner across individuals. At the same time, the absence of strong interaction effects highlights how modelling choices may shape the detection of heterogeneity. In contrast to studies using continuous and non-linear specifications of BMI, the present approach may not fully capture complex patterns within the BMI distribution. While the estimated patterns provide partial support for the existing literature, the non-linear relationship for males is less clearly identified in this specification.

These findings suggest that the relationship between metabolic risk and post-viral health outcomes may differ across population groups, highlighting the importance of considering heterogeneity when evaluating long COVID risk factors.

5.4 Regional and Country-Level Differences

The analysis is extended to explore whether the association between pre-pandemic BMI and long COVID varies across geographic contexts. Differences in healthcare systems, pandemic management, and pre-existing public health conditions may influence both infection severity and post-acute outcomes. In particular, cross-country variation in obesity prevalence and related health policies provides a relevant framework for interpreting heterogeneous patterns in long COVID risk.

Fig. 5.9 Regional Differences

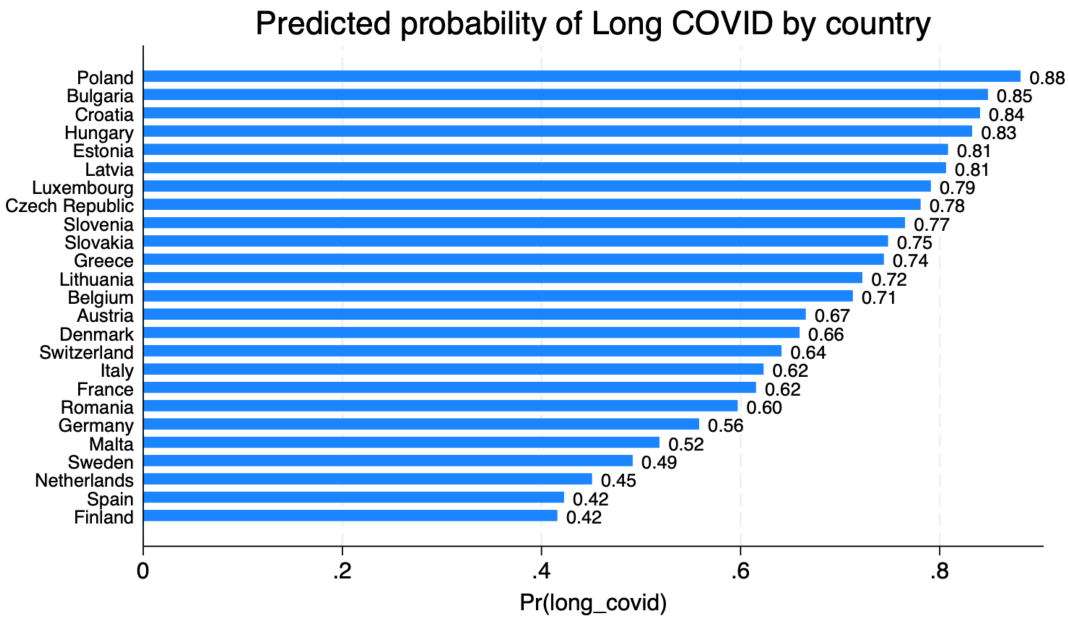


The interaction between BMI and macro-regional classification (Figure 5.9) does not reveal statistically significant heterogeneity ($\chi^2 = 4.44$, $p = 0.618$). The estimated probabilities show a consistent positive gradient across BMI categories in all regions, with higher values observed for overweight and obese individuals relative to those with normal weight.

Despite the absence of statistically significant interaction effects, descriptive differences across regions are evident. Eastern Europe displays systematically higher predicted probabilities of long COVID across all BMI categories, while Northern Europe exhibits comparatively lower levels. Western and Southern regions lie in between, with similar trajectories across BMI groups.

These patterns suggest that geographic variation primarily reflects differences in baseline risk rather than a differential effect of BMI. In other words, BMI operates in a broadly homogeneous way across regions, while contextual factors shift the overall level of long COVID prevalence.

Fig. 5.10: Country-level Differences in the Impact of COVID-19



To further investigate geographic variation, a specification including country fixed effects is estimated in Figure 5.10. The joint test confirms strong cross-country heterogeneity (χ^2 = 141.30, $p < 0.001$), indicating that national context plays a substantial role in shaping long COVID outcomes.

Predicted probabilities vary markedly across countries. Higher values are observed in several Eastern European countries, including Poland, Bulgaria, and Croatia, where the estimated probability exceeds 0.80. In contrast, substantially lower levels are found in countries such as Finland, Spain, and the Netherlands, with predicted probabilities around or below 0.45.

Importantly, these differences persist after controlling for individual characteristics, comorbidities, socioeconomic status, and COVID severity. This indicates that country-level factors are associated with differences in long COVID risk beyond individual-level characteristics.

From a policy perspective, these findings are consistent with the uneven distribution of obesity and related health interventions across Europe. Countries with higher long

COVID probabilities may also reflect broader differences in population health, healthcare systems, and reporting practices, although these mechanisms cannot be directly identified within the present analysis. While BMI does not exhibit region-specific effects, its contribution to long COVID risk operates within broader structural environments. These findings highlight how pre-pandemic metabolic risk constitutes a general vulnerability factor that is not confined to acute health outcomes but extends to post-infection conditions.

From a broader perspective, this evidence is consistent with the economic rationale for preventive health policies discussed in Chapter III. If metabolic risk increases susceptibility not only to chronic diseases but also to long-term consequences of health shocks, policies aimed at reducing obesity may contribute to strengthening population resilience over time. However, it is important to emphasize that the present analysis does not provide causal evidence on the effectiveness of such policies, but rather supports their relevance within a broader preventive framework.

5.5 The role of Sugar Tax

To further investigate whether structural policy environments are associated with differences in long COVID risk, the analysis is extended by introducing a policy-based classification of countries. In particular, countries are grouped according to the presence of a national-level sugar-sweetened beverage (SSB) tax, allowing for a comparison between policy and non-policy contexts.

This classification builds directly on the institutional framework discussed in Chapter III, where selected European countries were analyzed in terms of their implementation of preventive fiscal policies targeting excessive sugar consumption. In line with that discussion, countries are classified as "sugar tax" if they have implemented a well-established national SSB tax aimed at reducing metabolic risk through price incentives.

Within the present sample, this group includes France, Belgium, Hungary, and Portugal, which represent consolidated examples of fiscal intervention in nutrition policy. All remaining countries are classified as non–sugar tax environments. Countries with partial, subnational, or discontinued policies, such as Italy, are excluded from the treatment group in order to ensure a consistent and comparable definition of policy exposure.

The objective of this specification is to assess whether the association between pre-pandemic BMI and the probability of long COVID differs across these two policy environments. While the analysis does not aim to establish causal effects, it provides an exploratory assessment of whether preventive fiscal frameworks are associated with different patterns of vulnerability to post-acute health outcomes.

The model controls for the same set of demographics, clinical, socioeconomic, behavioral, and regional covariates included in the full specification, so that any observed difference can be interpreted conditional on individual-level characteristics.

The results of the model including the interaction between BMI categories and sugar-tax status. The joint specification is statistically significant ($p < 0.01$), confirming the overall explanatory power of the model.

The interaction between BMI and sugar-tax status does not reveal strong evidence of heterogeneity. The coefficients associated with the interaction terms are not statistically significant at conventional levels for overweight individuals (OR = 0.73, $p = 0.339$), while the coefficient for obesity is only marginally significant (OR = 0.54, $p = 0.094$). This suggests that the association between pre-pandemic BMI and long COVID is broadly similar across countries with and without sugar tax policies.

Nonetheless, the direction of the interaction effects points to a potentially weaker association between high BMI and long COVID in countries that have implemented sugar taxes. In particular, the negative coefficient for the obesity interaction term indicates that the excess risk associated with obesity may be attenuated in these policy environments, although this effect is not estimated with sufficient precision to draw definitive conclusions.

The main effects remain consistent with previous specifications. Obesity is positively and significantly associated with long COVID (OR = 1.39, $p < 0.05$), while the coefficient for overweight is only marginally significant. The effect of gender remains strong and stable, with women exhibiting significantly higher odds of persistent symptoms (OR = 1.44, p <

0.01).

Clinical severity continues to represent the dominant determinant of long COVID. Hospitalization is associated with a more than threefold increase in the likelihood of persistent symptoms ($OR = 3.45$, $p < 0.01$), confirming the central role of acute disease severity in shaping long-term outcomes. Among comorbidities, heart disease remains positively associated with long COVID, while other conditions do not display statistically significant effects.

In conclusion, these findings suggest that preventive fiscal policies such as sugar taxes are not associated with substantial differences in the BMI–long COVID relationship at the aggregate level. However, the direction of the estimated effects is consistent with the hypothesis that policy environments targeting metabolic risk may contribute to mitigating vulnerability to post-acute health conditions. Importantly, these results should not be interpreted as causal, but rather as indicative of broader structural differences across countries.

5.5.1 Mediation Role of Sugar Tax

The previous analysis has explored whether the association between pre-pandemic BMI and the probability of developing long COVID differs across countries with and without sugar tax policies. While this heterogeneity approach provides useful descriptive insights, it does not fully capture the underlying mechanisms through which policy environments may influence health outcomes.

Sugar taxes are not designed to directly affect post-viral conditions such as long COVID. Rather, their primary objective is to reduce excessive sugar consumption and, over time, to influence metabolic risk factors at the population level. Among these, body mass index (BMI) represents a key intermediate outcome, reflecting cumulative exposure to unhealthy dietary behaviors and playing a central role in shaping individual vulnerability to a wide range of chronic and infectious conditions.

This suggests the presence of a potential indirect pathway linking policy environments to post-acute health outcomes. Instead of acting directly on long COVID risk, sugar tax policies may operate through their effect on BMI distribution, which in turn has been shown to be significantly associated with the likelihood of persistent symptoms. From this perspective, BMI can be interpreted as a mediating variable through which preventive fiscal policies exert their influence on health outcomes.

Building on this intuition, the present section moves beyond a purely interaction-based framework and adopts a mediation perspective. The objective is to assess whether sugar tax policies are associated with differences in long COVID risk through their impact on metabolic conditions, rather than through direct effects. This approach allows for a more structural interpretation of the role of policy environments, highlighting the importance of upstream determinants of health in shaping population resilience to post-pandemic conditions.

While the analysis remains observational and does not aim to establish causal relationships, it provides an exploratory assessment of the extent to which preventive fiscal policies may contribute to reducing vulnerability to long COVID by influencing intermediate health risk factors.

5.5.1.1 Methodology: Bivariate Probit Model

The empirical strategy employs a bivariate probit model to jointly estimate the relationship between metabolic risk and the probability of developing Long COVID, while accounting for potential correlation in unobserved determinants affecting both outcomes.

This modelling approach is widely used in applied health economics when dealing with interdependent binary outcomes and potential endogeneity. Recursive bivariate probit models allow for a simultaneous specification of two equations while explicitly modelling the correlation between their error terms (ρ), capturing unobserved factors that jointly influence both processes, as shown in Giachello et al. (2024). Similarly, multivariate probit frameworks have been used to account for complex interrelations between health behaviors, healthcare utilization, and outcomes, highlighting the role of unobserved heterogeneity (Di Novi et al., 2020)). A fundamental requirement of the bivariate probit model is that both dependent variables must be binary. For this reason, the original categorical BMI variable is transformed into a dichotomous indicator

(BMI_binary), equal to 1 for individuals classified as overweight or obese (BMI ≥ 25) and 0 otherwise.

The model is specified through two latent equations:

- Long COVID equation:

$$LongCovid = \beta_1 BMI\_binary + \beta_2 SugarTax + X\gamma + \varepsilon_1$$

- BMI equation:

$$BMI\_binary = \alpha_1 SugarTax + X\delta + \varepsilon_2$$

Observed outcomes correspond to the binary realizations of these latent variables. X includes demographic, clinical, socioeconomic, and behavioral controls.

A key feature of the model is the allowance for correlation between the error terms:

$$corr(\varepsilon_1, \varepsilon_2) = \rho$$

A statistically significant ρ indicates the presence of shared unobserved determinants, justifying the use of a joint estimation framework instead of separate probit models.

Within this framework, the specification follows a recursive structure, where metabolic risk (BMI) enters the Long COVID equation as a predictor. Sugar tax is included in both equations with distinct roles: it captures the influence of the policy environment on metabolic risk in the BMI equation, and it is included as a control in the Long COVID equation to isolate any direct, non-mediated effects.

BMI is therefore interpreted as a potential mediating channel linking policy exposure to health outcomes. Evidence consistent with mediation arises if sugar tax significantly affects BMI and BMI, in turn, significantly predicts Long COVID risk.

The model is estimated via maximum likelihood under the assumption of joint normality of the error terms. All specifications include the same set of controls used in previous regressions to ensure consistency.

However, while the bivariate probit framework partially addresses endogeneity concerns through the correlation structure of the error terms, the results remain associational rather than causal, as no external instrument or identification strategy is implemented. Findings should therefore be interpreted as suggestive of underlying mechanisms rather than definitive causal effects.

5.5.1.2 Results and Discussion

The estimation of the bivariate probit model provides a joint assessment of the determinants of Long COVID and metabolic risk, while accounting for potential correlation in unobserved factors affecting both outcomes. The model is globally significant (Wald $\chi^2 = 329.36$, $p < 0.001$), indicating that the set of covariates included in the specification has strong explanatory power.

Tab 5.5.1 Bivariate Probit Estimates for Long COVID and BMI

Variables	**Long COVID equation**	**BMI equation**
BMI_binary	1.067* (0.589)	-
SugarTax	0.030 (0.120)	0.024 (0.116)
Female	0.182** (0.081)	0.023 (0.061)
Age 65-74	-0.196** (0.095)	0.030 (0.087)
Age 75+	-0.096 (0.151)	-0.236** (0.106)
Constant	-0.063 (0.529)	0.081 (0.305)

*Notes: The table reports the estimated coefficients from the bivariate probit model. The values reported outside parentheses correspond to the estimated coefficients, while the values in parentheses represent the associated standard errors. Statistical significance is indicated by asterisks: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. ρ (rho) -0.643 (0.397); Wald test ($\rho = 0$) $\chi^2 = 1.27$ ($p = 0.259$); Log-likelihood -2236.98; Observations 1,862.*

For ease of presentation, only a subset of key variables is reported in the table, including BMI_binary, Sugar Tax, Female, and age categories (65–74 and 75+), along with the constant term. The remaining control variables are included in the estimation but not displayed.

The correlation coefficient between the error terms (ρ), the Wald test of independence ($\rho = 0$), the log-likelihood, and the number of observations is reported as summary statistics of the model without further transformation.

Focusing on the Long COVID equation, metabolic risk emerges as a relevant predictor of post-acute health outcomes. Individuals classified as overweight or obese exhibit a higher probability of developing persistent symptoms, with the estimated coefficient being positive and marginally statistically significant ($\beta = 1.07$, $p = 0.070$). This result provides weak but consistent evidence of a positive association between pre-pandemic BMI and Long COVID vulnerability, in line with previous specifications and with existing findings in the literature (Wilk et al., 2024). The magnitude of the coefficient suggests that excess body weight operates as a risk factor, although the limited statistical significance calls for cautious interpretation.

Turning to the BMI equation, the results do not support the hypothesis that sugar tax policies significantly influence metabolic risk within the sample. The estimated coefficient associated with the sugar tax variable is small and not statistically significant ($\beta = 0.024$, $p = 0.833$), indicating that the presence of national sugar-sweetened beverage taxation is not associated with a lower probability of being overweight or obese. This finding suggests that cross-country variation in fiscal policies is not sufficient to explain differences in individual metabolic conditions in the short run. Among the control variables, education stands out as a significant determinant: higher educational attainment, particularly at the tertiary level, is associated with a lower likelihood of being in a high BMI category ($\beta = -0.31$, $p = 0.001$), consistent with well-established socioeconomic gradients in health.

These results have important implications for the proposed mediation framework. While BMI remains a relevant predictor of Long COVID, the absence of a statistically significant relationship between sugar tax and metabolic risk implies that the first link in the hypothesized causal chain is not empirically supported. Consequently, no evidence of a mediation mechanism through BMI can be established within this specification. The policy variable does not appear to operate through the expected pathway, at least within the time horizon and data structure considered.

A key feature of the bivariate probit model is the estimation of the correlation coefficient between the error terms of the two equations. The estimated ρ is negative (ρ = -0.64), suggesting that unobserved factors that increase the probability of being overweight or obese may be associated with a lower probability of developing Long COVID, and vice versa. However, the coefficient is not statistically significant (p = 0.259), implying that the null hypothesis of no correlation between the error terms cannot be rejected. This indicates that, conditional on observed covariates, there is no strong evidence of systematic unobserved heterogeneity jointly affecting both outcomes.

The lack of statistical significance of ρ has further implications for model interpretation. It suggests that the efficiency gains from joint estimation may be limited, and that separate probit models could provide similar results. This finding points to a scenario in which the relationship between BMI and Long COVID is primarily captured by observed characteristics rather than by shared latent factors.

Overall, the results highlight a clear and relatively stable role of BMI as a determinant of Long COVID risk, while failing to detect any meaningful effect of sugar tax policies on metabolic conditions. The association between pre-pandemic health status and post-acute outcomes appears to be largely driven by individual-level demographic and clinical characteristics, rather than by macro-level fiscal interventions. Although the empirical evidence does not support the existence of a mediation mechanism through BMI in this setting, it does not exclude the possibility of longer-term or indirect policy effects that are not captured by the available data.

Finally, it is important to emphasize that, despite the use of a joint modelling framework, the results remain associational. In the absence of a formal identification strategy or exogenous variation in policy exposure, the estimated relationships should be interpreted as indicative of potential mechanisms rather than as causal effects.

Chapter VI: Discussion and Policy Recommendations

6.1 Summary of Findings

The empirical analysis provides robust evidence that pre-pandemic body mass index (BMI) is significantly associated with the probability of developing Long COVID, even after controlling for a comprehensive set of demographics, clinical, and socioeconomic characteristics. This finding indicates that pre-existing metabolic risk constitutes an independent determinant of post-COVID health outcomes.

A clear and consistent gradient emerges across weight categories: overweight and obese individuals exhibit progressively higher probabilities of persistent symptoms compared to those with normal weight. This pattern remains stable across model specifications and after the inclusion of multiple controls, confirming that excess body weight represents a key correlate of Long COVID risk.

Gender differences represent an additional and persistent dimension of heterogeneity. Women display systematically higher probabilities of Long COVID relative to men, and this effect remains stable across all empirical specifications. This suggests that sex-related factors operate primarily through differences in baseline risk rather than through modifications of the relationship between BMI and long-term outcomes.

Clinical severity emerges as the strongest predictor within the empirical framework. Individuals who experienced COVID-19-related hospitalization show a substantially higher likelihood of persistent symptoms, indicating that the intensity of the acute phase plays a central role in shaping long-term health conditions. Importantly, hospitalization shifts the overall level of risk without altering the underlying association between BMI and Long COVID.

The analysis of interaction effects does not reveal statistically significant heterogeneity in the BMI–Long COVID relationship across subgroups. Neither gender nor hospitalization status significantly modifies the slope of the association, suggesting that the impact of metabolic risk is broadly consistent across individuals.

Regarding the policy dimension, the empirical evidence does not support the presence of statistically significant differences in Long COVID risk between countries with and without sugar-sweetened beverage taxes. Interaction effects are not significant, indicating that the policy environment is not associated with observable variations in the BMI–Long COVID relationship at the individual level.

This result is reinforced by the bivariate probit analysis, which confirms that BMI remains a statistically significant predictor of Long COVID even when accounting for potential endogeneity. In contrast, the policy variable does not exhibit a significant association with BMI, providing no evidence in support of the hypothesized indirect pathway through which fiscal policy would affect Long COVID risk via metabolic health.

Taken together, these findings suggest that Long COVID risk is primarily driven by individual-level factors, particularly pre-existing metabolic vulnerability, and the severity of the acute infection. While fiscal policies such as sugar-sweetened beverage taxes may influence population health over longer time horizons, their effects are not directly observable within the scope of this empirical framework and do not appear to mediate the relationship between BMI and post-COVID outcomes.

6.2 Interpreting the BMI–Long COVID Relationship: Biological and Economic Mechanisms

Pre-pandemic body mass index can be interpreted as a proxy for accumulated metabolic vulnerability, reflecting long-term exposure to behavioral and environmental risk factors. Within this framework, Long COVID does not emerge as an isolated outcome of infection, but as the result of an interaction between an acute health shock and pre-existing individual susceptibility. The positive association observed between BMI and persistent symptoms therefore suggests that excess body weight captures a structural dimension of vulnerability that extends beyond chronic disease and shapes resilience to post-acute conditions.

From a biological perspective, elevated BMI is associated with chronic low-grade inflammation and altered immune responses, which may impair the body’s ability to recover fully after infection. Rather than determining the occurrence of the initial shock, metabolic risk appears to influence its persistence, increasing the likelihood that symptoms extend over time. In this sense, higher BMI can be understood as amplifying the duration and intensity of the health impact generated by COVID-19, contributing to prolonged functional impairment.

Gender differences further refine this interpretation. The evidence indicates that women exhibit a consistently higher probability of Long COVID, while the relationship between BMI and persistent symptoms remains similar across sexes. This suggests that gender operates primarily through a shift in baseline risk rather than through a modification of the underlying mechanism linking metabolic vulnerability to post-acute outcomes. The observed gap may reflect a combination of biological factors and differences in symptom reporting but does not imply a fundamentally different role of BMI across male and female individuals.

The role of clinical severity provides a complementary dimension to this framework. Hospitalization, as a proxy for the intensity of the acute infection, emerges as the strongest predictor of Long COVID, substantially increasing the probability of persistent symptoms. However, its effect operates primarily by shifting the overall level of risk rather than altering the relationship between BMI and long-term outcomes. This distinction suggests that while metabolic vulnerability shapes the underlying susceptibility to prolonged symptoms, the severity of the initial shock determines the magnitude of the resulting health impact.

The absence of statistically significant interaction effects further supports this interpretation. The relationship between BMI and Long COVID remains stable across subgroups defined by gender and hospitalization status, indicating that excess body weight exerts a broadly uniform influence on the persistence of symptoms. Differences across individuals are therefore captured mainly by variations in baseline risk rather than by changes in the slope of the association, reinforcing the view of BMI as a general and structural determinant of post-acute vulnerability.

The analysis of policy environments introduces an additional, albeit indirect, perspective. The lack of strong statistical evidence for differential effects across countries with and without sugar-sweetened beverage taxes suggests that such policies do not immediately translate into observable differences in Long COVID risk. Nonetheless, the direction of the estimated effects points toward a potential attenuation of BMI-related risk in countries with established preventive frameworks. This pattern is consistent with the interpretation of fiscal interventions as mechanisms that operate gradually by influencing metabolic risk profiles over time, rather than by directly affecting short-term health outcomes.

These findings can be directly connected to the broader framework of preventive fiscal policies discussed in Chapter III. Sugar-sweetened beverage taxes are not designed to influence health outcomes in the short run, but to modify the economic environment in which dietary choices are made. By affecting relative prices and incentives, such policies gradually reshape consumption patterns and, consequently, the distribution of metabolic risk within the population. In this perspective, their relevance lies in their capacity to act upstream before vulnerability fully materializes.

From an economic standpoint, body mass index can be interpreted as a component of individual health capital, influencing both health trajectories and the capacity to absorb adverse shocks. Long COVID represents a form of prolonged impairment that carries potential consequences in terms of reduced productivity, increased healthcare utilization, and diminished well-being. The observed association between BMI and persistent symptoms therefore extends the economic relevance of metabolic risk beyond chronic disease, highlighting its role in shaping resilience to unexpected health events.

Overall, the results support a broader interpretation of preventive policy as an investment in structural resilience. If higher BMI increases susceptibility to long-term health consequences following systemic shocks, interventions aimed at reducing metabolic risk may generate benefits that go beyond traditional disease-specific outcomes. In this sense, preventive strategies can be understood not only as tools for improving average health, but also as mechanisms for mitigating the impact of future health crises on individuals and healthcare systems.

6.3 Policy Implications: Targeted Prevention and Post-pandemic Health Strategies

The findings of this study suggest that Long COVID should not be interpreted solely because of acute infection, but as an outcome shaped by pre-existing structural vulnerabilities within the population. In particular, the positive association between pre-pandemic BMI and persistent symptoms indicates that metabolic risk constitutes a key dimension of susceptibility to post-acute health conditions. This perspective implies that policies aimed at improving population health prior to the occurrence of systemic shocks may play a critical role in reducing the long-term burden of disease. Preventive action therefore acquires relevance not only for chronic disease control, but also for strengthening resilience to unexpected health crises.

Within this framework, interventions targeting metabolic risk emerge as a central policy priority. Reducing the prevalence of overweight and obesity is not simply a matter of improving average health outcomes, but of modifying the distribution of vulnerability across individuals. Policies that influence dietary behavior, access to healthier food environments, and long-term lifestyle patterns can contribute to shifting population-level BMI distributions, thereby lowering the proportion of individuals exposed to elevated risk of persistent post-infection conditions. The empirical evidence supports the interpretation of metabolic prevention as a structural intervention with potential benefits that extend beyond traditional disease-specific endpoints.

Fiscal instruments such as sugar-sweetened beverage taxes can be understood within this broader preventive logic. Although the empirical analysis does not provide strong statistical evidence of differential effects on Long COVID across policy environments, the direction of the estimated coefficients is consistent with a potential attenuation of BMI-related risk in countries with established fiscal measures. This suggests that such policies operate through indirect and cumulative mechanisms, influencing consumption patterns and metabolic risk profiles over time rather than producing immediate changes in health outcomes. Their policy relevance therefore lies in their capacity to act upstream, shaping the conditions that determine vulnerability before adverse health events occur.

The identification of specific high-risk groups further highlights the importance of targeted intervention strategies. The empirical evidence indicates that individuals with higher BMI, women, and those who experienced severe COVID-19 episodes face systematically greater probabilities of developing persistent symptoms. This suggests that uniform policy approaches may be insufficient to address the distribution of post-acute health risks. Instead, more efficient strategies should combine population-wide prevention with targeted measures aimed at particularly vulnerable groups, including enhanced screening, early identification of symptoms, and tailored follow-up care. Such an approach would allow healthcare systems to allocate resources more effectively while reducing long-term health deterioration among high-risk individuals.

The role of hospitalization as a strong predictor of Long COVID also underscores the need for structured post-infection management. Individuals who experienced severe disease represent a group with persistently elevated risk, requiring continuous monitoring and integrated care pathways beyond the acute phase of illness. This implies a shift in healthcare organization toward models that account for the long-term consequences of infectious diseases, including multidisciplinary follow-up programs and coordinated services addressing both physical and functional impairment. Strengthening post-COVID care is therefore essential to mitigate the long-term burden on healthcare systems and to prevent the progression of persistent symptoms into chronic conditions.

More broadly, these findings point to the need for a systemic approach to health policy that integrates prevention, acute care, and long-term management within a unified framework. The interaction between pre-existing metabolic vulnerability and acute health shocks suggests that resilience is not determined solely by the capacity to respond to crises, but also by the underlying health profile of the population. Policies aimed at improving baseline health conditions can therefore enhance the ability of societies to absorb and recover from future shocks. In this sense, strengthening population health becomes a central component of broader strategies aimed at ensuring both healthcare sustainability and economic stability in the face of emerging health risks.

Beyond its clinical implications, Long COVID also carries significant economic consequences. Persistent symptoms may translate into reduced productivity, increased absenteeism, and a lower capacity to participate in the labor market, particularly among older individuals. At the same time, prolonged health impairment generates sustained demand for healthcare services, placing additional pressure on already constrained public systems. Within this framework, the observed relationship between BMI and Long COVID highlights how metabolic risk can influence not only health outcomes but also economic performance. Preventive interventions aimed at reducing excess body weight can therefore be interpreted as long-term investments, capable of lowering future healthcare costs and preserving individual productive capacity.

At the same time, the policy implications derived from this analysis should be interpreted with appropriate caution. The empirical results identify conditional associations rather than causal effects, and the evidence on sugar-sweetened beverage taxation does not allow for definitive conclusions regarding its impact on Long COVID risk. The absence of strong statistical significance in the interaction terms suggests that any potential effect operates indirectly and may not be captured within the timeframe of the analysis. For this reason, policy recommendations should be understood as consistent with the observed patterns, rather than as direct evidence of policy effectiveness.

In conclusion, the findings support a broader view of preventive health policy as a tool for reducing structural vulnerability within the population. Interventions targeting metabolic risk do not only contribute to lowering the incidence of chronic diseases but may also enhance the capacity of individuals and healthcare systems to withstand future health shocks. In this sense, improving baseline health conditions becomes a central component of strategies aimed at strengthening both public health resilience and long-term economic sustainability.

6.4 Limitations and Future Research Directions

This study is subject to several limitations that should be considered when interpreting the results. While the empirical analysis provides consistent evidence of an association between pre-pandemic BMI and the probability of developing Long COVID, the findings should be understood within the constraints of the data and the methodological approach adopted.

First, the empirical strategy identifies conditional associations rather than causal effects. Although the models control for a wide range of demographic, clinical, and socioeconomic characteristics, the possibility of unobserved heterogeneity cannot be fully excluded. Factors not captured in the data, such as differences in health behaviors or access to healthcare, may influence both pre-pandemic BMI and the likelihood of persistent symptoms. As a result, the estimated relationships should be interpreted as indicative of systematic correlations rather than definitive evidence of causality.

Second, the measurement of body mass index is subject to potential limitations. BMI is derived from self-reported height and weight and is therefore exposed to reporting errors and measurement inaccuracies. In addition, the use of categorical BMI indicators, while facilitating interpretation, may obscure non-linear patterns within weight groups and reduce the precision of the estimated relationships. These aspects may affect the ability of the analysis to capture the full complexity of the association between metabolic risk and Long COVID.

Third, the dependent variable capturing Long COVID is based on self-reported information on persistent symptoms, which may be subject to reporting bias and cross-country differences in perception and diagnosis. The heterogeneity of symptoms and the absence of a universally standardized clinical definition further complicate the measurement of the outcome. Consequently, variation in reported Long COVID prevalence may reflect not only differences in underlying health conditions, but also differences in reporting behavior and survey interpretation.

A further limitation concerns the structure of the analytical sample. The empirical analysis is conducted on individuals who experienced COVID-19, implying that the results are conditional on infection and do not capture the determinants of exposure to the virus. This restriction is necessary to study the persistence of symptoms, but it limits the scope of inference to post-infection outcomes. As a result, the findings should not be interpreted as evidence on the overall health risk associated with BMI, but specifically on its relationship with the likelihood of developing Long COVID among infected individuals.

Additional caution is required in interpreting the results related to policy environments. The classification of countries according to the presence of sugar-sweetened beverage taxes is based on a simplified binary indicator, which does not fully capture differences in policy design, intensity, or duration. Variations in tax structure, implementation timing, and complementary interventions may influence their effectiveness, but cannot be fully accounted for within the present framework. For this reason, the evidence on the role of fiscal policies should be considered exploratory and indicative rather than conclusive.

The temporal dimension of the analysis represents a further constraint. Preventive policies targeting metabolic risk are likely to produce effects over extended time horizons, while the available data capture a relatively limited period surrounding the COVID-19 pandemic. This mismatch implies that the long-term impact of such policies on population health and resilience may not be fully observable within the timeframe considered. Consequently, the absence of strong empirical effects should not be interpreted as evidence of ineffectiveness, but rather as a reflection of the gradual nature of these processes.

Future research could build on these findings by adopting methodologies better suited to identifying causal relationships, such as quasi-experimental designs or policy evaluation frameworks exploiting variation in the timing and intensity of fiscal interventions. In addition, the use of continuous measures of BMI and longer-term longitudinal data would allow for a more detailed analysis of non-linear effects and dynamic health trajectories. Expanding the analysis to incorporate more granular information on policy design and behavioral responses would further contribute to understanding the mechanisms linking preventive interventions, metabolic risk, and post-acute health outcomes.

This thesis highlights the relevance of pre-pandemic metabolic vulnerability in shaping the persistence of post-COVID symptoms, while emphasizing the role of acute disease severity and broader structural factors in determining individual health outcomes. The findings suggest that Long COVID cannot be fully understood without considering the interaction between pre-existing conditions and systemic shocks, pointing to the importance of integrated approaches to health policy.

More broadly, the results support a reinterpretation of preventive health policy as an investment in resilience. By influencing the distribution of metabolic risk within the population, interventions targeting obesity and related behaviors may contribute not only to reducing the burden of chronic disease, but also to strengthening the capacity of individuals and healthcare systems to withstand future health crises. In this sense, improving baseline health conditions emerges as a key component of strategies aimed at promoting both long-term public health and economic sustainability.

Despite these limitations, the analysis provides consistent evidence of a systematic association between pre-pandemic BMI and Long COVID. This supports the relevance of metabolic risk as a key dimension in post-pandemic health outcomes. Ultimately, strengthening population health before crises emerge is not only a medical priority, but a fundamental economic strategy for reducing long-term vulnerability and ensuring sustainable health systems.

References

Allcott H., Lockwood B.B., Taubinsky D., 2019, "Regressive sin taxes, with an application to the optimal soda tax", *Quarterly Journal of Economics*, Vol. 134, n. 3, pp. 1557–1626.

Arrow K.J., 1963, "Uncertainty and the welfare economics of medical care", *American Economic Review*, Vol. 53, n. 5, pp. 941–973.

Berardi N., Sevestre P., Tepaut M., Vigneron A., 2016, "The impact of a 'soda tax' on prices: Evidence from French micro data", *Applied Economics*, Vol. 48, n. 41, pp. 3976–3994.

Bíró A., 2015, "Did the junk food tax make the Hungarians eat healthier?", *Food Policy*, Vol. 54, pp. 107–115.

Blüher M., 2019, "Obesity: global epidemiology and pathogenesis", *Nature Reviews Endocrinology*, Vol. 15, pp. 288–298.

Börsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., & Zuber, S. (2013). *Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE).* International Journal of Epidemiology, 42(4), 992–1001.

Börsch-Supan, A. (2021). *Survey of Health, Ageing and Retirement in Europe (SHARE) Corona Survey Release Guide.*SHARE-ERIC.

Börsch-Supan, A. (2022). *Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 8 Release Guide.* SHARE-ERIC.

Colchero M.A., Popkin B.M., Rivera J.A., Ng S.W., 2016, "Beverage purchases from stores in Mexico under the excise tax on sugar sweetened beverages: observational study", *BMJ*, Vol. 352, h6704.

Di Novi C., Leporatti L., Montefiori M., 2020, "Older patients and geographic barriers to pharmacy access: When nonadherence translates to an increased use of other components of health care", Health Economics, pp. 1–13.

Drummond M.F., Sculpher M.J., Claxton K., Stoddart G.L., Torrance G.W., 2015, *Methods for the Economic Evaluation of Health Care Programmes*, 4th ed., Oxford University Press, Oxford.

Effertz T., Engel S., Verheyen F., Linder R., 2016, "The economic costs of obesity in Germany", *European Journal of Health Economics*, Vol. 17, pp. 539–549.

Emery C., Dinet J., Lafuma A., 2017, "The economic burden of obesity in France", *Health Economics Review*, Vol. 7, Article 10.

GBD 2019 Risk Factors Collaborators, 2020, "Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019", *The Lancet*, Vol. 396, n. 10258, pp. 1223–1249.

Giachello M., Leporatti L., Levaggi R., Montefiori M., 2024, "The illness trap: The impact of disability benefits on willingness to receive HCV treatment", Economics and Human Biology, Vol. 52, 101352.

Goiana-da-Silva F., Severo M., Cruz e Silva D., Gregório M.J., Allen L.N., Muc M., Morais Nunes A., Torres D., Miraldo M., Ashrafian H., Rito A., Wickramasinghe K., Breda J., Darzi A., Araújo F., Lopes C., 2020, "Projected impact of the Portuguese sugar-sweetened beverage tax on obesity incidence across different age groups: A modelling study", *PLoS Medicine*, Vol. 17, n. 3, e1003036.

Grossman M., 1972, "On the concept of health capital and the demand for health", *Journal of Political Economy*, Vol. 80, n. 2, pp. 223–255.

Gruber J., Kőszegi B., 2001, "Is addiction 'rational'? Theory and evidence", *Quarterly Journal of Economics*, Vol. 116, n. 4, pp. 1261–1303.

HM Treasury (2016). *Soft Drinks Industry Levy*. London: UK Government.

Liu, S., Veugelers, P.J., Liu, C., & Ohinmaa, A. (2022). *The cost effectiveness of taxation of sugary foods and beverages: A systematic review of economic evaluations*. Applied Health Economics and Health Policy, 20, 185–198.

Nalbandian A., Sehgal K., Gupta A., Madhavan M.V., McGroder C., Stevens J.S., Cook J.R., Nordvig A.S., Shalev D., Sehrawat T.S., 2021, "Post-acute COVID-19 syndrome", *Nature Medicine*, Vol. 27, n. 4, pp. 601–615.

Ng M., Fleming T., Robinson M., Thomson B., Graetz N., Margono C., Mullany E.C., 2014, "Global, regional, and national prevalence of overweight and obesity 1980–2013", *The Lancet*, Vol. 384, n. 9945, pp. 766–781.

OECD, 2019, *The Heavy Burden of Obesity: The Economics of Prevention*, OECD Publishing, Paris.

OECD, 2024, *The impacts of long COVID across OECD countries*, OECD Health Working Papers No. 167, OECD Publishing, Paris.

O'Donoghue T., Rabin M., 2006, "Optimal sin taxes", *Journal of Public Economics*, Vol. 90, n. 10–11, pp. 1825–1849.

Popkin B.M., Du S., Green W.D., Beck M.A., Algaith T., Herbst C.H., Alsukait R.F., Alluhidan M., Alazemi N., Shekar M., 2020, "Individuals with obesity and COVID-19: A global perspective", *Obesity Reviews*, Vol. 21, n. 11.

Turchetti G., Vitiello N., Palla I., 2012, “The economic burden of obesity in Italy”, *European Journal of Public Health*, Vol. 22, n. 5, pp. 684–690.

Vandevijvere S., Berger N., De Ridder K., Fiolet T., 2020, “The impact of Belgium’s sugar-sweetened beverage tax on beverage purchases”, Public Health Nutrition, Vol. 23, n. 6, pp. 1073–1083

WHO, 2000, *Obesity: Preventing and Managing the Global Epidemic*, WHO Technical Report Series No. 894, World Health Organization, Geneva.

WHO, 2016, *Fiscal policies for diet and prevention of noncommunicable diseases*, World Health Organization, Geneva.

WHO, 2021, *A clinical case definition of post COVID-19 condition by a Delphi consensus*, World Health Organization, Geneva.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data.* MIT Press.

Wilk P., Stranges S., Cuschieri S., 2024, “Does sex modify the effect of prepandemic body mass index on the risk of Long COVID? Evidence from the longitudinal analysis of the Survey of Health, Ageing and Retirement in Europe”, *International Journal of Obesity*, Vol. 48, n. 6, pp. 821–829.

Withrow D., Alter D.A., 2011, “The economic burden of obesity worldwide: a systematic review of the direct costs of obesity”, *Obesity Reviews*, Vol. 12, n. 2, pp. 131–141.