UNIVERSITAT DE BARCELONA

Essays on the Economics of Obesity

Athina Raftopoulou

PhD in Economics

Essays on the Economics of Obesity

Athina Raftopoulou

2018

PhD in Economics | Athina Raftopoulou

UNIVERSITAT DE BARCELONA

PhD in Economics

Thesis title:

Essays on the Economics of Obesity

PhD student:

Athina Raftopoulou

Advisor:

Joan Gil Trasfi

Date:

December 2018

UNIVERSITAT DE BARCELONA

Abstract

This thesis comprises five chapters in total, starting with a general introduction that raises the issue of obesity as well as a brief description of the basic research questions of the dissertation, three main chapters involving the analysis of the Body Mass Index (BMI) with a view to investigate the social, economic, cultural and environmental factors driving and sustaining health disparities in obesity in Spain and a chapter of concluding remarks stemming from the analysis made. In particular, the second chapter examines the evolution of obesity as well as the income-related inequality in obesity over the past two decades in Spain, splitting by gender. It also evaluates income inequality in obesity (measured by distribution sensitive measures) by breaking it down to its main contributors. The results indicate that obesity prevalence rates have been increasing over the last twenty years among the Spanish population, as in most developed countries, however income-related inequality in obesity status, depth and severity has a declining trend mainly among women. These findings may imply a switch in the basic determinants of obesity across the income distribution; that is, BMI status might not be linked only to individual attributes, but changes in environmental influences across income groups may be important as well. This is inextricably linked to the third chapter, where we seek to understand the basic determinants of individual body weight and obesity risk, by concurrently examining individual and regional characteristics within a multilevel approach, to conclude that not only personal attributes but also environmental characteristics (i.e., criminality and lack of green spaces) affect positively individual and women s BMI and obesity. Driven by the spatial pattern of BMI that is observed in this third chapter, according to which southern regions of Spain tend to exhibit higher BMI levels than the northern ones, we proceed with chapter four. In this fourth chapter we aim to contribute to the North to South health divide in Spain, using decomposition techniques to analyse the main contributors of the BMI gap between the North and the South of Spain. Our findings indicate that North to South differences are significant only for women and that the largest share of this gap is attributed to differences in endowments (mainly education) to the detriment of women living in the South.

Acknowledgments

I would first like to express my sincere gratitude to my supervisor Joan Gil Trasfi for his great support and guidance not only in the dissertation phase but during the whole PhD process. He encouraged me to develop my own ideas and this work would not have been possible without his insight and advice on daily basis, since the start of our collaboration. I am also very grateful to Antonio Di Paolo, who contributed to my research with his valuable inputs, comments and suggestions, from which I benefited enormously. Special thanks to Dr. Nikolaos Giannakopoulos for his helpful advice and encouragement even before I started this journey as well as to Dra Concepcion Patxot Cardoner for her support.

Furthermore, I thank the Fundacio Bosch i Gimpera (FBG) and the Centre for Economic Analysis and Social Policies (CAEPS) for funding my research, as well as all the collaborators to the Verkami Platform (linked to the “Alumni Wall”) for contributing to my research project: Ramon Adell Ramon, Josep Ainaud i Escudero, Ramon Alemany Leira, Salvador Alemany Mas, Alexauea, Pere Aragonès i Garcia, Enric Arderiu Gras, Georgina Arderiu Munill, Andreu Arenas Jal, Manuel Artís Ortuño, Eduard Arruga i Valeri, Jordi Berguido Covas, Guiem Bosch, Aïda Bravo Almirall, Bartomeu Caldentey, Ramona Canals Puy, Marc Casañas, Albert Castellanos, David Ceballos Hornero, Daniel Daví Arderius, Màrius Domínguez i Amorós, Alejandro Estruch Manjón, Roser Fernández Alegre, Francesc Foraster i Arespacochaga, Jordi Garcia i Serra, Carol Gillanders, Laura Guitart Tarrés, Francesc Granell Trias, Joan Hortalà i Arau, Francesc Iglesies i Riumalló, Xavier Martinez Gil, Mariona Mas, Andreu Mas-Colell, Jordi Oliveras Vilallonga, Elisenda Paluzie i Hernàndez, Josep Piqué i Camps, M. Cristina Poblet Farrés, Marc Puigmulé Solà, Dídac Ramírez i Sarrió, Pau Ras i Jansà, Mar Raventós Chalbaud, Jordi Roca Solanelles, Xavier Rodríguez i Guasch, Carles Rojas, Gloria Rubert Adelantado, Joaquim Silvestre i Benach, Pere Sitjà Miquel, Anna Sitjà Tejedor, Montserrat Termes Rifé, Ramon Tremosa Balcells and Juan Tugores Ques.

I would also like to thank my friends who motivated me and provided me with all the support I needed during this phase of my life. Specially, I would like to thank my partner for his unconditional support, encouragement and understanding during the completion of my PhD thesis. Last but not least, I would like to thank my family for their infinite support during all these years of being abroad.

Preface

Chapter 2 Is a work in progress, co-authored with my supervisor Joan Gil Trasfi and in the process to be sent for presentation to Conferences and Worskhops. **Chapter 3** is a single-authored paper which is already published at the high-impact journal Economics and Human Biology (Economics and Human Biology, Vol 24, p:185-193, 2017). **Chapter 4** is a co-authored work with my supervisor Dr. Joan Gil Trasfi and Dr. Antonio Di Paolo. The paper is already published as a working paper at the IREA-AQR working paper series and is sent for publication to one of the top Health Economics Journals.

Contents

List of Figures

List of Tables

1 Introduction

This PhD Thesis deals with a public health issue that is among the most burdensome faced by the modern societies. In spite of intense public health focus to curtail growing levels of obesity worldwide, obesity remains one of the greatest contemporary challenges to the health of the world population. Statistics indicate that obesity rates have continued to rise the last decade, although at a lower pace than before (OECD, 2017). The result is severe health implications since obesity is associated with heightened risk of nearly every chronic disease (from diabetes, to poor mental health), as well as substantial costs for the economy (direct and indirect) and society. Obesity and related health problems, such as increased risk of heart attack and stroke, drive up healthcare costs, reduce worker productivity and increase obesity-related absenteeism. A large spectrum of intervention policies including unhealthy (healthy) food taxing (price subsidy) measures, transport policies (e.g. subsidies for active commuting instead of cars), school-based interventions, etc. have been proposed and implemented, however OECD projections show a steady increase in obesity rates until at least 2030 (OECD, 2017).

Obesity is most commonly defined using the body mass index (BMI) which is calculated based on one s weight in relation to their height, that is, as the ratio of weight in kg to the square of height in meters (known as the Quetelet index). In turn, body weight is determined by a combination of factors that include genetics, metabolism, as well as socioeconomic, cultural, and environmental influences, making obesity a complex and multifaceted disease. In the present dissertation we will mainly deal with the relationship between obesity and two of these determinants, the socioeconomic and the environmental factors, for two basic reasons. First, because they are among the largest contributors to overweight and obesity risk and second, because they provide the greatest opportunity for policy interventions.

Spain is one of the countries where growth trends in overall obesity might be levelling– off, however obesity and morbid obesity levels continue to climb. Adult obesity rates in Spain are higher than the OECD average and childhood obesity rates are amongst the highest in the OECD. This calls for further research focusing in unmasking the interplay of different forces contributing to excess weight gain. Aiming at providing as much information as we can with regard to the drivers of obesity in Spain, we examine obesity from three different standpoints that are though interconnected, with a view to provide relevant and useful information for policy makers.

In Chapter 2 we focus on the relationship between Socio-economic Status (SES) and obesity, and more specifically on the way income as a measure of SES and obesity are associated to one another to describe the health status (proxied here by obesity prevalence and two distribution-sensitive obesity measures, i.e. depth and severity of obesity) of the poor in comparison to the rich. There is a large volume of research that has identified a negative SES gradient in obesity (Sobal and Stunkard, 1989; Nayga, 1999; Chou et al., 2004; McLaren, 2007; Garcia-Villar and Quintana-Domeque, 2009; Sassi et al., 2009; Lakdawalla and Philipson, 2009; Baum and Ruhm, 2009), while when focusing on the Spanish case, several studies report income-related inequalities in obesity as well (Costa-Font and Gil, 2008; Costa-Font et al., 2010a; Rodriguez-Caro et al., 2016; Merino Ventosa and Urbanos-Garrido, 2016). Aiming to provide first European evidence, we do not evaluate income inequality only in obesity status, but also in its depth and severity, as the health risks associated with being obese are increasing even more at the top of the BMI distribution (Willett et al., 1999; Freedman et al., 2002; Zabarsky et al., 2018). We start by examining both changes in obesity rates as well as in income-related inequality in obesity over time using Spanish individual-level and cross-section data from the past two decades. In order to observe the main factors explaining income-related inequality in obesity, we decompose the overall inequality in obesity status, depth and severity to its main contributing factors (as in Bilger et al. (2017)). To achieve that, we make use of the Spanish data from the National Health Survey of Spain (ENSE), comparing two points in time (2017 with 1997) and splitting by gender. Our findings show that SES inequality in obesity status considerably differs by gender in agreement with the literature, while based on our decomposition results income is remarkably the most important (negative) contributor to the overall income inequality in obesity status, playing education a more modest role. Overall, obesity is concentrated among the poor in Spain since the CI is always negative and statistically significant, but the degree of inequality has a declining trend, even though all three measures of obesity increase over the years. This finding may reflect significant changes on the key determinants of individual obesity risk over the past years. A possible change could be the increasing importance of the characteristics of the built and food (obesogenic) environment in shaping the lifestyle and eating habits of the individuals. An emerging question then is, what can be inferred when we concurrently account not only for personal but also social environmental characteristics in the analysis of individual obesity risk?

With a view to answer this question, we move to Chapter 3 of the Thesis. In this chapter, we seek to understand the determinants of individual body weight status and obesity risk in Spain by concurrently examining individual and regional characteristics. An influential strand of the obesity literature argues that environmental influences (e.g., distance to grocery stores, parks, neighbourhood safety, green areas) represent the public health arm of the obesity problem (Egger and Swinburn, 1997; Mark Austin and Spine, 2002; Cubbin et al., 2006). In addition, according to Costa-Font and Gil (2008), sociocultural contexts of obesity are recognized as key factors that account for the development of an individual s weight. To reach our purposes, we carry out a multilevel analysis using data from the National Health Survey of Spain (ENSE) for the year 2011–2012. Our objective is to disentangle the different influences on individual weight status and obesity risk with our contribution lying on the model specification. It is our belief that previous studies have failed to take into account regional and individual characteristics in tandem when dealing with obesity. In this study, therefore, we control for personal characteristics as well as higher level geographic variations that may cause the body mass index (BMI) to rise above normal levels. Our contribution is to illustrate how individual weight status and obesity risk are explained by individual and regional characteristics – both in the immediate environment and in the broader setting – exploiting the hierarchical structure of the data, in a multilevel (ML) regression model. We conclude that both group and individual effects play a key role in understanding BMI and obesity and we provide evidence that our proxies of the social environment (criminality and green spaces) have a positive and statistically significant effect on female BMI and the prevalence of obesity. Another particularly interesting aspect of this application is that the institutional setting of Spain is very suitable for this kind of analysis, i.e. for an analysis that takes into account geographical influences. Notably, because of the country's decentralized health care system with health competencies having been devolved to the Autonomous Communities (ACs), disparities are conspicuous even within Spain, making the country even more interesting to study. More specifically, the obesity epidemic seems to be affected by the country's diverse population and severe income inequality, something captured by the random effects we include in the analysis as well as by the mapping of the BMI and Income distribution. The data confirm that the average self-reported BMI at the provincial level is higher in the southern part of Spain than in the northern part of the country (while income distribution presents the opposite pattern), which constitutes the main motivation of Chapter 4.

In this fourth Chapter, we decompose regional differentials in BMI between northern and southern Spanish regions. First, we decompose the observed average gap into the part attributed to differences in observable determinants of BMI (i.e. the endowments) and the part that is due to differences in the return to observable characteristics, using the classical Oaxaca-Blinder (OB) decomposition. Second, as long as important differences in BMI occur away from the average, we proceed with a distributional analysis by applying the Recentered Influence Function (RIF) regression and the corresponding decomposition (Firpo et al., 2009; Fortin et al., 2011). The RIF regression enables obtaining evidence along the unconditional distribution of BMI, which is especially important for the design of health and food policies. Our main contribution lies in the fact that we decompose regional differences in BMI along its unconditional distribution and this way, we are able to observe what happens at every part of the distribution and subsequently draw conclusions for the more interesting tails: the upper one (obesity, severe obesity) and the lower (underweight) where relationships might vary. The analysis is carried out separately by gender, as the underlying mechanisms that affect BMI and health outcomes in general appear to be different for women and men. Our findings indicate that the South to North gap in BMI is mostly driven by women, whereas it is lower and not statistically significant for men. The distributional analysis reveals that the South to North gap in BMI for Spanish women tends to increase over its unconditional distribution, with observable factors (especially schooling) making a growing contribution in explaining the differential across the quantiles of BMI. Given that the education gradient in obesity seems to be much stronger in women than in men (as in Devaux et al. (2011)), efforts aimed at improving (years of) schooling for women in the South would substantially mitigate differences in overweight and obesity between the two groups of regions. Such a policy intervention would additionally reduce differences in obesity-related diseases and/or improve health in general, inasmuch as obesity constitutes a key risk factor for many chronic conditions and health complications. These results appear to be robust to alternative scenarios dealing with missing information, BMI bias and alternative grouping of regions.

2 Decomposing income-related inequalities in obesity status, depth and severity in Spain.

2.1 Introduction

The world is threatened by a huge epidemic of obesity with data showing disappointing results in the coming years as well. According to recent statistical evidence, almost 39% of adults were overweight and 13% obese in 2016 (WHO, 2017). Paralleling the obesity crisis in the US, Europe is confronted by a similar obesity challenge and even though researchers and health authorities argue that they have advanced their understanding on this epidemic, obesity rates still remain in extremely alarming levels. Recent data shows that Spaniards rank among the first in obesity rates in Europe, while steep rises in obesity are projected for the following decade as well (OECD, 2017). These trends come with adverse health related consequences and are accompanied with direct and indirect costs. At the increased risk of morbidity we find several types of cancers (breast, prostate and colon cancers), hypertension, stroke, type 2 diabetes or respiratory problems (Rose, 1998). From the economic point of view, obesity is associated with excess health care expenditure and indirect costs in the form of foregone productivity and economic growth or obesity related absenteeism (Wolf and Colditz, 1998; Vazquez-Sanchez and Lopez-Alemany, 2002; Sander and Bergemann, 2003; Finkelstein et al., 2004; Raebel et al., 2004; Finkelstein et al., 2005; Borg et al., 2005; Von Lengerke et al., 2006; Nakamura et al., 2007; Muller-Riemenschneider et al., 2008; van Baal et al., 2008; Cawley and Meyerhoefer, 2012; Wolfenstetter, 2012; Mora et al., 2015; Specchia et al., 2015; Goettler et al., 2017).

In light of these evidence, there has been a push to study the link between obesity and its most important determinants. Researchers from several fields have approached the issue from different standpoints, however in this paper we will focus on the relationship between socioeconomic status (SES) and Body Mass Index

(BMI) and more specifically on the way income and BMI categories are associated to one another to describe the health status (proxied here by obesity measures) of the poor in comparison to the rich. The mechanism is clear, as there are various pathways by which income can relate to excess body weight and obesity. Better-off individuals for example, can afford healthier food, and high income is linked to higher health literacy which in turn, is positively related to health-promoting behaviors (i.e. healthy diet, regular physical activity, etc.). The mediating effect of stress in shaping the association between socio-economic position and obesity has been also obserevd in previous literature (Moore and Cunningham, 2012). The fact that obesity is predominant among individuals with low SES is also supported empirically, since a large volume of research (e.g. Lakdawalla and Philipson (2009); Jolliffe (2011) has identified that a socioeconomic gradient exists in obesity which is however different by gender (Ergin et al., 2012; Devaux and Sassi, 2013; Markwick et al., 2013). Many studies have considered possible gender differences in the association between socio-economic indices and obesity prevalence as diversity in both biological and social attributes has given rise to different health outcomes between genders. But when analysing obesity, there is a need to account for the long right tail of the BMI distribution (overweight, obese and morbidly obese statuses) to properly evaluate the cost of this epidemic, however just a handful of studies analyse what happens beyond the obesity threshold.

That said, the purpose of this study is twofold. We firstly want to examine both changes in obesity rates as well as in income-related inequality in obesity over time splitting by gender for the case of Spain. These are relevant questions not only for the health economics literature but also for the development economics literature, as there are very important policy implications linked to correctly understanding these relationships, i.e. the distribution of obesity between the female and male populations represents a useful proxy variable for measuring gender equality at the country level[1]. Second, we seek to understand which are the main factors that explain income inequality in obesity, which is their share in the overall inequality, as well as whether their contribution changes at different points in time, comparing our results for 2017 with 1997. For the measurement of obesity, we use distribution sensitive measures which allows us to have a clear idea on the relationships along the whole distribution of BMI.

[1] According to Ferretti and Mariani (2017) focusing on measuring gender inequality and in womens empowerment is essential to understand the determinants of gender gaps, evaluate policies and monitor countriess progress.

To tackle these goals we use cross-section and individual-level data gathered from several waves of the Spanish National Health Survey (ENSE) and the European Health Interview Survey (EHIS) spanning the last two decades in Spain, from 1997 to 2017. Specifically, this long time horizon covering diverse stages of the economic cycle with booms and busts is used to show trends in obesity prevalence as well as the evolution of the income-related inequality in obesity. In the decomposition analysis, we first focus on the latest available wave of ENSE- 2017 to exhibit which are the main contributing factors to the overall inequality in Spain, differentiating by gender and thence compare our results to the main contributing factors to the overall inequality in obesity back in 1997.We contribute on that we make use of distribution sensitive measures of obesity being this the first application with European data, so as to be able to compare these new evidence with the US findings. Analyzing the whole distribution of BMI is of extreme importance as there are increased health risks the further we move from the obesity threshold. We find that depth and severity of obesity is much greater for the poor than the non-poor in Spain, as well as that income inequality in obesity has decreased substantially during the past two decades for the case of women.

2.2 Literature Review

The literature has widely evidenced a positive income-health gradient (Johnston et al., 2009; Powdthavee, 2010; Carrieri and Jones, 2016; Frankenberg et al., 2016; Davillas et al., 2017). More relevant to our study, the existence of a negative gradient or relationship between SES (measured basically through the income or education) and BMI or obesity has been largely witnessed as well (Nayga, 1999; Chou et al., 2004; Lakdawalla and Philipson, 2009; Baum and Ruhm, 2009). In their seminal paper Sobal and Stunkard (1989) argue that obesity is associated with lower SES in women in high-income countries, whereas this pattern is not observed in men. Unlike, in low/ middle-income countries obesity can be positively associated with well-being. In a similar vein, McLaren (2007); Garcia-Villar and Quintana-Domeque (2009); Sassi et al. (2009) argue that the SES gradient in obesity is only observed in women in high-income countries.

Making use of the Concentration Index (CI), Zhang and Wang (2004) find that the inverse association between SES and obesity status is stronger in women than in men (and weaker in minority groups). Among other studies Ljungvall and Gerdtham (2010) analyse SES-related inequalities in obesity using longitudinal data for a Swedish cohort and find that among females, inequalities in obesity favor the rich, but the inequality declines over time. They attribute this finding mainly in increased obesity prevalence, because in absolute terms obesity has increased uniformly across income groups. They also find that income is the main driving factor behind obesity inequality. On the basis of absolute and relative inequality indexes, Devaux and Sassi (2013) observe large and persistent social inequalities in obesity and overweight by education level and socio-economic status in OECD countries, that are larger in women than men.

For the case of Spain, several studies report the existence of SES-related inequalities in obesity (Costa-Font and Gil, 2008; Costa-Font et al., 2010a; Rodriguez-Caro et al., 2016; Merino Ventosa and Urbanos-Garrido, 2016). For example, Costa-Font and Gil (2008) and Costa-Font et al. (2014) document persistent income-related inequalities in obesity in Spain. More specifically, examining cross-country trends in income inequalities in unhealthy lifestyles (obesity, smoking and alcohol intake) between Spain and the UK, they show that inequality in obesity appears to increase in Spain by 50% among females from 1987 to 2006. Merino Ventosa and Urbanos-Garrido (2016) contribute to the literature by providing complementary evidence to previous estimations regarding SES inequalities in obesity, using path analysis to disentangle the direct and indirect effects of SES on obesity. They use corrected concentration indices (CCI) to measure inequality and find that significant pro-rich inequality exists in obesity, particularly for women. They make use of dietary patterns, physical activity and sleep habits that act as mediator variables and find that the indirect effects of SES on obesity (those transmitted via mediator variables) are quite modest.

A limitation of these studies however (Rodriguez-Caro et al. (2016) being the exception) is that they focus on the prevalence of obesity and hence on the obesity threshold, neglecting the BMI distribution beyond this point and their implied inequalities. Certainly, a continuous BMI measure is converted into a discrete outcome using a standard cut-off point (i.e., BMI $\geqslant 30$) which is translated to a loss of important information (Jolliffe, 2004). However, it is well-known that the health risks associated with being obese are increasing even more at the top of the BMI distribution (Willett et al., 1999; Freedman et al., 2002; Zabarsky et al., 2018). That said, to avoid this loss of information there is a need to look beyond the obesity threshold (Madden, 2006). To overcome this issue, Jolliffe (2004, 2011) borrowed heavily from the poverty literature and slightly modified the Foster-Greer-Thorbecke (FGT) index, which was originally introduced by Foster et al. (1984) to measure poverty. Specifically, he sets up a class of overweight indices expressing the depth and severity of the problem as the excess BMI above the overweight threshold and the squared excess, respectively. This method addresses two important weaknesses of the prevalence measure. Firstly, excess BMI gap and squared BMI gap addresses the problem of too much emphasis placed at the selected threshold, i.e. these measures are more robust to measurement error near the threshold. Secondly, a distinction of the just slightly overweight or obese and the completely or morbidly obese is possible, which is more insightful in terms of public health policy recommendation. This analysis was further extended by Bilger et al. (2017) who contributes to the literature by combining the FGT measure with the standard concentration index (CI) to gain further insight on the SES gradient of obesity, when obesity is not only measured in prevalence terms, but when its depth and severity indices are also analyzed. They utilize the US NHANES data from 1971-2012 to evidence that income inequality in obesity prevalence has almost disappeared during these years; however, when distribution sensitive measures of obesity are analyzed, depth and severity of obesity seems to continue disproportionally affecting the poor. They continue with decomposing these FGT - concentration indices (FGT-CIs), to the contribution of the basic factors responsible for the overall inequality.

Following Bilger et al. (2017) and with a view to analyze the income-related inequality in obesity and at the same time not to limit our analysis only to obesity status but also investigate the often neglected distribution of BMI above the obesity threshold, we combine these two measures the following way; i.e. applying the FGT transformation to our variable of interest (obesity) and then calculating the CI of this transformed variable after ranking the observations according to our SES status variable of choice (income).

2.3 Method

As already mentioned, we use distribution-sensitive measures of obesity for the worse-off and the better-off individuals as in Jolliffe (2011) and Bilger et al. (2017). To achieve that, we start by using the modified FGT index as follows:

$$Y = \begin{cases} (BMI - c)^{\alpha} & if BMI \geqslant c \\ 0 & otherwise \end{cases} \tag{2.1}$$

where c is the obesity threshold and α is a parameter weighting the deviation above the obesity cut-off point. When α=0, Y provides a measure of obesity status indicating whether the individual is obese or not. When α =1, Y measures how far above the obesity threshold BMI of obese individuals are, providing a measure of depth of obesity, while in the case where α=2, Y yields a severity measure of obesity and is measured as the squared excess BMI over the obesity threshold. In this paper, we are interested in the distribution of the above three measures according to income [2] .

The next basic step of the analysis is the measurement of inequality in the distribution of the above three obesity measures. The concentration index (CI) has become the standard tool to quantify income-related inequalities in a health measurement (Kakwani, 1977; Kakwani et al., 1997; Wagstaff et al., 1991; Wagstaff and van Doorslaer, 2000; O'Donnell et al., 2008, 2016) and is frequently used in the obesity literature (Zhang and Wang, 2004; Costa-Font and Gil, 2008). This index relates the concentration of a health variable with the cumulative rank of the income distribution. It can be computed with individual-level data as follows:

$$CI = \frac{2}{N\mu}\sum_{i=1}^{N} h_i R_i - 1 \tag{2.2}$$

where h_i (i=1,..,N) is the obesity rate (or health variable) of the individual i, μ denotes its mean value and R_i is the relative income rank of the person i. Notice that the CI is a measure of relative inequality, so that a small change of everyone s obesity (health) leaves the CI index unchanged. The CI ranges between -1 and +1 which makes the comparison of inequality between years, groups as well as populations possible. In the case where the health variable of interest corresponds to a measure of ill health (i.e., obesity), negative values of the CI indicate that this variable is concentrated among the worse-off, which means that inequalities favor the high-SES individuals (pro-rich). A zero value of the index represents an equal distribution of obesity across income.

As a next step, we combine these two measures (i.e. the FGT and CI) as in Bilger et al. (2017) and compute FGT-CIs separately for the status, depth and severity of obesity. These inequality indexes are calculated, as usual, on the basis of a convenient regression formula in which a fractional rank variable is created (Kakwani et al., 1997; O'Donnell et al., 2008). Due to the fact that obesity prevalence is a binary measure and because of problems associated with bounded variables, Wagstaff (2005) and Erreygers (009a) propose different correction methods with a view to overcome these issues[3]. Both corrections satisfy the mirror condition

[2] The value of these measures is really important and can be depicted for instance, in the situation where an already obese individual gains weight and increases her BMI. In such a case, the weight gain affects the individuals health, something that is not reflected in the prevalence measure of obesity, in contrast to the other two measures.

[3] These drawbacks have been previously discussed in relevant literature and include 1) problems associated to the minimum and maximum values of the CI calculated on the basis of a binary variable that depend upon the mean of this variable making comparison of populations with different mean health levels problematic (Erreygers, 009a), 2) the non-satisfaction of the mirror property (Clarke

(Erreygers and Van Ourti, 2011), but following (Costa-Font et al., 2014) we apply the Erreygers correction [4] for the CI of obesity status which is calculated as follows:

$$CCI = \frac{4\mu}{h^{max} - h^{min}} * CI \tag{2.3}$$

where h^{max} and h^{min} denote the maximum and minimum values or bounds of the prevalence of obesity, respectively, and CI is the concentration index calculated as indicated.

It should be noted that in contrast to the CI for obesity status, the CI for the depth and severity of obesity is not only affected by the rank of the income distribution of the individuals that exceed the threshold, but also from the extent in which this threshold is exceeded. Given the non-linearity of the indices (problem of excess of zeros or non-obese and highly right-skewed distribution of positive values) a Two Part Model (TPM) approach is used (e.g. Duan et al. (1983); Pohlmeier and Ulrich (1995)), as in Bilger et al. (2017). This is the appropriate method for such analyses, as the values that lie below the obesity threshold do not contribute to the FGT measures (i.e. the zero values taken are true zeros). We proceed and use the van Doorslaer and Koolman (2004) approximation of the Wagstaff et al. (2003) decomposition of the CI as follows:

$$CI_Y = \sum_{k=1}^{k} \frac{\partial \bar{\epsilon}(Y|X)}{\partial x_k} \frac{\bar{x}_k}{\mu} CI_k + GC_\varepsilon \tag{2.4}$$

CI_Y is the CI of the variable Y, CI_k is the CI of the factor χ_k, μ is the sample average of Y, $\bar{x}_k$ the sample average of factor x_k , $\frac{\partial \bar{\epsilon}(Y|X)}{\partial x_k}$, the average marginal effect of factor x_k on Y obtained from the TPM estimates and finally GC_ε the generalized CI of the regression residuals.

We continue with the decomposition of the estimated FGT-CIs into their main contributors, that is, factors that are correlated both with obesity and income. The statistical package available from Bilger et al. (2017) in Stata allows us to decompose

et al., 2002; Erreygers et al., 2012), 3) the arbitrary nature of the index in the cases where the health variable has a qualitative nature, 4) the fact that when the health variable is binary, the limits of the CI are not necessarily -1 and +1 (Wagstaff, 2005)

[4] Erreygers corrected CI has been found insensitive to any feasible equal addition to the health variable (Erreygers and Van Ourti, 2011)

the overall income-related inequality in obesity into the contribution of each factor by means of the TPM which determines how our control variables are associated with overall income inequality. Regarding the specification of the first part of the TPM, we apply a Logit model as in Bilger et al. (2017). For the second part, we choose a generalized linear model (GLM, Nelder and Wedderburn (1972)) as this family offers many alternatives to the linear model that are suitable to skewed data[5] .

2.4 Data & Model Specification

We base our analysis on combining several waves of cross-section data gathered from the ENSE (Spanish National Health Survey) and from the Spanish version of the EHIS (European Health Interview Survey) ranging from 1997 to 2017. The National Health Survey of Spain (ENSE) is a periodic study conducted by the Ministry of Health, Consumption and Social Welfare and collects health information on the entire population on health status, personal, social and environmental determinants of health and the use and access to health services. The European Health Interview Survey (EHIS) has very similar characteristics and aims at measuring on an harmonized basis and with a high degree of comparability among Member States (MS) the health status (including disability), health determinants (including environment), use and limitations in access to health care services of the EU citizens, as well as background variables on demography and socio-economic status. Specifically, with this 20 years of data we calculate trends in obesity status, depth and severity as well as the evolution of income related inequalities for these three obesity indexes. Weight and height of the respondents is self-reported [6] and this information is used to calculate the individual BMI according to the usual formula, i.e. (weight in kilograms divided by the square of height in meters). We use household income to compute equivalent net income (in logs) which is used as our rank variable [7].

For the purpose of the decomposition analysis we first analyze the latest available

[5]We used the Box- Cox (Box and Cox, 1964) and Park (Park, 1966; Manning and Mullahy, 2001) tests to determine the GLM link function and distribution family.

[6]The correcting formulas for Spain provided by Gil and Mora (2011) are used in order to make sure the data do not suffer from significant measurement error.

[7]As household earnings are measured as a categorical variable with 8 response categories in our datasets, we employ an interval regression model based on information of the head of the household (age, gender, education, SES and region of residence) to obtain a continuous income measure. Once predicted, we divide it by an equivalence factor (equal to the number of household members elevated to 0.5) to adjust for differences in household size.

data for 2017 and subsequently compare them to the 1997 data aiming to draw conclusions for a sufficiently long time span where obesity grows quite a lot and very importantly to compare the contribution of each inequality determinant. Concerning the 2017 wave, interviews on 29,195 individuals are carried out and correspond to 23,089 adults (15 and over) and 6,106 minors (0-14 years). We discard individuals with no information on either weight or height (649 obs.). A number of observations (5,570 obs.) have missing income information which leaves us with 17,519 observations in total. There are 6,400 individuals interviewed in 1997, out of which we discard 860 due to the fact that they have missing information on either weight or height. We also discard the individuals with missing information on income which leaves us to a sample of 4,331 persons. We decompose the previously estimated FGT-CIs of obesity differentiating by gender. As our basic control variables (Table 2.7 Appendix) we use dummies of age cohorts, marital status, employment status, completed level of education (compulsory education, secondary post-compulsory education, tertiary education), equivalent net monthly income (in logs), sedentarism, as well as dummies of daily smoking and alcohol, fruit, pasta, vegetable, legume and meat consumption, controlling also for region of residence.

2.5 Results

2.5.1 Trends in Prevalence, Depth and Severity, 1997-2017

Figure 2.2 (Appendix) presents the evolution of the three obesity measures (pooled sample and by gender) demonstrating an increasing trend in their evolution from 1997 until 2017. The rise is more pronounced in depth and severity compared to obesity status and larger for women than men when observing the distributional measures, witnessing that while roughly the same amount of women and men are obese, women lie further beyond the obesity threshold than men.

2.5.2 Trends in Concentration indices of Status, Depth and Severity, 1997-2017

Table 2.1 exhibits trends of income-related inequality in status, depth and severity of obesity over the years (measured by the CI (CCI for status) separately for women and men). First of all, as can be seen all measures are negative and statistically significant evidencing that obesity disproportionately affects the poor in Spain. Secondly, although the prevalence of obesity rises continually across the years, our results show that income-related inequality has remarkably decreased over the years under examination in all three measures of obesity (-0,10 vs -0,08, -0,22 vs -0,16 and -0,28 vs -0,18 respectively). Interestingly, when differentiating by gender, we evidence that the decrease in inequalities over time are basically driven by women and is sharper for severity and depth compared to the status measure.

Table 2.1: Trends in the CI

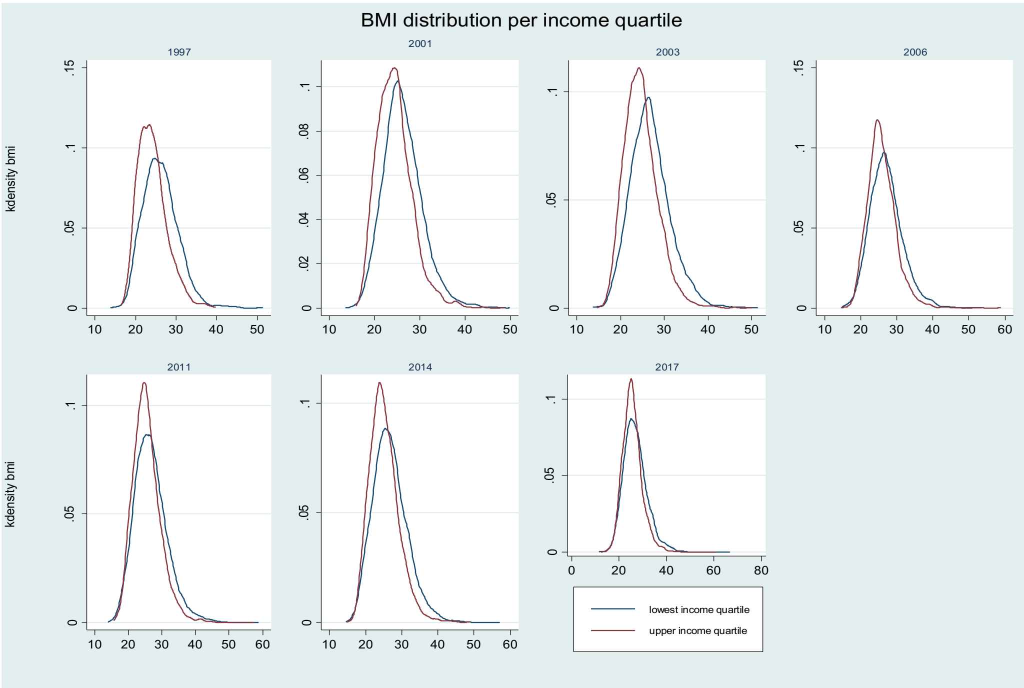
	Status*			**Depth**			**Severity**		
	Females	*Males*	*Pooled sample*	*Females*	*Males*	*Pooled sample*	*Females*	*Males*	*Pooled sample*
1997	-0,16	-0,05	-0,10	-0,27	-0,13	-0,22	-0,30	-0,12	-0,26
2001	-0,11	-0,06	-0,08	-0,14	-0,08	-0,20	-0,15	-0,13	-0,24
2003	-0,11	-0,03	-0,08	-0,20	-0,09	-0,16	-0,21	-0,07	-0,17
2006	-0,14	-0,06	-0,09	-0,26	-0,11	-0,16	-0,25	-0,09	-0,15
2011	-0,11	-0,07	-0,09	-0,20	-0,13	-0,17	-0,24	-0,15	-0,21
2014	-0,12	-0,05	-0,09	-0,17	-0,13	-0,16	-0,16	-0,19	-0,18
2017	-0,12	-0,05	-0,08	-0,19	-0,13	-0,16	-0,18	-0,17	-0,18

**: The CI for obesity status is corrected using the Erreygers correction formula.*

Overall, data shows that obesity is concentrated among the poor in Spain since the CI is always negative and statistically significant, but the degree of inequality has a declining trend, even though obesity prevalence, as well as depth and severity increase over the years.

A thorough understanding of trends in the relationship between SES and obesity will provide useful insights for developing effective intervention programs and policies. In order to have a general picture of the shape of the BMI distribution for different levels of income, Figure 2.1 which depicts the BMI distribution of individuals belonging to the lowest income quantile versus the BMI distribution of the ones belonging to the highest income quantile is presented for the whole period analyzed. The results clearly indicate that there are important differences in the shape of the distributions between the two populations, with the BMI distribution of the worse-off being less-peaked indicating greater spread in the tails and as expected, shifted towards the right. We also notice important changes in the distribution of BMI between the poor and non-poor over time. Specifically, the shape of the two distributions appears to become more and more similar over time, resulting in just slight differences between the two populations in 2017 (distribution of the worse-off has a slightly higher dispersion than the one of the better-off individuals) in highly contrast with 1997 where differences in shapes are much more pronounced.

Figure 2.1: Distribution of BMI for poor vs non-poor over time



2.5.3 Decomposition Results: 2017

For more insight, we proceed with the decomposition of total inequality in obesity status, depth and severity. Initially, we focus on the latest data available, i.e. the 2017 wave. As we notice important differences between women and men, we report the results differentiating by gender. In general, overall income inequality in obesity status is much higher for women (-0,119) than men (-0,053), in line with Bilger et al. (2017). Surprisingly, when we move from obesity prevalence to depth and severity, inequalities for women increase with a lower pace than in men (from -0.12 to -0.19 and -0.18 vs from 0.05 to -0.13 and -0.17, respectively). In more detail, we analyze income-related inequality in obesity by breaking down the estimated FGT-CIs into their main conditioning factors and check the contribution of each control to the overall inequality. As a last step of our analysis, we compare these results with the results obtained by the 1997 wave (detailed decomposition results for 1997 are shown in the tables 2.8 to 2.11 of the Appendix).

Obesity Status

In Tables 2.2 and 2.3 we present the detailed results that correspond to the decomposition of the income inequality in obesity status (measured by the CCI) separately for women and men in 2017. The contribution of the controls/factors to the overall inequality can be negative or positive. A negative (positive) contributor is interpreted as if such factor was evenly distributed across income where the corresponding CCI would fall (rise) in the magnitude of the coefficient size. For the case of women, having tertiary education, equivalent income and being employed are the basic negative contributors to the income inequality for obesity status, with income constituting the most important factor with the largest contribution to the overall inequality (66.3%). The negative contribution of education is expected, as higher education is concentrated among the better-off individuals and higher BMI is associated with low educational attainment, however its importance is strongly modest (roughly 9%). The same applies for the role of equivalent income as well.

Belonging to the oldest age group, being inactive and tertiary education are important negative contributing factors to the income inequality in obesity status for the case of men. The negative contribution of income is large (37.6%) and significant as in the female case, though with a slightly lower contribution to the overall inequality. Unlike the female case, the contribution of daily smoking and sedentarism is significant and positive. Indeed, high income individuals are usually occupied in positions that do not require physical activity and even though sedentarism is positively associated with higher BMI levels, its contribution to the overall inequality is positive.

Table 2.2: Decomposition of the income related inequality in obesity status-Women.

Women, 2017

	elasticity	concentration	contribution	%
Age2	0,03	0,05	0,00	0,37
Age3	0,07	0,04	0,00	0,73
Age4	0,07	0,05	0,00	0,83
Age5	0,24	-0,15	-0,04	8,55
Married	0,05	0,20	0,01	**2,56**
Widowed	0,02	-0,14	0,00	0,62
Divorced/separated	0,00	-0,07	0,00	0,01
Demographics				***13,66***
Loginc	-1,98	0,14	-0,28	66,28
Secondary	-0,05	0,11	-0,01	1,48
Tertiary	-0,10	0,31	-0,03	**7,26**
Employed	-0,04	0,39	-0,02	**4,23**
Inactivity	-0,03	-0,22	0,01	1,52
SES				***80,77***
Dailysmoker	-0,05	0,02	0,00	0,28
Dailydrink	0,00	0,05	0,00	0,01
Sedentarism	0,08	-0,01	0,00	0,14
Dailycarne	0,01	0,01	0,00	0,01
Dailypasta	0,00	-0,03	0,00	0,02
Dailyfruits	0,01	0,09	0,00	0,21
Dailyveg	0,04	0,19	0,01	2,07
Dailyleg	0,00	0,00	0,00	0,08
Lifestyle				***2,82***
ccaa_2	-0,01	0,00	0,00	0,01
ccaa_3	0,00	0,04	0,00	0,03
ccaa_4	-0,01	0,01	0,00	0,01
ccaa_5	0,00	-0,03	0,00	0,02
ccaa_6	-0,02	0,00	0,00	0,02
ccaa_7	-0,02	0,01	0,00	0,03
ccaa_8	0,00	0,00	0,00	0,00
ccaa_9	-0,02	0,03	0,00	0,18
ccaa_10	0,00	-0,01	0,01	2,13
ccaa_11	0,00	-0,02	0,00	0,01
ccaa_12	0,00	-0,04	0,00	0,03
ccaa_13	-0,01	0,05	0,00	0,11
ccaa_14	-0,01	-0,01	0,00	0,01
ccaa_15	0,00	0,04	0,00	0,02
ccaa_16	-0,01	0,07	0,00	0,15
ccaa_17	0,00	0,02	0,00	0,00
ccaa_18	0,00	-0,01	0,00	0,00
ccaa_19	0,00	-0,01	0,00	0,00
Region				***2,75***
CCI:-0,119				

Note: Statistically significant contributors in bold.

Table 2.3: Decomposition of the income related inequality in obesity status-Men.

Men, 2017

	elasticity	concentration	contribution	%
Age2	0,06	0,07	0,00	2,83
Age3	0,10	0,01	0,00	0,90
Age4	0,09	0,03	0,00	1,80
Age5	0,11	-0,12	-0,01	**8,54**
Married	0,07	0,04	0,00	1,79
Widowed	0,01	-0,01	0,00	0,03
Divorced/separated	0,00	0,02	0,00	0,02
Demographics				***15,90***
Loginc	-0,40	0,15	-0,06	**37,62**
Secondary	-0,02	0,03	0,00	0,50
Tertiary	-0,06	0,35	-0,02	**14,29**
Employed	0,01	0,42	0,01	3,51
Inactivity	0,09	-0,21	-0,02	**12,50**
Ses				***68,42***
Dailysmoker	-0,03	-0,09	0,00	**1,60**
Drink	0,00	0,02	0,00	**0,01**
Sedentarism	0,05	0,08	0,00	2,60
Dailycarne	0,02	0,04	0,00	0,52
Dailypasta	-0,01	-0,01	0,00	**0,05**
Dailyfruits	-0,06	0,12	-0,01	4,44
Dailyveg	0,03	0,11	0,00	1,85
Dailyleg	0,00	-0,01	0,00	2,39
Lifestyle				***13,47***
ccaa_2	-0,01	0,03	0,00	0,17
ccaa_3	0,00	0,02	0,00	0,03
ccaa_4	-0,01	0,02	0,00	0,06
ccaa_5	0,00	-0,03	0,00	0,05
ccaa_6	-0,02	-0,02	0,00	0,16
ccaa_7	-0,02	0,00	0,00	0,01
ccaa_8	0,00	-0,02	0,00	0,05
ccaa_9	-0,02	0,05	0,00	0,68
ccaa_10	0,00	-0,01	0,00	0,01
ccaa_11	0,00	-0,06	0,00	0,07
ccaa_12	0,00	-0,02	0,00	0,04
ccaa_13	-0,01	0,07	0,00	0,40
ccaa_14	-0,01	-0,01	0,00	0,05
ccaa_15	0,00	0,05	0,00	0,06
ccaa_16	-0,01	0,06	0,00	0,36
ccaa_17	0,00	0,02	0,00	0,01
ccaa_18	0,00	-0,01	0,00	0,01
ccaa_19	0,00	0,00	0,00	0,00
Region				***2,21***
CCI : -0,054				

Note: Statistically significant contributors in bold.

Depth and severity of obesity

We proceed with the decomposition of our FGT-CIs for depth and severity of obesity into their determinants across the income distribution (shown in the following Tables 2.4 and 2.5) to see how relationships change when we take into account how far above the threshold obese individuals are. In more detail, the CI_k measures income inequality in the factors themselves, η_k is the elasticity of the FGT-CIs obesity measure with respect to factor and quantifies the association between such factors and the obesity measure. Last, both tables exhibit the contribution of each factor to the overall FGT-CI in values as well as in percentages. Again, significant differences are observed between women and men, leading to an analysis differentiated by gender.

For women, tertiary education and income are the main negative contributors with the contribution of education becoming now larger than the direct effect of income. Being married and having a sedentary job are the basic positive contributors to the overall income inequality in depth as well as severity of obesity. Inactivity also contributes positively to the overall inequality in depth and when moving from depth to severity its contribution becomes even stronger. The main negative contribution comes in principle from tertiary education and income for the case of men, with inactivity being also negative and statistically significant. Sedentarism is statistically significant and positive both for depth and severity, but being married has a negative contribution unlike the case of women.

In general, the main drivers of the differences between the two genders are marital status and employment status as being married has a positive (negative) contribution for women (men) while inactivity contributes positively for women and negatively for men. On the other hand, tertiary education is the most significant negative contributor to the overall inequality for all measures of obesity and both genders, with the contribution becoming stronger and stronger as one moves form status to depth and severity.

Table 2.4: Decomposition of the FGT-CI of depth and severity of obesity-Women.

Women, 2017

		Depth			Severity		
Variable	CI_k	η_k	$CI_{Y\,k}$	%	η_k	$CI_{Y\,k}$	%
Age group(reference : age1)							
Age2	0,052	0,060	0,003	**1,10**	0,044	0,002	**0,58**
Age3	0,103	0,121	0,012	**4,41**	0,125	0,013	**3,25**
Age4	0,089	0,072	0,006	**2,27**	0,040	0,004	0,90
Age5	-0,105	0,130	-0,014	4,83	0,061	-0,006	1,62
Married	0,089	0,119	0,011	**3,75**	0,127	0,011	**2,86**
Widowed	-0,152	0,011	-0,002	0,59	0,014	-0,002	0,54
Divorced/separated	-0,189	0,010	-0,002	0,67	0,007	-0,001	0,33
Education (reference: Primary)							
Secondary	0,020	-0,150	-0,003	1,06	-0,207	-0,004	1,05
Tertiary	0,472	-0,199	-0,094	**33,23**	-0,243	-0,115	**28,99**
Employed	0,190	-0,014	-0,003	0,94	-0,191	-0,036	9,17
Inactivity	-0,106	-0,100	0,011	**3,75**	-0,860	0,091	**23,04**
Loginc	0,046	-1,680	-0,077	**27,34**	-1,067	-0,049	**12,41**
Dailysmoking	0,008	-0,098	-0,001	**0,28**	-0,116	-0,001	0,23
Drink	-0,057	0,489	-0,028	9,86	0,590	-0,034	8,50
Sedentarism	0,030	0,190	0,006	**2,02**	0,295	0,009	**2,24**
Dailymeat	0,015	0,012	0,000	0,06	0,008	0,000	0,03
Dailypasta	-0,120	-0,018	0,002	0,76	-0,032	0,004	0,97
Dailyfruit	0,047	-0,042	-0,002	0,70	-0,113	-0,005	1,34
Dailyvegetables	0,087	0,019	0,002	0,58	0,024	0,002	0,53
Dailylegumes	-0,130	-0,001	0,000	0,03	-0,001	0,000	0,03
cc_aa2	0,087	-0,012	-0,001	0,07	-0,020	0,000	0,05
cc_aa3	0,099	0,004	0,000	0,03	-0,002	0,000	0,01
cc_aa4	0,102	-0,012	-0,001	0,09	-0,018	0,000	0,06
cc_aa5	-0,206	0,005	-0,001	0,07	0,008	0,000	0,05
cc_aa6	0,048	-0,004	0,000	0,01	0,003	0,000	0,00
cc_aa7	0,015	-0,048	-0,001	0,05	-0,076	0,004	0,04
cc_aa8	-0,055	0,001	0,000	0,00	-0,008	0,000	0,01
cc_aa9	0,105	-0,018	-0,002	0,13	-0,039	0,001	0,13
cc_aa10	-0,078	-0,001	0,000	0,00	-0,003	0,000	0,01
cc_aa11	-0,196	-0,028	0,006	0,39	-0,044	0,001	0,27
cc_aa12	-0,082	-0,014	0,001	0,08	-0,031	0,000	0,08
cc_aa13	0,186	-0,014	-0,003	0,18	-0,019	0,000	0,11
cc_aa14	-0,065	-0,026	0,002	0,12	-0,056	0,001	0,12
cc_aa15	0,286	0,002	0,001	0,05	0,006	0,000	0,05
cc_aa16	0,202	-0,021	-0,004	0,30	-0,037	0,001	0,24
cc_aa17	0,119	-0,013	-0,001	0,11	-0,026	0,000	0,10
cc_aa18	-0,271	0,002	0,000	0,03	-0,003	0,000	0,03
cc_aa19	-0,150	-0,002	0,000	0,03	-0,009	0,000	0,04
Residuals			0,016			0,016	
Overall CI			**-0,185**			**-0,180**	

CI_k , concentration index of factor k.

η_k , elasticity of the FGT measure Y with respect to factor k.

$CI_{Y\,k}$, contribution made by factor k to the overall FGT-CI

CI, concentration index.

Note: Statistically significant contributors in bold.

Table 2.5: Decomposition of the FGT-CI of depth and severity of obesity-Men.

Men, 2017

		Depth			Severity		
Variable	CI_k	η_k	$CI_{Y\,k}$	%	η_k	$CI_{Y\,k}$	%
Age group(reference : age1)							
Age2	0,082	0,101	0,008	**4,09**	0,117	0,009	**3,65**
Age3	0,030	0,165	0,005	**2,55**	0,197	0,007	**2,84**
Age4	0,012	0,110	0,001	**0,51**	0,094	0,001	0,41
Age5	-0,075	0,077	-0,005	2,55	0,043	-0,003	1,22
Married	0,013	-0,128	-0,001	**0,51**	-0,357	-0,004	**1,62**
Widowed	-0,014	-0,013	0,000	0,05	-0,038	0,001	0,20
Divorced/separated	0,050	-0,018	-0,001	0,46	-0,050	-0,002	**0,81**
Education (reference: Primary)							
Secondary	-0,003	-0,005	0,000	0,05	-0,040	0,000	0,04
Tertiary	0,510	-0,119	-0,061	**31,00**	-0,134	-0,068	**27,56**
Employed	0,162	0,006	0,001	0,51	-0,009	-0,001	0,41
Inactivity	-0,108	0,124	-0,013	**6,64**	0,139	-0,015	**6,08**
Loginc	0,045	-1,520	-0,069	**35,24**	-2,100	-0,096	**38,91**
Dailysmoking	-0,084	-0,040	0,003	**1,53**	-0,030	0,002	**0,81**
Drink	-0,050	-0,023	0,001	0,51	-0,023	0,001	0,41
Sedentarism	0,067	0,184	0,012	**6,13**	0,270	0,018	**7,30**
Dailymeat	0,026	0,027	0,001	0,36	0,023	0,001	0,24
Dailypasta	-0,030	-0,007	0,000	0,10	0,003	0,000	0,04
Dailyfruit	0,033	-0,030	-0,001	0,51	-0,006	0,000	0,08
Dailyvegetables	0,057	0,039	0,002	1,02	0,042	0,002	0,81
Dailylegumes	-0,098	0,001	0,000	0,05	0,002	0,000	0,08
cc_aa2	0,129	-0,009	-0,001	0,25	-0,012	-0,002	0,29
cc_aa3	0,106	0,004	0,000	0,09	-0,002	0,000	0,04
cc_aa4	0,112	-0,006	-0,001	0,15	-0,006	-0,007	1,27
cc_aa5	-0,240	-0,011	0,003	0,56	-0,010	0,003	0,46
cc_aa6	-0,101	-0,001	0,000	0,03	0,005	-0,001	0,09
cc_aa7	0,009	-0,025	0,000	0,05	-0,033	0,000	0,05
cc_aa8	-0,081	0,010	-0,001	0,17	0,018	-0,001	0,27
cc_aa9	0,128	-0,050	-0,006	1,37	-0,041	-0,005	0,95
cc_aa10	-0,069	-0,012	0,001	0,18	-0,028	0,002	0,35
cc_aa11	-0,236	0,011	-0,003	0,58	0,014	-0,003	0,62
cc_aa12	-0,099	-0,009	0,001	0,19	-0,014	0,001	0,25
cc_aa13	0,191	-0,010	-0,002	0,40	0,001	0,000	0,04
cc_aa14	-0,046	-0,017	0,001	0,16	-0,019	0,001	0,16
cc_aa15	0,307	0,005	0,001	0,29	0,008	0,002	0,44
cc_aa16	0,206	-0,019	-0,004	0,84	-0,015	-0,003	0,58
cc_aa17	0,106	-0,007	-0,001	0,15	-0,011	-0,001	0,20
cc_aa18	-0,213	0,003	-0,001	0,12	0,010	-0,002	0,38
cc_aa19	0,031	0,005	0,000	0,03	0,008	0,000	0,04
Residuals			-0,005			-0,011	
Overall CI			**-0,128**			**-0,170**	

CI_k , concentration index of factor k.

η_k , elasticity of the FGT measure Y with respect to factor k.

$CI_{Y\,k}$, contribution made by factor k to the overall FGT-CI

CI, concentration index.

Note: Statistically significant contributors in bold.

2.5.4 Comparison of the drivers of inequality between two points in time.

So as to be able to understand the possible changes in the drivers of income inequality in obesity over the years, we proceed by comparing our results for 2017 and 1997 [8]. That is, we compare inequalities between the two years, however not performing a decomposition analysis of changes in FGT-Cis inequalities by means of the Oaxaca-Blinder decomposition. We document (see Table 2.6) that while inequalities remained quite stable between these two time periods for men, differences for women are huge.

Starting with the case of women, the direct effect of income is the main contributor to the total inequality in obesity status for both years. The difference between the two time periods is on the role of employment status, as being employed is statistically significant and negative in 2017 unlike 1997 where it contributes positively to the income inequality in obesity. Moving on to depth and severity, we observe that the largest contribution to the total inequality comes from SES (education, income, employment status) for 2017, whereas it is more evenly divided between SES and demographic characteristics in 1997.

For the case of men, the largest contributors to the negative socioeconomic inequality in obesity status are income and education for 2017. In contrast, in 1997 the direct effect of income prevails, while the role of education is way less important and offset by demographic characteristics. The basic difference between the two time periods in depth and severity of obesity, lies to the fact that demographic factors constitute an important part of the negative contribution to the total inequality in 1997 along with SES status, in contrast to 2017 where SES controls are by far the more important contributors to the total inequality.

[8] Due to some incompatibilities in the questionnaires between the two time periods, there is no available information on the food habits of the individuals for 1997. That said, the contribution of lifestyle characteristics for 1997 corresponds to the effect of smoking, drinking and sedentary behavior to the total inequality.

Table 2.6: % Contribution of groups of variables to the overall SES inequality

	Women			Men		
				1997		
Variables	**Status**	**Depth**	**Severity**	**Status**	**Depth**	**Severity**
Demographics	17,16	29,68	33,14	22,25	21,96	6,46
Income	61,32	16,46	9,65	42,37	27,69	17,59
Education	6,28	24,14	17,32	8,33	7,06	11,78
Employment	9,05	5,78	12,42	17,65	24,08	28,33
Lifestyle	4,24	9,47	10,43	6,43	7,49	9,71
Region	1,95	14,47	17,04	2,98	11,72	12,25
CCI	***-0,157***	***-0,270***	***-0,300***	***-0,046***	***-0,133***	***-0,125***
				2017		
Demographics	13,66	17,62	10,08	15,90	10,73	10,74
Income	66,28	27,34	12,41	37,62	35,24	38,91
Education	8,74	34,29	30,04	14,79	31,05	27,60
Employment	5,75	4,69	32,22	16,01	7,15	6,49
Lifestyle	2,82	14,30	13,87	13,47	10,21	9,77
Region	2,75	1,75	1,39	2,21	5,62	6,49
CCI	***-0,119***	***-0,185***	***-0,180***	***-0,054***	***-0,128***	***-0,170***

2.6 Conclusion

Our results indicate, that even though all three measures of obesity have been increasing over the past 20 years among the Spanish population, income related inequality in obesity status, depth and severity has a clear decreasing trend, especially among women. While the decreasing pattern in obesity status inequalities has been previously detected in other European countries (Costa-Font et al., 2014), this work is the first, as far as we are concerned, to evidence an decreasing trend of inequalities in depth and severity of obesity in an EU country.

This finding is in line with results from the US and more specifically with Bilger et al. (2017) who observe that the more severe cases of obesity are becoming more and more equally distributed according to income over the years. One of the explanations given is that these results insinuate a deeper change in the basic determinants of individuals' obesity risk occurred in the last decades. In more detail, it might be the case that individual obesity risk is no more linked only to individual attributes and that a non negligible part of the effect may come from environmental influences (obesogenic environment) as well.

The fact that the declining trend of income related inequalities in obesity status, depth and severity is mostly driven by women, may be the result of changes in the social standards; that is women might be more affected by the societal attitude towards obesity than men and try to pursue a thinner silhouette in the recent years compared to the past. Based on our results, we also witness that income related disparities in depth and severity of obesity tend to be noticeably larger than the ones in obesity status both for women and men. This is in line with similar findings arguing that highest levels of BMI are often observed among the worse-off and poorly educated and more generally among those in disadvantaged socio-economic circumstances.

Decomposing the income inequality of FGT-CIs obesity measures into its main determinants, we observe that for obesity status, income is the most important single contributor to the overall inequality both for women and men. As we move to depth and severity, income continues to be very important for both genders, but education becomes the most important contributing factor to the overall inequality for women. The protective role of education in health outcomes has been observed in previous literature as well Cutler et al. (2015). Furthermore, when comparing the two time periods, we notice that the basic difference lies to the fact that in 1997, SES inequality in obesity is a result of the confounding factors of demographics and SES, in contrast to 2017 where SES controls are by far the most important contributors to the total inequality in obesity. We infer that overall, income related inequalities in obesity are still a reality in Spain especially for depth and severity of obesity, something that could aggravate even more the socioeconomic gradient in health.

2.7 Appendix

Table 2.7: Description of variables

Variables	Description
Dependent Variables	
Obesity	1 when BMI >=30, 0 otherwise
Depth of obesity	excess BMI beyond the obesity threshold
Severity of obesity	squared excess BMI beyond the obesity threshold
Independent Variables	
Demographics	
Age1	1 when aged 18-35 , 0 otherwise
Age2	1 when aged 36-45, 0 otherwise
Age3	1 when aged 46-55, 0 otherwise
Age4	1 when aged 55 65, 0 otherwise
Age5	1 when aged more than 65, 0 otherwise
Married	1 when married, 0 otherwise
Widowed	1 when widowed, 0 otherwise
Divorced/separated	1 when divorced/separated, 0 otherwise
SES	
Primary	1 when primary education is the highest education level achieved, 0 otherwise
Secondary	1 when secondary education is the highest education level achieved, 0 otherwise
Tertiary	1 when tertiary education is the highest education level achieved, 0 otherwise
Employed	1 when employed, 0 otherwise
Inactivity	1 when inactive, 0 otherwise
Loginc	equivalent net monthly income in logs
Lifestyle	
Dailysmoking	1 if daily smoker, 0 otherwise
Drink	daily alcohol consumption
Sedentarism	1 if working in a sedentary job, 0 otherwise
Dailymeat	1 if consumes meat more than 4 times per week, 0 otherwise
Dailypasta	1 if consumes pasta more than 4 times per week, 0 otherwise
Dailyfruit	1 if consumes fruit more than 4 times per week, 0 otherwise
Dailyvegetables	1 if consumes vegetables more than 4 times per week, 0 otherwise
Dailylegumes	1 if consumes legumes more than 4 times per week, 0 otherwise
Region	
cc_aa	Autonomous Community the individual belongs (17 regions: cc_aa1-cc_aa17)

Table 2.8: Decomposition of the income related inequality in obesity status-Women.

Women, 1997				
	elasticity	concentration	contribution	%
Age2	0,102	0,120	0,012	1,92
Age3	0,133	0,023	0,003	0,48
Age4	0,150	-0,112	-0,017	**2,61**
Age5	0,153	-0,265	-0,041	**6,33**
Married	0,271	0,123	0,033	5,20
Widowed	0,019	-0,208	-0,004	0,62
Divsep	-0,001	-0,025	0,000	0,00
Demographics				**17,16**
Loginc	-2,221	0,177	-0,394	**61,32**
Secondary	-0,093	0,206	-0,019	2,99
Tertiary	-0,080	0,263	-0,021	**3,29**
Employed	0,158	0,288	0,045	**7,07**
Inactivity	-0,046	-0,278	0,013	1,98
SES				**76,65**
Dailysmoker	-0,082	0,176	-0,014	2,24
Drink	-0,021	0,243	-0,005	0,78
Sedentarism	0,077	0,102	0,008	1,22
Lifestyle				**4,24**
cc_aa2	0,001	0,010	0,007	1,02
cc_aa3	0,003	0,047	0,000	0,02
cc_aa4	-0,009	0,018	0,000	0,02
cc_aa5	-0,011	-0,031	0,000	0,05
cc_aa6	0,000	-0,002	0,000	0,03
cc_aa7	-0,016	-0,021	0,000	0,05
cc_aa8	-0,016	-0,032	0,001	0,08
cc_aa9	-0,051	0,035	-0,002	0,28
cc_aa10	-0,015	-0,038	0,001	0,09
cc_aa11	0,011	-0,019	0,000	0,03
cc_aa12	-0,012	-0,043	0,000	0,08
cc_aa13	-0,006	0,157	-0,001	**0,14**
cc_aa14	0,009	-0,015	0,000	0,02
cc_aa15	0,003	0,020	0,000	0,01
cc_aa16	0,001	0,013	0,000	0,00
cc_aa17	-0,009	0,016	0,000	0,02
Region				**1,95**
CCI= -0,157				

Note: Statistically significant contributors in bold.

Table 2.9: Decomposition of the income related inequality in obesity status-Men.

Men,1997				
	elasticity	concentration	contribution	%
Age2	0,103	0,061	0,006	1,40
Age3	0,100	0,063	0,006	**1,40**
Age4	0,185	-0,127	-0,024	**5,24**
Age5	0,162	-0,219	-0,035	**7,89**
Married	0,355	-0,079	-0,028	**6,22**
Widowed	0,002	-0,038	0,000	0,02
Divsep	-0,042	-0,009	0,000	0,08
Demographics				**22,25**
Loginc	-1,154	0,165	-0,191	**42,37**
Secondary	-0,051	0,230	-0,012	2,60
Tertiary	-0,087	0,295	-0,026	**5,73**
Employed	-0,021	0,380	-0,008	1,80
Inactivity	-0,290	-0,246	0,071	**15,85**
SES				**68,34**
Dailysmoker	-0,054	-0,019	0,001	0,22
Drink	-0,021	0,107	-0,002	0,49
Sedentarism	0,152	0,170	0,026	**5,72**
Lifestyle				**6,43**
cc_aa2	-0,005	0,017	0,000	0,02
cc_aa3	0,003	0,041	0,000	0,03
cc_aa4	-0,010	0,016	0,000	0,04
cc_aa5	-0,026	-0,026	0,001	0,15
cc_aa6	0,004	-0,003	0,000	0,00
cc_aa7	-0,019	-0,013	0,000	0,05
cc_aa8	-0,033	-0,026	0,001	0,19
cc_aa9	-0,104	0,039	-0,004	0,90
cc_aa10	-0,018	-0,063	0,001	0,26
cc_aa11	0,011	-0,027	0,000	0,07
cc_aa12	-0,015	-0,016	0,000	0,06
cc_aa13	-0,035	0,148	-0,005	1,14
cc_aa14	-0,010	-0,012	0,000	0,03
cc_aa15	0,002	0,013	0,000	0,01
cc_aa16	0,007	0,017	0,000	0,03
cc_aa17	0,006	0,010	0,000	0,01
Region				**2,98**
CCI :-0,046				

Note: Statistically significant contributors in bold.

Table 2.10: Decomposition of the FGT-CI of depth and severity of obesity-Women.

Women, 1997

			Depth			Severity	
Variable	CI_k	η_k	CI_{Yk}	%	η_k	CI_{Yk}	%
Age group(reference : age1)							
Age2	0,160	0,207	0,033	**5,50**	0,348	0,056	**6,58**
Age3	0,034	0,272	0,009	**1,55**	0,417	0,014	**1,70**
Age4	-0,204	0,209	-0,043	**7,07**	0,261	-0,053	**6,31**
Age5	-0,382	0,237	-0,090	15,02	0,311	-0,119	**14,06**
Married	0,052	0,026	0,001	0,22	-0,210	-0,011	1,28
Widowed	-0,482	-0,002	0,001	0,14	-0,041	0,020	2,32
Divorced/separated	-0,224	-0,005	0,001	0,17	-0,034	0,008	0,90
Education (reference: Primary)							
Secondary	0,214	-0,229	-0,049	**8,11**	-0,278	-0,059	**7,02**
Tertiary	0,523	-0,185	-0,097	**16,03**	-0,167	-0,087	**10,30**
Employed	0,265	0,086	0,023	3,78	0,244	0,065	**7,65**
Inactivity	-0,111	0,109	-0,012	**2,00**	0,365	-0,040	4,78
Loginc	0,051	-1,931	-0,099	16,46	-1,588	-0,082	9,65
Dailysmoking	0,161	0,066	0,011	1,76	0,173	0,028	**3,29**
Drink	0,166	-0,203	-0,034	**5,57**	-0,259	-0,043	**5,07**
Sedentarism	0,103	0,125	0,013	**2,14**	0,168	0,017	**2,06**
cc_aa2	0,056	-0,024	-0,001	0,22	-0,039	-0,002	0,25
cc_aa3	0,327	-0,050	-0,016	2,71	-0,088	-0,029	3,42
cc_aa4	0,237	-0,043	-0,010	1,70	-0,071	-0,017	2,01
cc_aa5	-0,185	-0,021	0,004	0,66	-0,050	0,009	1,09
cc_aa6	-0,042	-0,011	0,000	0,07	-0,018	0,001	0,09
cc_aa7	-0,092	-0,086	0,008	1,30	-0,120	0,011	1,30
cc_aa8	-0,136	-0,020	0,003	0,44	-0,019	0,003	0,30
cc_aa9	0,059	-0,153	-0,009	1,48	-0,274	-0,016	1,90
cc_aa10	-0,084	-0,085	0,007	1,18	-0,124	0,010	1,23
cc_aa11	-0,146	-0,010	0,001	0,23	-0,023	0,003	0,40
cc_aa12	-0,179	-0,069	0,012	2,05	-0,122	0,022	2,59
cc_aa13	0,281	-0,019	-0,005	0,88	-0,025	-0,007	0,82
cc_aa14	-0,162	0,008	-0,001	0,23	-0,001	0,000	0,02
cc_aa16	0,326	-0,004	-0,001	0,23	-0,010	-0,003	0,39
cc_aa17	0,129	-0,050	-0,006	1,08	-0,081	-0,010	1,23
Residuals			0,060			0,005	
Overall CI			**-0,270**			**-0,300**	

CI_k , concentration index of factor k.

η_k , elasticity of the FGT measure Y with respect to factor k.

CI_{Yk} , contribution made by factor k to the overall FGT-CI

CI, concentration index.

Note: Statistically significant contributors in bold.

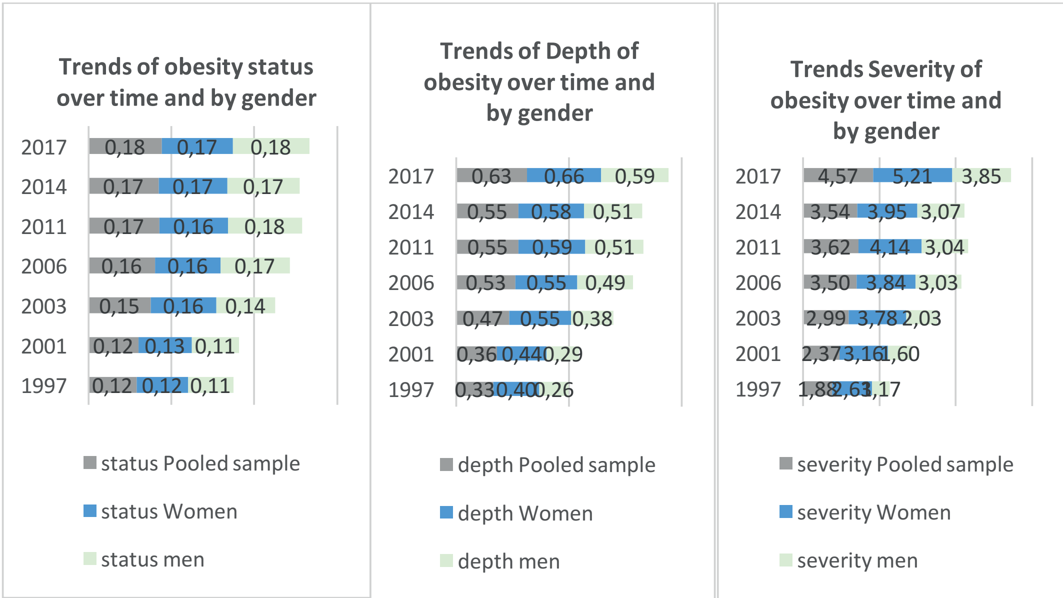
Table 2.11: Decomposition of the FGT-CI of depth and severity of obesity-Men.

Men, 1997

		Depth			Severity		
Variable	CI_k	η_k	CI_{Yk}	%	η_k	CI_{Yk}	%
Age group(reference : age1)							
Age2	0,076	0,053	0,004	1,23	-0,029	-0,002	0,42
Age3	0,111	0,104	0,012	**3,46**	0,066	0,007	1,37
Age4	-0,223	0,085	-0,019	**5,72**	0,118	-0,026	4,96
Age5	-0,356	0,064	-0,023	**6,87**	0,131	-0,047	8,82
Married	-0,032	0,311	-0,010	**2,96**	0,496	-0,016	**2,97**
Widowed	-0,371	0,015	-0,005	1,63	0,022	-0,008	**1,55**
Divorced/separated	-0,133	-0,002	0,000	0,08	-0,010	0,001	0,26
Education (reference: Primary)							
Secondary	0,201	0,027	0,006	1,66	0,140	0,028	5,32
Tertiary	0,530	-0,034	-0,018	5,40	-0,065	-0,034	6,46
Employed	0,176	-0,200	-0,035	**10,58**	-0,392	-0,069	**12,99**
Inactivity	-0,169	-0,263	0,045	13,50	-0,480	0,081	**15,33**
Loginc	0,048	-1,936	-0,092	**27,69**	-1,961	-0,093	17,59
Dailysmoking	-0,013	0,002	0,000	0,01	0,097	-0,001	0,25
Drink	0,037	-0,110	-0,004	1,21	-0,315	-0,012	**2,17**
Sedentarism	0,111	0,187	0,021	**6,27**	0,347	0,039	**7,29**
cc_aa2	0,164	0,016	0,003	**0,81**	0,040	0,007	**1,24**
cc_aa3	0,330	0,019	0,006	1,84	0,025	0,008	1,57
cc_aa4	0,208	-0,016	-0,003	1,01	-0,023	-0,005	0,92
cc_aa5	-0,148	-0,044	0,006	1,94	-0,073	0,011	2,04
cc_aa6	-0,017	-0,005	0,000	0,02	-0,015	0,000	0,05
cc_aa7	-0,047	-0,011	0,001	0,16	0,002	0,000	0,02
cc_aa8	-0,122	0,013	-0,002	0,48	0,024	-0,003	0,56
cc_aa9	0,076	-0,123	-0,009	2,82	-0,194	-0,015	2,79
cc_aa10	-0,137	0,008	-0,001	0,34	0,026	-0,004	0,67
cc_aa11	-0,195	0,014	-0,003	0,81	0,013	-0,003	0,49
cc_aa12	-0,066	-0,002	0,000	0,04	0,016	-0,001	0,20
cc_aa13	0,267	0,010	0,003	0,83	-0,005	-0,001	0,24
cc_aa14	-0,127	0,006	-0,001	0,21	0,006	-0,001	0,14
cc_aa15	0,227	-0,002	-0,001	0,16	-0,016	-0,004	0,70
cc_aa16	0,122	0,002	0,000	0,09	0,004	0,001	0,10
cc_aa17	0,385	-0,001	0,000	0,14	-0,007	-0,003	0,52
Residuals			-0,003			0,037	
Overall CI			**-0,133**			**-0,125**	

CI_k , concentration index of factor k.
η_k , elasticity of the FGT measure Y with respect to factor k.
CI_{Yk} , contribution made by factor k to the overall FGT-CI
CI, concentration index.
Note: Statistically significant contributors in bold.

Figure 2.2: Trends in obesity measures



3 Geographic determinants of individual obesity risk in Spain: A multilevel approach

This paper seeks to understand the determinants of individual body weight status and obesity risk in Spain by concurrently examining individual and regional characteristics. The data are drawn from the National Health Survey of Spain for the year 2011-2012 (INE-National Statistical Institute of Spain) and contain information for a representative sample of 12,671 adults across 50 provinces in Spain. A multilevel analysis is carried out to examine the determinants of individual weight status and obesity, controlling not only for the individual effects and those of the immediate environment but also for the broader setting to which individuals and their immediate environment belong. Our findings suggest that attributes from all three levels of analysis have an effect on individual weight status and obesity. Lack of green spaces and criminality taken as proxies of the social environment positively affect individual and women's BMI and obesity, respectively.

3.1 Introduction

Obesity is a highly complex condition with numerous dimensions that affects all groups of people, irrespective of their age or social status. The condition threatens to overwhelm both developed and developing countries and has been identified as a risk factor for major chronic diseases, such as high blood pressure, type II diabetes and many types of cancer (affecting, among others, the kidney, thyroid and pancreas), resulting in millions of deaths annually. Latest data suggest that the proportion of adults that are overweight (pre-obese) and obese has increased substantially in recent years (WHO, 2015). Despite the fact that obesity constitutes one of the most serious public health problems of the century, its main determinants have yet to be fully clarified.

Likewise, obesity is on the rise in Europe, and Spain is no exception here. Indeed, Spain is one of the EU countries with the highest prevalence of overweight and obese individuals (the highest in child obesity) and it is also one of the countries in which the condition has increased most (OECD, 2015). Over the last few years, this phenomenon has attracted the interest of many researchers from different fields (economists, doctors, nutritionists, geographers and policy makers) who have identified several interrelated determinants underpinning this epidemic. In addition to personal and socioeconomic factors, geographic patterns of high obesity rates have been observed across Spain (Gutierrez-Fisac et al., 1999). In such instances of significant variations being observed within a country, a fall in the prevalence of the condition might be expected through the drawing up of purpose-designed, public health intervention policies. However, an important preliminary step is understanding and accounting for the determinants of this variability and the factors explaining these variations, in order to be able to recommend specific prevention policies.

In this paper, our objective is to disentangle the influences on individual weight status and obesity risk, by employing realistic models of disease causation. Based on the fact that these determinants might be aspects of the sociocultural environment, the physical environment or an after-effect of an economic policy implemented in a region, we employ a hierarchical model with three levels of analysis that correspond to the individual, the surrounding area and the Autonomous Community (AC) or region. It is our belief that previous studies have failed to take into account regional and individual characteristics in tandem when dealing with obesity. In this study, therefore, we control for personal characteristics as well as higher level geographic variations that may cause the body mass index (BMI) to rise above normal levels. Our contribution is to illustrate how individual weight status and obesity risk are explained by individual and regional characteristics - both in the immediate environment and in the broader setting - exploiting the hierarchical structure of the data, in a multilevel (ML) regression model. In such circumstances, ML models are applicable and usually preferable to other models. An ML model assumes that individuals (lower level) belonging to a particular region (higher level) are not independent of each other because they share the similar characteristics of that region; thus, the model considers an intra-regional correlation. Our results suggest that attributes from all three levels have an effect on individual weight status and obesity.

The rest of the paper is structured as follows. In section 2 we review the previous literature and in section 3 we justify why Spain makes a particularly interesting case study. We detail the theoretical model in section 4 and the data and variables used are discussed in section 5. Finally, we present the results of the econometric analysis in section 6 and section 7 concludes.

3.2 Literature

The complexity of the condition has attracted numerous researchers who have analysed obesity from a range of perspectives. This paper, however, focuses on two major strands in this literature given their relevance to the analysis conducted here. On the one hand, one line of research has established a strong correlation between socioeconomic status (SES) and health outcomes. The graded association between various indicators of SES and health holds across all ages and for all countries in which it has been studied, as a person's health behaviour (e.g. smoking, drinking and level of physical activity) is believed to be significantly influenced by their socioeconomic status, creating a mechanism that links SES to health (Antonovsky, 1967; Anderson, 1995; Link et al., 1998; Subramanian and Kawachi, 2004; Cutler et al., 2012). According to Link and Phelan (1995), people of higher SES have access to a wide range of resources to influence their health and are, therefore, at an advantage when their health is threatened. Data from the US and Canada show that levels of health tend to be higher among the richer, better educated and more privileged and then to gradually deteriorate down the rungs of the social ladder (Humphries and van Doorslaer, 2000; Kosteniuk and Dickinson, 2003). This relationship is referred to as the SES health gradient.

A growing body of influential literature here documents the specific relationship between personal weight status and SES. Sobal and Stunkard (1989) and Sobal (1991) find clear-cut evidence of an association between socio-economic position and obesity. Moreover, Drewnowski and Specter (2004) and Drewnowski (2003) demonstrate that population groups with high poverty rates and low education exhibit the highest obesity rates and that wealth and poverty have profound effects on diet structure, nutrition and health. The significant long-run effect of education on obesity is patently obvious in Kim (2016), who investigates how education is associated with BMI in later stages of life using data from the US. Other researchers find pro-rich inequality in obesity in adults, while the main drivers of this finding are educational attainment and income (Eberth and Gerdtham, 2008; Devaux et al., 2011). According to Monteiro et al. (2000, 2004) and Merino Mantosa and Urbanos-Garrido (2016), obesity is increasing faster in low-SES subpopulations. In a recent study, Costa-Font et al. (2014) present evidence of income-related inequalities in obesity in England and Spain for the years 1987-2006, although patterns differ by gender and age-group.

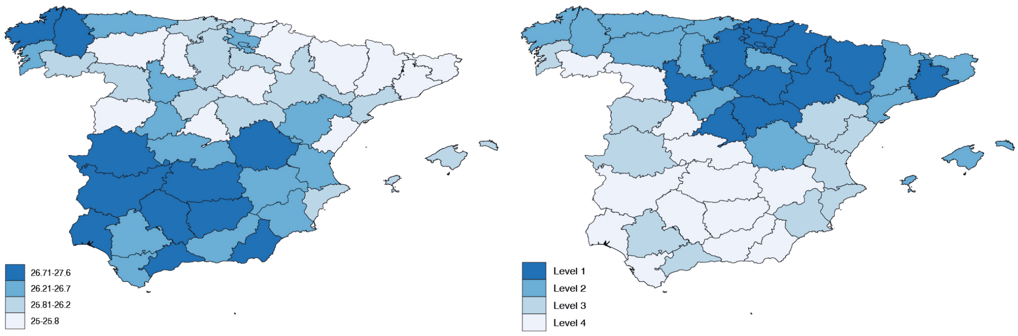
The second strand in the literature with a bearing on this study identifies environments that tend to encourage obesity-related behaviours (e.g. unhealthy eating and low rates of physical exercise) more than others. These are what Egger and Swinburn (1997) first referred to as "obesogenic environments". According to these authors, environmental influences (distance to grocery stores, parks, neighbourhood safety, green areas) represent the public health arm of the obesity problem. If the surrounding environment is obesogenic, obesity will become more prevalent and programs aimed at influencing individual behaviour can be expected to have only limited effects. They conclude by recommending a broader public health approach to the obesity epidemic. A large number of other studies find a strong correlation between area-level variables and individual obesity (Mark Austin and Spine, 2002; Cubbin et al., 2006). Obesity patterns are associated with individual but also contextual socioeconomic and environmental factors, and the residential context might influence the SES-health gradient in complex ways. According to Costa-Font and Gil (2008), sociocultural contexts of obesity are recognized as key factors that account for the development of an individual's weight.

The geography of obesity is complex, given the interrelation of different factors at both the individual and contextual level. If we understand that environmental features and natural amenities are geographically located, it can be deduced that health problems, such as obesity, are geographically clustered according to socioeconomic and regional factors and that regional spillovers also need to be taken into account in such analyses. Bailey and Gatrell (1995) argue that the prevalence of overweight and obese individuals in one region is likely to correlate with the prevalence in nearby regions, indicating the presence of geographic clusters. Costa-Font and Pons-Novell (2007) point out the need to control for potential geographic dependency, which might be especially important in heterogeneous countries like Spain. Several other papers seek to investigate the relationship between socioeconomic status and obesity risk and report that in most countries obesity contributes to health disparities, as the condition is not evenly distributed across geographic areas and social groups (Black and Macinko, 2009; McLaren, 2007; Wang and Beydoun, 2007). When analysing health outcomes, such as overweight and obesity, taking into account regional characteristics and geographic spillovers is vital, especially when socioeconomic development and policies in the areas of education and health care can affect these outcomes.

3.3 Spanish Case

Spain provides an especially interesting case study for such an analysis for two main reasons. First, adult obesity rates in Spain are relatively high compared to those in the rest of Europe and the OECD (OECD, 2014), while among EU countries the impact of obesity on avoidable mortality is particularly high in Spain (Costa-Font and Gil, 2008). Additionally, obesity rates have increased rapidly compared to those in countries of similar characteristics (Costa-Font et al., 2010b). Second, because of the country's decentralized health care system with health competencies having been devolved to the ACs, disparities are conspicuous even within Spain, making the country even more interesting to study. More specifically, the obesity epidemic seems to be affected by the country's diverse population and severe income inequality, making Spain one of the most suitable institutional settings in which to examine regional inequalities (Costa-Font and Gil, 2008). The existence of a pattern of geographic heterogeneity - captured by the random effects we include in the analysis - is evident from the following Figure 3.1.

Figure 3.1: *BMI vs Income Distribution*



The data confirm that the average self-reported BMI at the provincial level is higher in the southern part of Spain than in the northern part of the country, while income distribution presents the opposite pattern.

3.4 The model

3.4.1 The rationale of using multilevel modelling

The adoption of a multilevel approach (also known as hierarchical, mixed and random-effects, covariance components or random-coefficient regression) is especially suited to certain contexts. First, it is highly appropriate when the individual health outcome is anticipated to be clustered, and the source of clustering is a geographic area, such as a census block-group or tract. It is especially relevant for use by governments as it indicates the level at which policy actions should be implemented and the relative importance of each level in predicting outcomes. Second, it is also appropriate for use when the exposure variable is measured at multiple levels and the interest is in evaluating the relative importance of a variable at different levels (Leyland and Groenewegen, 2003).

These specific conditions identified above apply to our case study and therefore make the use of ML analysis highly appropriate. Our data document substantial regional variations in obesity rates in Spain. These geographic variations have a number of causes, some of which may reflect characteristics of the regions or characteristics of the individuals that live in these areas. Thus, as we seek to observe not only the locational and environmental but also the socioeconomic attributes of the region under investigation, adopting an ML statistical approach becomes essential. This framework allows us to incorporate into our model both regional and environmental factors, while studying body weight variations. In short the main advantage of our approach is being able to analyse the determinants of BMI variation and obesity prevalence by focusing on three levels of analysis, corresponding to individual influences, environmental attributes and regional effects.

Ignoring the ML structure of variations does not simply risk disregarding the importance of regional effects, but it also has implications for statistical validity. ML models have three basic advantages over OLS models. First, the use of conventional regression for clustered data results in the underestimation of standard errors, because this model does not consider the similarity of responses among observations within the same cluster. ML models resolve this problem by including random components of cluster effects in the statistical model and, therefore, they divide the total variance in the dependent variable into between-cluster and within-cluster parts. The variability of random effects across clusters and the importance of clusters can also be evaluated (Teachman and Crowder, 2002; Ross and ling Wu, 1995). Second, both individual-level and cluster-level covariates can be included in ML models and so, the relationship between observation-level and cluster-level covariates can be examined in order to determine whether cluster-level characteristics moderate individual-level relationships. Third, aggregation bias might occur in the OLS model, but it is absent from ML models. This means that the results from aggregated variables at the cluster level may differ from those at the individual level (Kreft and De Leeuw, 1998; Ross and ling Wu, 1995). That is, ML models separate the estimated effects in the covariates into different levels, which can be interpreted as individual-level effects (within a cluster) and cluster-level effects (across clusters), respectively.

3.4.2 Model Specification

The general hypothesis underpinning our analysis is that people interact within their local contexts and, therefore, they are influenced by the social groups to which they belong. Thus, we hypothesize that an individual's weight status is influenced by both personal and contextual characteristics. Exploiting the hierarchical structure of our data, we employ a three-level model as we seek to explain personal weight status and obesity risk by taking into account both individual (level 1) and regional (level 2 and 3) characteristics. Defining the relevant group and the relevant group-level variables is a crucial component of ML analysis. The groups should not be arbitrary or convenient groupings of individuals, but rather groups that are hypothesized as being meaningful in any explanation of the outcome. Hence, each level of analysis chosen is essential as it enables us to observe the determinants of obesity from three different standpoints. [1]

At the first level of analysis, our basic determinants of individual weight status and obesity are personal characteristics. The census tract constitutes the second level of our analysis and so, we include variables corresponding to the individual's immediate environment. This level of analysis is important here as it allows us to observe the relationship between an individual's weight status and their environment. Finally, the third level of analysis is comprised by the ACs. In a decentralized country like Spain, the ACs become even more spatially meaningful with respect to specific target policies and health outcomes. As such, the use of variables measured at the regional level can be an intuitive task for policy intervention, while at the same time it allows for policy evaluation (Costa-Font and Pons-Novell, 2007).

We test two econometric models: a linear ML model for analysing individual BMI and a non-linear logit ML model for analysing obesity. Specifically we estimate the following equation:

$$Y_{i,j,k} = \alpha + \beta X_{ijk} + \theta_j + \mu_k + \epsilon_{ijk} \tag{3.1}$$

where $Y_{i,j,k}$ is the outcome (BMI and obesity) variable of individual i, in census tract j and Autonomous Community k. The parameter α is the overall intercept coefficient; β stands for the effect of the fixed covariates; while θ and μ denote the level 2 and level 3 random effects, respectively, and ϵ is the error term. The vector X of regressors includes level 1 variables and regional characteristics (level 2 and level 3 variables). Notwithstanding we will run the OLS model for comparative purposes.

[1] According to Flowerdew et al. (2008) given the uncertainty concerning neighborhood boundaries, it is more reasonable to estimate neighborhood effects using multiple geographic units of analysis.

3.5 Data and Variables

For our empirical analysis we make use of individual health data and additional socioeconomic characteristics for the year 2011-2012, obtained from the Spanish National Health Survey (SNHS), which is conducted jointly by the Ministry of Health, Social Services and Equality (MSSSI) and the National Statistics Institute (INE). A stratified tri-stage sample type is used, where the first-stage units are the census tracts and the second-stage units are the main family dwellings. All households whose regular residence is within said dwellings are studied. Within each household, one adult person is selected to complete the individual questionnaire. The sample is distributed among the ACs, assigning one portion uniformly and another in proportion to the size of the ACs, so that, besides being representative at the national level, it is also representative at the AC level. Information on socio-demographic characteristics, health status, health care and health determining factors are available in this dataset. The original dataset contains 26,502 interviews with information about 21,007 adults (15+) and 5,495 children. Children and individuals with missing information about their weight and height were excluded from the sample and we were left with 18,649 observations. We then had to discard all those for which we had no information about their region of residence (5,978 interviewees), and so ended up with a joint dataset of 12,671 observations. The regional data with information at the NUTS2 and NUTS3 levels are taken from Eurostat.

3.5.1 Dependent Variables

Our two outcome variables are BMI and obesity. We measure BMI using self-reported weight and height data. [2] Specifically, BMI is defined as a person's weight (in kilograms) divided by the square of his or her height (in metres). A person with a BMI of 30 or more is generally considered obese, while a person with a BMI equal to or more than 25 is considered overweight (WHO, 1995). A normal-weight person exhibits BMI scores ranging between 18.5 and 24.9, while an underweight individual has BMI scores below 18.5.

[2] Given that self-reported anthropometric data suffer from reporting bias (i.e weight tends to be under-declared and height over-reported), we re-run the same ML models after correcting BMI using the standard procedure (Cawley, 2000; Lakdawalla and Philipson, 2009; Chou et al., 2004; Cawley and Burkhauser, 2008), although this is not free from criticism (e.g., Courtemanche et al. (2015)). In this case, the correcting formulas for Spain provided by Gil and Mora (2011) were used. Interestingly, we find that the resulting ML estimates are roughly the same.

3.5.2 Independent Variables

The individual-level control variables include demographic and socioeconomic characteristics, as well as health-related behaviours. Among the demographic factors, we include age, age squared, gender, ethnicity and marital status. Based on the literature, we use education and equivalent net income as proxies for the individual's SES status.[3] Concerning health-related behaviours, we control for tobacco consumption (daily smokers), sedentarism and physical exercise in leisure time (see Table 3.1).

The level of urbanization is measured using the urban-rural dichotomy.[4] Criminality is taken as a proxy for neighbourhood safety and is traditionally used in studies examining environmental factors and their influence on individual health outcomes (Zhao et al., 2014). Living in a neighbourhood which is perceived as unsafe can contribute to obesity and generally higher BMI levels in a number of ways, including lower rates of walking or other types of outdoor physical activity and higher rates of stress-related eating. In our case, the criminality rate is interacted with gender, [5] and so crime is a factor variable measuring whether the respondent believes the violence/criminality/vandalism in their neighbourhood is high or not and to what extent. [6] A variable capturing the existence of green spaces in the neighbourhood (census tract) is also included in the analysis.[7] The last regional variable that we include in the model is poverty risk, capturing the impact of regional socioeconomic status on individual obesity risk and weight status.

[3] The OECD equivalence scale is used.

[4] This depends on the number of residents in the specific census tract of residence.

[5] Sampson et al. (2002) use criminality as a proxy of neighbourhood attributes within ML analysis

[6] This variable was defined at an individual level in the household survey but we measure it at the census tract level.

[7] Some studies demonstrate that green areas are associated with lower levels of BMI (Dadvand et al., 2014) and quite often the existence of green spaces is used as a predictor of overweight and obesity.

Table 3.1: Description of dependent and independent variables

Variables	*Description*
Dependent Variables	
BMI	Self-reported body mass index
Obesity	1 if obese(BMI$\geq$ 30), 0 otherwise
Independent Variables	
Personal Characteristics	
female	1 when female, 0 otherwise
age	age
age^2	square of the age
single	1 if single, 0 otherwise
married	1 if married, 0 otherwise
widowed	1 if widowed, 0 otherwise
divorced	1 if divorced, 0 otherwise
native	1 if Spanish, 0 otherwise
low education	1 if primary/lower than primary education, 0 otherwise
secondary education	1 if secondary education, 0 otherwise
university education	1 if university education, 0 otherwise
income	equivalence net income in euros
daily-smoker	1 if daily smoker, 0 otherwise
sedentarism	1 if seated most of the time during work, 0 otherwise
physical activity in leisure time	1 if intense physical activity, 0 otherwise
Regional Characteristics	
urban	1 if census tract is urban, 0 otherwise
green0	1 if lack of green areas, 0 otherwise
green1	1 if moderate lack of green areas, 0 otherwise
green2	1 if no lack of green areas, 0 otherwise
crime0	low/no level of criminality in the area
crime1	moderate level of criminality in the area
crime2	high level of criminality in the area
poverty risk	poverty risk per Autonomous Community

Source: Spanish National Health Survey 2011-2012. INE: Statistical Institute of Spain, Eurostat

3.6 Results

3.6.1 Sample Characteristics

As Table 3.2 shows, average BMI levels are higher for men than for women in our sample (26.36 vs. 25.97). The adult obesity rate is 18%, but higher for women (19.5% vs. 18.1% men), whereas 39% of the sample population is overweight (41% for men vs.36% for women). The data document important gender differences in the BMI and obesity patterns. Specifically, we report that men smoke on a daily basis more than women (28.8% vs 23.4%), but that men exercise more. In our sample, men tend to be younger than women (51.14 and 54.09). There are also gender differences with regard to education. The data indicate that men are more highly educated than women (31.56% vs. 28.78%). Although not shown, our data indicate that more highly educated individuals tend to have higher incomes and to display lower BMI levels. We also notice that people belonging to upper social groups tend to be slimmer. According to the literature, urbanization level and weight status are strongly correlated, especially among women (Caliendo and Lee, 2013). In agreement with this, our results indicate that married women who live in rural/semi-urban regions exhibit the highest levels of BMI. In addition, less educated women exhibit higher levels of body weight.

Table 3.2: Table of Sample Characteristics

		Men		*Women*
	Mean	Frequency%	Mean	Frequency%
self-reported BMI	26.36		25.97	
obese		18.13		19.52
age	51.14		54.09	
single		31.34		27.05
married		55.18		47.23
widowed		6.68		18.51
separated/divorced		6.8		7.15
native		93.65		94.27
low		24.22		30.67
moderate		44.22		40.56
high		31.56		28.78
smoke		28.77		23.43
physical activity during work or free-time		49.4		32.4

Means calculated using sampling weights. Source: Spanish National Health Survey 2011-2012. INE

3.6.2 Econometric Analysis

The estimates of the linear (logit) ML models for BMI (obesity) are presented in Tables 3.3(3.4). [8] Both tables also include OLS estimates for comparative purposes. The likelihood-ratio test suggests the use of the random coefficient model, but only for the ML linear model. However, since the coefficients and standard errors are almost identical in both the random intercept and random coefficients model, we present the estimates based on the random intercept model.

As expected, our results show that BMI and obesity are associated with personal socioeconomic characteristics, health behaviours and regional characteristics using the two models. More specifically, net income is statistically significant and has a negative association both with BMI and obesity. That is, individuals with higher net income tend to exhibit lower levels of BMI and lower probability of being obese. The same applies for high-educated individuals who appear to have lower expected BMI and obesity levels compared to those of the low-educated. Nationality seems to be a risk factor as well, as natives tend to have lower levels of BMI and obesity prevalence. As for smoking habits, we observe that daily smokers have lower levels of BMI and obesity than non-smokers, which is an expected finding according to the literature. [9] Sedentarism and no physical exertion in leisure time are positively associated with BMI and obesity.

Interestingly, all the regional variables that we include in the model are statistically significant, which indicates the need to include attributes of the contextual environment in the study. Specifically, people that live in census tracts (approximate neighbourhood area) which lack green spaces seem to have higher BMI scores and a higher probability of being obese, compared to those who have access to green spaces in their areas and who engage in physical activity in their leisure time. Finally, poverty risk, as expected, is positively associated with BMI and obesity.

The use of the ML model allows us to examine whether social context influences the effect of a level-1 variable. We find that criminality/violence (proxy of social environment) is positive and statistically significant only for women. Thus, we report a positive association between high levels of neighbourhood insecurity and weight status for women that live in these neighbourhoods. In particular, we find that living in an unsafe area raises female BMI by 0.15 points and female obesity by almost

[8] In Table 4, we report odds ratios as opposed to coefficients.

[9] According to Dare et al. (2015), for example, current smokers are less likely to be obese than non-smokers.

10% (although significant only at the 10% level).

We should stress that most of the ML estimates are lower than those of OLS, which suggests that the individual level effects reported by the literature (where intercepts are restricted to being the same across regions) are due to BMI and obesity level differences between regions. By allowing intercepts to vary randomly, we have obtained a more accurate estimate of the effect of the covariates on BMI and obesity. We also find that ML and OLS estimates do not deviate greatly. However, the Hausman test suggests the superiority of the ML model.

In our case, the intra-class correlation coefficient (ICC) is 10.45% which means that this share of the total BMI variance is determined at the regional level. [10] Although the ICC is close to the lower bound recommended for running an ML (>10%), here it is relevant since the LR test rejects the traditional model in favour of a ML multilevel model with random effects.[11]

[10] The extent of variance that exists between groups, as opposed to within groups, can be described by an intra-class correlation. An ICC can be determined from an intercept-only model (an ML model with no covariates).

[11] Notice that the variance at the individual level is : σ_e^2, the variance at the census level is : σ_u^2 and the variance at the AC level is : σ_v^2

Table 3.3: Estimates of the ML Model

Dependent Variable	BMI	
	Random Intercept	OLS
Fixed Part		
Personal characteristics		
intercept	23.18(.56)**	21.58(.37)**
female		-.37(.07)**
age	.11 (.012)**	.12(.012)**
age^2	-.09(.01)**	-.10(.01)**
single (reference category)		
married	1.54(.10)**	1.58(.09)**
widowed	1.56(0.22)**	1.64(.14)**
separated/divorced	0.71(0.21)	.83(.016)*
native	-.06 (0.033)*	-.08(.03)*
primary education (reference category)		
secondary	-.62 (.10)**	-.63(.10)**
university	-1.23 (.14)**	-1.20(.13)**
income	-.02(.008)*	-.03(.01)**
daily-smoker	-.84(.10)**	-.84(.09)**
sedentarism	.69(.08)**	.71(.07)**
physical activity in leisure time	-.86(.14)**	-.89(.14)**
Regional characteristics		
urban	-.21(.13)*	-.42(.10)*
green0	.28(.14)**	.29(.14)**
green1	.16(.15)	.17(.15)*
green2 (reference category)		
poverty risk	.04 (.006)**	.03(.005)**
Interaction effect		
women*crime	.11(.03)**	.12(.06)*
Random Part		
σ_e^2	4.01	
σ_u^2	.93	
σ_v^2	.89	
ICC	10.45%	

*Note: Standard errors are in parentheses, *: p<.01, **: p<.005*

Table 3.4: Estimates of the ML Model

Dependent Variable	Obese	
	Random Intercept Model	Logistic
Fixed Part		
Personal characteristics		
intercept	.080(.07)**	.03(.02)**
female		.98(.05)
age	1.08(.09)**	1.08(.08)**
age^2	.99(.000)**	.99(.000)**
single (reference category)		
married	1.53(.10)**	1.53(.09)**
widowed	1.78(.14)**	1.75(.14)**
separated/divorced	1.03(.12)	1.05(.11)*
native	.095(.03)*	.095(.02)*
primary education (reference category)		
secondary	.65(.04)**	.65(.04)**
university	.39(.06)**	.37(.05)**
income	.95(.11)**	.89(.10)**
daily-smoker	1.07(.13)**	1.07(.13)**
sedentarism	1.49(.08)**	1.51(.08)**
physical activity in leisure time	.54(.06)**	.52(.06)**
Regional characteristics		
urban	.85(.08)	.97(.05)*
green0	1.17(.08)**	1.21(.08)**
green1	1.06(.06)	1.06(.06)
green2 (reference category)		
poverty risk	1.01(.013)*	1.01(.003)**
Interaction effect		
women*crime	1.09(.002)*	1.10(.002)**
Random Part		
ICC	9.01%	

*Note: Standard errors are in parentheses, *:p<.01, **: p<.005*

3.7 Conclusion

This paper seeks to explain the basic determinants of individual weight status and obesity risk by accounting for the interplay between individual and regional attributes in a multilevel framework. In line with the literature (Roux, 2000), we conclude that both group and individual effects play a key role in understanding BMI and obesity.

Our ML estimates confirm the expected individual-level and regional effects on BMI and obesity. In addition, we provide evidence that our proxies of the social environment (criminality and green spaces) have a positive and statistically significant effect on female BMI and the prevalence of obesity. The main finding of the paper stresses the fact that some health determinants can only be observed while investigating interactions between variables at different levels of analysis. ML modelling provides a more accurate and comprehensive description of relationships in clustered data than do conventional OLS models by correcting underestimated standard errors, by estimating components of variance at several levels and by estimating cluster-specific intercepts and slopes.

The findings of the paper are important for public health authorities, since we report that environmental and regional characteristics influence individual BMI and obesity. This means that local governments and local communities can play an important role in implementing specific policies, such as promoting environments that encourage and support healthy lifestyles.

We should point out that here we identify associational relationships in our estimates, while a handful of other studies have sought to identify through exogenous variations causal effects of such factors as education, income and the characteristics of the built environment on individual weight status (Cawley et al., 2010; Brion et al., 2011; Brunello et al., 2013; Martin et al., 2014). Nevertheless, we have employed a highly complex statistical tool (ML modelling) that allows us to separate out the effect of each variable, and so account for the effects from higher levels (environmental influences). Yet, having said that, even in the best circumstances, correlation is not the same as causation.

One potential limitation of our study is the possibility that people may self-select into neighbourhoods. That is, many of the idiosyncratic characteristics that affect obesity may also affect neighbourhood choices. For instance, someone with a particular dislike for walking is both more likely to be obese and to prefer living where they can easily get around by car (Eid et al., 2008). A few studies that have sought to observe this issue of reverse causality reached the conclusion that sorting rather than causation is the mechanism which drives observed differences in individual characteristics across places, or that correlation between sprawl and obesity does not imply causation (Plantinga and Bernell, 2015). Arguably, controlling for unobserved individual characteristics would be of great interest for this study; however, data unavailability is a hurdle which unfortunately prevents us from controlling for this, at least in the case of Spain. All in all, this analysis has added to the paucity of literature examining the relationship between regional characteristics and obesity, where much work remains to be done.

4 What drives regional differences in Body Mass Index? Evidence from Spain.

4.1 Introduction

The rapid increase of overweight and obesity around the globe has raised concerns both from a health perspective and from an economic point of view, as it represents a high risk factor for several chronic diseases like CVD, stroke, hypertension, diabetes, dyslipidaemia or some cancers (Malnick and Knobler, 2006). Overweight generates negative effects on labour market performance e.g. (Cawley, 2004; Morris, 2006; Lindeboom et al., 2010; Kinge, 2016) and, directly and indirectly, increases health care expenditure e.g. (Finkelstein et al., 2009; Tremmel et al., 2017).

Spain is one of the countries experiencing high trends in the prevalence of overweight and obesity compared to the OECD average (OECD, 2014). Specifically, 1 out of 6 adults is obese and more than 1 out of 2 is overweight (including obese) in Spain. Notwithstanding, strong regional discrepancies in excess body weight exist within the country, i.e., the residents of some regions exhibiting much higher average BMI rates than others (Gutierrez-Fisac et al., 1999; Valdes, 2014; Raftopoulou, 2017). Geographical disparities in health outcomes have been observed in other countries as well. For example, Ellis and Fry (2010) consider several health indicators, including life expectancy, childhood obesity, cancer deaths, smoking and alcohol consumption to document the existence of a divide between northern and southern regions of the UK, in favour of the latter. This result is also confirmed by Hacking et al. (2011), showing a northern excess in all-cause mortality that remained substantial and persistent over the four decades from 1965 to 2008 in England, affecting relatively more males than females.

Investigating the existence and magnitude of regional differentials in BMI and analysing the underlying determinants of such health disparities could be especially relevant for public health policy-makers, in their intent to meet the WHO target of halting the rise of obesity to its 2010 level by 2025 (WHO, 2017). In addition, in contexts where the NHS is decentralized and health competences are primarily the responsibility of the country s regions, as in the case of Spain, local decision-makers need to have evidence on health indicators at the regional level. Therefore, the ultimate goal of this work is to produce evidence regarding the drivers of regional disparities in BMI for the Spanish case.

In this paper we decompose regional differentials in BMI between northern and southern Spanish regions [1]. First, by means of OLS regression, we analyse the relationship between BMI and several potential conditioning factors (basically sociodemographic attributes, socioeconomic status and lifestyle characteristics) and examine whether their conditional correlation with BMI is different between the two groups of regions. Second, we decompose the observed average gap into the part attributed to differences in observable determinants of BMI (i.e. the endowments) and the part that is left unexplained and is due to differences in the return to observable characteristics, using the classical Oaxaca-Blinder (OB) decomposition. Moreover, as long as important differences in BMI occur away from the average, we proceed with a distributional analysis by applying the Recentered Influence Function (RIF) regression and the corresponding decomposition (Firpo et al., 2009; Fortin et al., 2011). The RIF regression enables obtaining evidence along the unconditional distribution of BMI, which is especially important for the design of health and food policies. Indeed, policy-makers are interested in targeting policies to individuals who are (unconditionally) either underweight or obese, rather than those who appear in the two cues of the conditional distribution of BMI (i.e. whether they are obese or underweight given their characteristics). Therefore, our main contribution lies in the fact that we decompose regional differences in BMI along its unconditional distribution into the contribution of the endowment of observable characteristics and the return to those characteristics. This way, we are able to observe what happens at every part of the distribution and subsequently draw conclusions for the more interesting tails, the upper one (obesity, severe obesity) [2] and the lower (underweight) where relationships might vary. The analysis is carried out separately by gender, as the underlying mechanisms that affect BMI and health outcomes in general appear to be different for women and men.

[1]The results obtained under different grouping of regions will be analysed in the robustness checks section.

[2]The WHO defines obesity as a BMI $\geqslant 30kg/\mathrm{m}^2$, while severe obesity corresponds to BMI $\geqslant 40kg/\mathrm{m}^2$.

Our findings indicate that the South to North gap in BMI is mostly driven by women, whereas it is lower and not statistically significant for men (0.55 points, z-stat 3.1 for females relative to 0.11 points, z-stat 0.72 for males). Around 73% of the cross-regional gap in BMI among women is accounted by differences in observable characteristics. More specifically, women residing in the South have lower education and income levels. The distributional analysis reveals that the South to North gap in BMI for Spanish women tends to increase over its unconditional distribution, with observable factors (especially schooling) making a growing contribution in explaining the differential across the quantiles of BMI.

4.2 Related Literature

Two lines of research can be distinguished within the health economics literature and specifically the economics of obesity, where decomposition methods have been extensively employed. The first is linked to the well-known literature on SES-related health inequalities [3] and refers to a set of studies aimed at quantifying and decomposing the extent of inequalities in obesity risk via the calculation of concentration indexes. This research, mostly focused on developed countries, tends to show that obesity is mainly concentrated among the poor, and inequality varies over time, with education, demographics, income and life-style being its main contributors (e.g. Zhang and Wang (2004); Costa-Font and Gil (2008); Nikolaou and Nikolaou (2008); Ljungvall and Gerdtham (2010); Hajizadeh et al. (2014); Davillas and Benzeval (2016)). The second line of research includes those studies concerned with decomposing average BMI differentials by applying the Oaxaca-Blinder method (Dutton and McLaren, 2011; Sen, 2014) or examining the entire BMI distribution using conditional quantile regression (Costa-Font et al., 2009) [4].

Our paper is related to the latter group of studies that analyse BMI differentials, adopting a geographical perspective. Although our study is not the first in decomposing BMI differentials [5] (Costa-Font et al., 2009, 2010a; Dutton and McLaren, 2011; Sen, 2014), to the best of our knowledge we are the first in performing a detailed decomposition based on the RIF method and providing evidence for a European country. Indeed, Dutton and McLaren (2016) used a similar technique utiliz-

[3] See for instance Kakwani et al. (1997); Wagstaff et al. (2003); van Doorslaer and Koolman (2004)

[4] This tool has also been applied to other health issues such as differences in objective health indices (Heger, 2016) or in low birth weight (Lhila and Long, 2012)

[5] Decomposition tools have been also applied to other health issues such as differences in objective health indices (Heger, 2016) or in low birth weight (Lhila and Long, 2012)

ing Canadian data to examine the importance of individual - level characteristics for explaining geographic variation in BMI distributions. They also performed quantile regression and the corresponding decomposition, however focusing only on the aggregated effects. We move a step further by presenting the detailed decomposition of the relevance of specific key factors for the design of interventions targeting overweight individuals (such as age, education or lifestyle and food habits) in accounting for the regional gap in BMI.

From the methodological point of view, our study is mostly based on the contributions of two seminal papers (Oaxaca, 1973; Blinder, 1973), which present a method to decompose inter-group differences in the mean levels of an outcome into explained (i.e. difference in the endowment of observable characteristics) and unexplained (i.e. difference in the returns to those characteristics) factors using the OLS regression estimates. This method has been widely applied within the field of labour economics when decomposing average wage differentials by gender or ethnicity (Reimers, 1983; O'Neill and O'Neill, 2006). However, an important limitation of this approach is the focus on average gaps, thus neglecting important differences at other points of the outcome's distribution [6]. Therefore, subsequent developments extended the decomposition methods to other moments than the mean, or even to the whole distribution of the outcome (Freeman, 1980, 1984; Juhn et al., 1993; DiNardo et al., 1996; Machado and Mata, 2005)[7].

In this paper, we apply a method that was proposed by Firpo et al. (2009), where the (Recentered) Influence Function (RIF) for the distribution statistic of interest is used – instead of the usual outcome variable – as the left-hand side variable in a regression. The basic advantages of this analysis are twofold. First, it is not affected by path dependency, and second, it enables a detailed decomposition. That is, applying the OB decomposition to the RIF allows disentangling the observed gap along the unconditional distribution of BMI into the contribution of composition and returns effects of single covariates (or group of covariates) included in the model. It seems worth noting that the use of the RIF-Regression decomposition is especially relevant in our framework, since providing evidence on the unconditional distribution of BMI makes the analysis much more informative for policy-makers who are interested in designing policies addressed to those who are either over- or

[6] Another drawback of the Oaxaca-Blinder (OB) decomposition is that it is path-dependent, which means that the decomposition relies on the ordering of the explanatory variables.

[7] This set of approaches does have some drawbacks though. For example, the DFL (DiNardo et al., 1996) method does not allow detailed decomposition, while the MM (Machado and Mata, 2005) approach that is based on the decomposition of differences along the conditional distribution suffers from the problem of path-dependence on top of being computationally demanding.

underweight (not conditionally to over- or underweight)[8].

4.3 Data and Descriptive statistics

This paper draws on data from the 2014 wave of the Spanish version of the European Health Interview Survey (EHIS), which covers the population aged 15 or more and contains several sociodemographic and health-related variables. Moreover, the Spanish data of the EHIS survey are representative at the regional level (NUTS2), which enables examining regional disparities in BMI and their determinants. The original sample contains 22,842 observations. We keep only native Spaniards aged 18-65 at the time of the survey [9] with valid information on the relevant variables.[10] We also discard observations from the Balearic and Canary Islands, as well as the autonomous cities of Ceuta and Melilla that are located on the northern coast of Africa. [11]

As standard, body mass index or BMI is calculated as weight in kilograms divided by the square of height in meters (kg/m^2). These anthropometric measurements are based on self-reported information. Notwithstanding, we will assess the extent to which our benchmark findings are affected by the (potential) bias in BMI due to the habitual misreporting of weight and height in self-reported survey data (e.g. Bostrom and Diderichsen (1997); Kuczmarski et al. (2001); Gil and Mora (2011). Better measurements of body fatness and associated health risks seem to exist, such as the waist circumference or the waist-to-hip ratio (Janssen et al., 2004), however this information is not available in our data set.[12]

Mostly following the existing literature on BMI, we divide the conditioning factors into three main groups, namely 1) sociodemographic variables, 2) socioeconomic

---

[8] i.e., with very high or very low residuals, given the observed characteristics.

[9] The results are unaffected by the inclusion of migrants in the sample and controls for being born abroad, having a foreign nationality and years since migration (quadratic). These results are available upon request. Old-age individuals are disregarded to reduce the bias arising through larger mortality among the more obese as well as the measurement error affecting self-declared weight and height (and hence BMI) which tends to rise with age (Gil and Mora, 2011)

[10] The exception here is the variable family income, which is missing for a non-trivial proportion of the sample (roughly 20%). Its adjustment and potential effects on our estimates will be examined in the robustness checks section.

[11] Ceuta and Melilla were excluded due to their very low representativeness in the dataset, as were the Balearic and Canary Islands since we assume they may have different influences due to their geographical position in contrast to Spanish inland territory.

[12] Notice that others question the superiority of these alternatives proxies of adiposity and conclude they tend to complement BMI (ProspectiveStudiesCollaboration, 2009).

status (SES), and 3) lifestyle variables (see Appendix: Table 4.8). Specifically, we consider several dummies for age cohorts, the number of children in the household and a dummy for being married for the first group of controls. For the second group we proxy socioeconomic status with years of schooling, net family income in intervals and with a dummy variable for being employed. Since both lifestyles and food habits have been identified as key obesity-risk factors in the literature, we also include indicators for sedentary behaviour at work, physical activity during leisure time, daily smoking, alcohol consumption and consumption of meat, fruits, vegetables and legumes as our last group of controls.[13]

Figure 4.1 exhibits average BMI by ACs for the pooled sample and by gender, where it is evident that geographical differences in BMI are much more pronounced for women. Since our aim consists in disentangling the BMI between northern and southern Spanish regions, we divided Spain into three groups. The group named "South" consists of the regions or Autonomous Communities of Andalusia, Extremadura and Murcia and the second group, named "North", comprises Asturias, Cantabria, Galicia, Navarra, the Basque Country and Rioja. The remaining continental Spanish regions are considered to form part of the centre of the country and are excluded from our empirical analysis. Notwithstanding, the results obtained under other grouping of regions will be analysed in the robustness checks section. Table 4.1 shows the resulting two groups of regions, with the corresponding observations contained in the estimation sample and some basic descriptive statistics for BMI. We report a statistically significant difference of 0.34 units in mean BMI between the South (25.96 kg/m^2) and the North (25.62 kg/m^2).

[13] Specifically, we measured the consumption of fruits, vegetables and legumes (meat) of between 4 to 6 times per week or higher intakes (less than once per week or never) as high (low) frequency of consumption.

Figure 4.1: Average BMI per region and by gender.

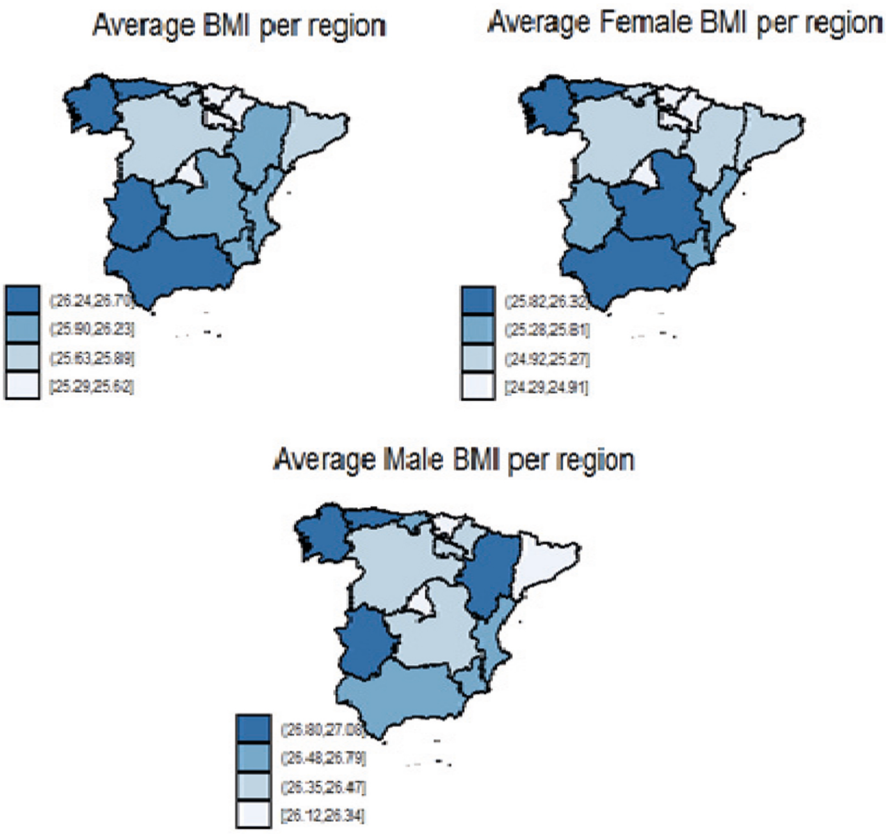


Table 4.1: Groups of regions

	Sample size	%	Mean BMI	S.D BMI
		South		
Andalucía	1.569	57.92	26.109	4.554
Extremadura	553	20.41	26.075	4.396
Murcia	587	21.67	25.728	4.201
Total	**2709**	**100**	**25.955**	**4.469**
		North		
Asturias	520	17.93	26.103	4.643
Cantabria	330	11.38	25.717	4.496
Galicia	585	20.17	26.504	4.833
Navarra	457	15.76	25.233	4.143
Basque Country	676	23.31	25.202	4.414
Rioja	332	11.45	25.004	3.986
Total	**2900**	**100**	**25.621**	**4.455**

4.3.1 Descriptive statistics

Tables 4.2 and 4.3 report the sample means of the BMI indicator and its determinants differentiating by regional group, for women and men respectively. As can be appreciated, there are substantial differences in the endowment of characteristics between the two groups of regions, which are generally statistically significant and more pronounced for women.

Specifically, in Table 4.2, we document a large and significant difference in mean weight level of around 1.59 kg (average height is roughly the same) between women in the South and the North. As a result, the South to North BMI gap amounts to a significant 0.55 kg/m^2 (0.12 standard deviations apart). In terms of household composition, a higher proportion of females in the South are married compared to those living in the North. Interestingly, the data show the existence of a large and significant difference in years of schooling, with females residing in the North having almost 1.5 extra years of schooling (11.09 vs 10.48). Similarly, noticeable differences to the detriment of females living in the South exist regarding income and working status endowments. With respect to lifestyle characteristics, women in the South are less likely to work in a sedentary job compared to their counterparts in the North and are more likely to smoke on a daily basis and drink less alcohol per week. In terms of food habits, women in the South tend to consume less red meat (26% vs 37%) and less fruit. Differences in the consumption of vegetables and legumes among women are not statistically significant between the two groups of regions. Table 4.3 exhibits the same descriptive statistics for males. Interestingly, we evidence the absence of any significant difference in BMI across the two areas. Less remarkable differences in endowments between the South and the North are shown as well.

Table 4.2: Descriptive statistics by groups of regions for Women

	South		North		
Variables	Mean	s.d.	Mean	s.d.	Diff. South-North
Height	162.03	6.40	161.86	6.33	0.17
Weight	66.30	12.09	64.71	12.33	1.59***
BMI	25.29	4.62	24.74	4.79	0.55***
Sociodemographic characteristics					
Age: 18-35	0.24	0.42	0.23	0.41	0.01
Age: 36-45	0.28	0.45	0.28	0.44	0.00
Age: 46-55	0.26	0.43	0.26	0.44	-0.00
Age: 55-65	0.22	0.41	0.23	0.43	-0.01
Household composition					
Married	0.62	0.48	0.57	0.49	0.05***
Kids	0.60	0.83	0.49	0.74	0.11***
Socioeconomic status					
Schooling	10.48	4.53	11.93	4.27	-1.45***
Income1	0.33	0.44	0.18	0.33	0.14***
Income2	0.27	0.41	0.21	0.36	0.05***
Income3	0.25	0.41	0.29	0.41	-0.04**
Income4	0.11	0.29	0.23	0.38	-0.12***
Income5	0.05	0.20	0.09	0.26	-0.04***
Working	0.46	0.49	0.61	0.48	-0.15***
Lifestyle variables					
Sedentary job	0.28	0.44	0.31	0.46	-0.04**
Weekly sport activities	0.10	0.30	0.09	0.27	0.01
Daily smoker	0.27	0.44	0.23	0.42	0.04**
Weekly alcohol consumption	2.30	5.15	3.24	6.16	-0.95***
Food habits variables					
Meat	0.26	0.43	0.37	0.47	-0.11***
Fruit	0.76	0.42	0.80	0.39	-0.05***
Vegetables	0.73	0.43	0.74	0.44	-0.01
Legumes	0.06	0.24	0.07	0.22	-0.01
Number of observations	**1366**		**1494**		

Table 4.3: Descriptive statistics by groups of regions for Men

	South		North		
Variables	mean	s.d.	mean	s.d.	Diff. south-north
Height	174.04	7.13	174.24	6.85	-0.20
Weight	81.01	13.21	80.95	12.61	0.06
BMI	26.76	4.17	26.65	3.83	0.13
Sociodemographic characteristics					
Age: 18-35	0.26	0.43	0.19	0.39	0.07***
Age: 36-45	0.28	0.44	0.29	0.44	-0.01
Age: 46-55	0.27	0.44	0.25	0.44	0.02
Age: 55-65	0.18	0.39	0.27	0.44	-0.08***
Household composition					
Married	0.63	0.48	0.59	0.49	0.04*
Kids	0.57	0.82	0.45	0.72	0.12***
Socioeconomic status					
Schooling	10.01	4.10	11.19	4.01	-1.18***
Income1	0.31	0.43	0.15	0.30	0.17***
Income2	0.27	0.42	0.22	0.37	0.05***
Income3	0.24	0.41	0.29	0.42	-0.05***
Income4	0.12	0.31	0.25	0.39	-0.13***
Income5	0.06	0.22	0.10	0.27	-0.04***
Working	0.62	0.48	0.66	0.47	-0.05**
Lifestyle variables					
Sedentary job	0.30	0.46	0.31	0.46	-0.01
Weekly sport activities	0.16	0.36	0.14	0.33	0.02
Daily smoker	0.33	0.47	0.30	0.46	0.03
Weekly alcohol consumption	8.47	13.38	9.74	14.02	-1.27**
Food habits variables					
Meat	0.31	0.45	0.41	0.48	-0.10***
Fruit	0.70	0.45	0.69	0.46	0.02
Vegetables	0.62	0.48	0.60	0.49	0.02
Legumes	0.07	0.23	0.09	0.27	-0.02**
Number of observations	**1343**		**1406**		

4.4 Empirical Methodology

4.4.1 Average BMI Differentials between groups of regions.

Since only the descriptive statistics do not give us a clear picture of the ceteris paribus effects, nor the contribution of each factor on the BMI difference between the groups, we proceed first by running a simple OLS regression which explains BMI as a function of a vector of control variables (X_i) divided into the three main groups we mentioned before, namely 1) sociodemographic variables, 2) SES, and 3) lifestyle variables. We estimate the equation separately for Southern and Northern regions, that is

$$bmi_i^S = \alpha^S + \beta^S X_{iS} + u_i^S \tag{4.1}$$

$$bmi_i^N = \alpha^N + \beta^N X_{iN} + u_i^N \tag{4.2}$$

where the superscripts S and N indicate that the corresponding estimates are allowed to be different for South and North, respectively. Next, with the aim of appreciating the contribution of the covariates on the observed BMI disparities between the groups of regions, we utilize the Oaxaca-Blinder (OB) decomposition (Oaxaca, 1973; Blinder, 1973). This widely used decomposition method disentangles average outcome differentials into the contribution of the (average) endowment of observable characteristics (i.e. the explained or composition component) and the contribution of unexplained factors or structure effect (which is captured by differences in the estimated coefficients). Furthermore, as suggested by Fortin (2008) and Fortin et al. (2011) , we estimate the non-discriminatory reference BMI structure from a pooled regression with all the selected regions together, imposing an identification restriction that ensures that the BMI advantage of one group of regions equals the disadvantage suffered by the other group, that is:

$$bmi_i = \alpha + \beta' X_i + \gamma_N I\left(N = 1\right) + \gamma_S I\left(S = 1\right) + u_i \tag{4.3}$$

$$subject\ to\ \ \gamma_S + \gamma_N = 0$$

Equation (2) is estimated using the pooled sample, and contains indicators for belonging to the North or South (L = 1 if South, 0 if North). The estimated vector of β coefficients thus represents the nondiscriminatory BMI structure that is used in the decomposition. From the estimates of equation (2) we decompose the raw BMI differentials between groups of regions into different components as follows:

$$\overline{bmi^S} - \overline{bmi^N} = \left(\overline{X^S} - \overline{X^N}\right)\widehat{\beta} + (\widehat{\gamma}_S - \widehat{\gamma}_N) + E\left[u_i|N=1\right] - E\left[u_i|S=1\right] =$$
$$= \left(\overline{X^S} - \overline{X^N}\right)\widehat{\beta} + [\underbrace{\left(\overline{X^S}\left(\widehat{\beta}_S - \widehat{\beta}\right) + (\widehat{\alpha}_S - \widehat{\alpha})\right)}_{\widehat{\gamma}_S} - \underbrace{\left(\overline{X^N}\left(\widehat{\beta}_N - \widehat{\beta}\right) + (\widehat{\alpha}_N - \widehat{\alpha})\right)}_{\widehat{\gamma}_N}] \tag{4.4}$$

The term $\left(\overline{X^S} - \overline{X^N}\right)\widehat{\beta}$ represents the composition effect (i.e. share of average BMI gap due to differences in observable characteristics), whereas the term $(\widehat{\gamma}_S - \widehat{\gamma}_N) = \left(\overline{X^S}\left(\widehat{\beta}_S - \widehat{\beta}\right) + (\widehat{\alpha}_S - \widehat{\alpha})\right) - \left(\overline{X^N}\left(\widehat{\beta}_N - \widehat{\beta}\right) + (\widehat{\alpha}_N - \widehat{\alpha})\right)$ corresponds to the part of the mean BMI differential that can be attributed to different coefficients or returns to observable characteristics across regions (including the intercept). [14]

It seems worth noticing that the term $E\left[u_i|N=1\right] - E\left[u_i|S=1\right] =$is assumed to be zero, which corresponds to the hypothesis that differences in unobservable determinants of BMI are the same between the two groups of regions. There are, however, at least two possible situations that might invalidate this hypothesis. First, unobservable characteristics such as cultural differences, and the quality of school and health services might be more favourable for residents in northern (and richer) regions. Second, there could be endogenous residential sorting, due to the fact that the movers with better unobservable characteristics could be more likely to migrate from the south to the north. In any of these two cases, the difference in the coefficients between northern and southern regions would be overestimated, due to the potential upward bias in the coefficients of the former group. Unfortunately, we are unable to explicitly gauge the relevance of this possible issue (and eventually address it) with the available data. However, as we show later, the contribution of differences in coefficients in rather limited in this specific application, which is reassuring. Moreover, in the discussion of the results we mostly focus on the analysis of differences in endowments between regions.

4.4.2 Distributional BMI Differentials

One limitation of the OB decomposition is that it provides evidence about average BMI differences across the groups of regions, whereas by focusing only on average gaps one may miss important differences at other points of the BMI distribution (es-

[14] Moreover, the two components of the OB decomposition can be further divided into the contribution of each specific covariate and the corresponding coefficient (detailed decomposition), which can eventually also be aggregated into subgroups (as explained later). However, the presence of categorical variables makes the results of the detailed decomposition dependent on the choice of the reference category. This issue can be avoided by "normalizing" the effects of discrete covariates as explained in Jann (2008).

pecially at the top, corresponding to obesity and severe/morbid obesity categories). Therefore, we investigate distributional BMI differences by means of the Unconditional Quantile Regression (UQR) method proposed by Firpo et al. (2009). [15] The UQR is based on the statistical concept of the Influence Function (IF), which represents the influence of an individual observation on a distributional statistic of interest (e.g. the quantile). By adding back the statistic to the corresponding IF, it is possible to obtain the Recentered Influence Function (RIF) for each quantile of the outcome. The RIF Regression estimates the marginal effects of a set of characteristics on an unconditional distributional statistic of an outcome variable. The RIF for the th quantile (q_τ) of BMI corresponds to,

$$RIF\,(bmi_i,\; q_\tau) = q_\tau + IF\,(bmi_i,\; q_\tau) = q_\tau + \frac{\tau - I(bmi_i \leq\; q_\tau)}{f_{bmi_i}(q_\tau)} \tag{4.5}$$

where $f_{bmi_i}(q_\tau)$ is the unconditional density of BMI evaluated at the th quantile and I an indicator function. By replacing the unknown elements of equation (4.5) by their sample estimators it is possible to obtain an estimate of the RIF, which is,

$$\widehat{RIF}\,(bmi_i,\; q_\tau) = \hat{q}_\tau + \widehat{IF}\,(bmi_i,\; q_\tau) = \hat{q}_\tau + \frac{\tau - I(bmi_i \leq\; \hat{q}_\tau)}{\hat{f}_{bmi_i}(\hat{q}_\tau)} \tag{4.6}$$

where $\hat{f}_{bmi}(\hat{q}_\tau)$ corresponds to a Kernel density estimator of the unconditional density function of the outcome. The RIF for a given quantile can be taken as a linear approximation of the nonlinear function of the quantile, and captures the change of the (unconditional) quantile of the outcome in response to a change in the underlying distribution of the covariates (Firpo et al., 2009). It can be shown that the expected value of the RIF for selected quantiles of the unconditional distribution of BMI ($\hat{q}_\tau$) can be modelled to be a linear function of explanatory variables, as in a standard linear regression.

Given the linear approximation of the conditional expectation of the RIF and the theoretical property stating that the average $\overline{RIF}\,(bmi_i,\; \hat{q}_\tau)$ is equal to the corresponding marginal quantile of the distribution of the outcome, is it possible to generalize the standard OB decomposition of average outcomes to a distributional decomposition applied to the unconditional distribution of the outcome (see Firpo et al., 2009 and Fortin et al., 2011 for technical details). In other words, it is possible to examine the contribution of both the endowment of observable characteristics and the returns to these characteristics, in explaining the estimated unconditional BMI gap across groups of regions, applying the decomposition for average outcomes de-

[15] Notice that the same potential pitfalls regarding unobservable characteristics and residential sorting highlighted for the OB decomposition could also apply to the one based on RIF regressions.

scribed by equation (3) to the RIF, that is:

$$\hat{q}_{S\tau} - \hat{q}_{N\tau} = \overline{RIF}\left(bmi^{S},\, \hat{q}_{\tau}\right) - \overline{RIF}\left(bmi^{N},\, \hat{q}_{\tau}\right) =$$

$$\left(\overline{X^{S}} - \overline{X^{N}}\right)\widehat{\beta}_{\tau} + \left[\left(\overline{X^{S}}\left(\widehat{\beta}_{S\tau} - \widehat{\beta}_{\tau}\right) + \left(\widehat{\alpha}_{S\tau} - \widehat{\alpha}_{\tau}\right)\right) - \left(\overline{X^{N}}\left(\widehat{\beta}_{N\tau} - \widehat{\beta}_{\tau}\right) + \left(\widehat{\alpha}_{N\tau} - \widehat{\alpha}_{\tau}\right)\right)\right] \tag{4.7}$$

Here $\widehat{\beta}_{\tau}$ corresponds to the nondiscriminatory BMI structure (estimated from a pooled RIF regression) at quantile estimated in a similar fashion as equation (4.3). Similar to equation (4.4), the term $\left(\overline{X^{S}} - \overline{X^{N}}\right)\widehat{\beta}_{\tau}$ represents the composition effect and the term $\left(\overline{X^{S}}\left(\widehat{\beta}_{S\tau} - \widehat{\beta}_{\tau}\right) + (\widehat{\alpha_{S\tau} - \alpha_{\tau}})\right) - \left(\overline{X^{N}}\left(\widehat{\beta_{N\tau} - \beta_{\tau}}\right) + (\widehat{\alpha_{N\tau} - \alpha_{\tau}})\right)$ captures the unexplained component of BMI differential evaluated at the -quantile of the unconditional distribution of BMI. There are several advantages of this method. Its computational cost is minimal and it provides path independent detailed decompositions of both components.

4.5 Results

OLS Estimates

Table 4.4 shows the OLS estimates of the BMI determinants separately for the South and the North and distinguishing by gender. The findings point out the existence of a heterogeneous pattern of correlates between BMI and its covariates both across regions and by gender. That is, control variables affect individual BMI differently depending on the group of regions the person belongs to and on whether they are females or males. Certainly, Table 4.4 evidences a positive age gradient in mean BMI in both regions and for both genders, however this effect is comparatively stronger for females. In terms of household composition, being married is only significant for men in the South group of regions, while number of children in the household is a significant control (with a negative impact on BMI) only for women and men in Northern regions. Schooling exerts the expected negative effect on mean BMI, while on the contrary, family income barely affects BMI regardless of geographical location and gender.Statistically significant coefficients and with the expected sign are also found for other key BMI determinants. More specifically, working in a sedentary job has a positive and statistically significant conditional association with BMI for females in both areas of the country, although for males this variable is only significant in the South. Regarding physical activity during leisure time, the estimates for females indicate a negative association with BMI only in Northern regions, whereas for males doing sports at least once per week is negatively associated with BMI, with a stronger effect in the South. Finally, daily smokers exhibit lower BMI levels, as widely reported in the literature (Dare et al., 2015).

Table 4.4: BMI Estimations: OLS Results

	Women		**Men**	
	South	**North**	**South**	**North**
Constant	26.515***	26.695***	26.090***	26.655***
	(0.543)	*(0.654)*	*(0.459)*	*(0.499)*
Sociodemographic characteristics				
Age: 18-35 (reference category)	reference category			
Age: 36-45	0.514	0.670*	1.043***	1.756***
	(0.365)	*(0.335)*	*(0.310)*	*(0.308)*
Age: 46-55	1.499***	1.218***	1.276***	2.107***
	(0.357)	*(0.351)*	*(0.331)*	*(0.307)*
Age: 55-65	2.465***	1.397**	1.683***	1.984***
	(0.409)	*(0.440)*	*(0.390)*	*(0.335)*
Household composition				
Married	0.450	0.557	1.109***	0.438
	(0.290)	*(0.289)*	*(0.289)*	*(0.249)*
Kids	-0.153	-0.595***	0.018	-0.503***
	(0.175)	*(0.167)*	*(0.152)*	*(0.145)*
Socioeconomic status				
Schooling	-0.220***	-0.240***	-0.076*	-0.115***
	(0.036)	*(0.041)*	*(0.032)*	*(0.030)*
Income1 (reference category)	reference category			
Income2	0.250	-0.259	0.655*	-0.163
	(0.326)	*(0.409)*	*(0.320)*	*(0.383)*
Income3	-0.221	-0.632	0.151	0.159
	(0.338)	*(0.419)*	*(0.351)*	*(0.378)*
Income4	-0.269	-1.159**	-0.477	-0.234
	(0.471)	*(0.444)*	*(0.399)*	*(0.394)*
Income5	-0.637	-1.534**	0.137	0.287
	(0.507)	*(0.566)*	*(0.494)*	*(0.504)*
Working	-0.367	-0.265	-0.653*	0.021
	(0.261)	*(0.267)*	*(0.275)*	*(0.258)*
Lifestyle variables				
Sedentary job	0.771**	0.781**	0.641*	0.223
	(0.288)	*(0.301)*	*(0.259)*	*(0.227)*
Weekly sport activities	-0.580	-0.724*	-1.272***	-1.047***
	(0.373)	*(0.348)*	*(0.246)*	*(0.238)*
Daily smoker	-0.521	-0.807**	-0.820***	-0.673**
	(0.278)	*(0.287)*	*(0.241)*	*(0.238)*
Weekly alcohol consumption	-0.046*	0.005	0.013	-0.001
	(0.021)	*(0.018)*	*(0.009)*	*(0.008)*
Food habits variables				
Meat	0.334	0.401	0.136	0.174
	(0.287)	*(0.252)*	*(0.248)*	*(0.203)*
Fruits	-0.145	0.707*	-0.286	0.112
	(0.311)	*(0.299)*	*(0.256)*	*(0.230)*
Vegetables	0.248	0.265	0.546*	-0.312
	(0.281)	*(0.284)*	*(0.226)*	*(0.214)*
Legumes	-0.326	-0.795	-0.033	0.137
	(0.447)	*(0.466)*	*(0.481)*	*(0.362)*
R-squared	0.156	0.138	0.114	0.097
Number of Observations	1366	1494	1343	1406

OB decomposition results

Tables 4.5 and 4.6 present the aggregated and detailed OB decomposition results respectively, differentiating by gender. The decomposition analysis shown in Table 4.4 evidences that up to 73% (0.40 BMI units) of the overall South to North mean BMI gap for women (0.55 BMI units) is due to differences in endowments, whereas the remaining 27% (0.15 BMI units) is due to the differences in coefficients or returns to BMI determinants. This finding indicates that a policy intervention addressed to equalize certain endowments across regions (particularly schooling) would reduce the mean BMI gap among women quite significantly. Interestingly, the results show that while the explained part is mostly driven by more disadvantaged SES endowments of women living in the South, lifestyle characteristics are to the detriment of females living in the North. Moving on to the detailed decomposition (see Table 4.5), we identify differentials in average years of schooling as by far the single most important contributor in explaining the greater mean BMI level for Southern women. Differences in income are also relevant factors in the explained part to the detriment of women in the South, but with a more modest contribution. In contrast, healthy (unhealthy) lifestyles such as low consumption of meat (daily smoking), as well as the number of children in the household are in favour of women living in the South (though their contribution is low). As shown in Table 4.5, as a whole, the unexplained part or returns to certain characteristics, which accounts for 27% of the total gap for females, is not statistically significant. The OB decomposition analysis suggests that the average BMI differential across regions for males is small (0.11 BMI points) and insignificant. The contribution of explained factors is also insignificant for males, since the more advantaged endowment of SES variables for those residing in the north of the country tends to compensate with their unfavourable distribution of sociodemographic and lifestyle variables (relative to men residing in southern regions).

Table 4.5: Oaxaca Decomposition (Aggregated Results)

	Women		**Men**	
Overall decomposition	Mean BMI	z-stat	Mean BMI	z-stat
South regions	25.29	203.18	26.76	235.69
North regions	24.74	198.12	26.65	257.39
BMI Difference (south-north)	**0.55**	**3.11**	**0.111**	**0.72**
Explained difference	**0.40**	**4.68**	**-0.04**	**-0.58**
Sociodemographic characteristics	-0.04	-1.08	-0.12	-3.19
SES	0.53	8.02	0.15	3.39
Lifestyle	-0.09	-2.65	-0.07	-2.60
Unexplained difference	**0.15**	**0.87**	**0.15**	**0.99**
Sociodemographic characteristics	0.14	0.62	0.64	3.02
SES	0.07	0.10	0.05	0.10
Lifestyle	-0.72	-1.58	0.40	1.09
Constant	0.66	0.74	-0.94	-1.31

Table 4.6: Oaxaca Decomposition (Detailed Results)

	Women				Men			
Overall Decomposition	Mean BMI		z-stat		Mean BMI		z-stat	
South regions	25.29		203.18		26.76		235.69	
North regions	24.74		198.12		26.64		257.39	
BMI Difference (south-north)	**0.55**		**3.11**		**0.11**		**0.72**	
	Explained		**Unexplained**		**Explained**		**Unexplained**	
	Coef.	**z-stat**	**Coef.**	**z-stat**	**Coef.**	**z-stat**	**Coef.**	**z-stat**
	0.40	**4.68**	**0.15**	**0.87**	**-0.04**	**-0.58**	**0.15**	**0.99**
Sociodemographic characteristics	-0.03	-1.08	0.14	0.62	-0.12	-3.19	0.64	3.02
Age: 18-35	-0.01	-0.32	-0.07	-0.94	-0.08	-3.94	0.10	1.62
Age: 36-45	-0.00	-0.32	-0.13	-1.36	-0.00	-0.20	-0.07	-0.90
Age: 46-55	-0.00	-0.16	-0.00	-0.06	0.01	0.90	-0.09	-1.42
Age: 55-65	-0.01	0.01	0.17	2.01	-0.48	-2.98	0.03	0.48
Married	0.02	1.79	-0.06	-0.26	0.02	1.68	0.41	2.77
Kids	-0.04	-2.35	0.23	1.83	-0.25	-1.74	0.25	2.48
SES	0.53	8.02	0.07	0.10	0.15	3.39	0.05	0.10
Schooling	0.33	6.09	0.23	0.37	0.11	3.87	0.41	0.90
Income1	0.06	2.14	-0.14	-1.35	-0.16	-0.52	-0.01	-0.14
Income2	0.02	2.13	-0.01	-0.12	0.01	1.34	0.17	2.42
Income3	-0.00	-0.33	-0.03	-0.43	-0.01	-0.87	-0.02	-0.34
Income4	0.04	1.73	0.05	0.99	0.05	2.52	-0.05	-1.15
Income5	0.03	2.30	0.02	0.80	-0.00	-0.58	-0.02	-0.61
Working	0.05	1.71	-0.55	-0.28	0.01	1.37	0.43	-1.80
Lifestyle variables	-0.09	-2.65	-0.72	-1.58	-0.07	-2.60	0.40	1.09
Sedentary job	-0.03	-1.83	-0.00	-0.02	-0.00	-0.55	0.13	1.22
Weekly sport activities	-0.00	-0.68	0.01	0.29	-0.02	-1.53	-0.03	-0.66
Daily smoker	-0.03	-1.99	0.07	0.72	-0.02	-1.43	-0.05	-0.44
Weekly alcohol consumption	0.02	1.25	-0.13	-1.95	-0.01	-0.96	0.13	1.27
Meat	-0.04	-1.89	-0.20	-0.18	-0.02	-1.09	-0.01	-0.10
Fruits	-0.01	-1.14	-0.66	-1.99	-0.00	-0.33	-0.27	-1.16
Vegetables	0.00	-0.49	-0.12	-0.04	0.00	0.62	0.52	2.77
Legumes	0.00	0.57	0.03	0.73	-0.00	-0.25	-0.01	-0.28
Constant			0.66	0.74			-0.94	-1.31

In what follows, we move a step ahead from the simple decomposition of average differentials and by means of RIF-regressions we disentangle the factors behind the North to South gap for males and females over the entire unconditional distribution of BMI.

4.5.1 RIF decomposition results

Figure 4.2 presents the aggregated RIF decomposition results separately for women and men at the different deciles of the unconditional distribution of BMI. Since we obtained no evidence of significant regional gaps at any point of the BMI distribution for men, we show the tables of the aggregated and the detailed RIF-decomposition results only for women (Table 4.7 and Table 4.9 of the Appendix [16] respectively). As shown in Table 4.7, BMI differences in women between the two sets of regions appear to be quite stable from Q2 to Q8 (except Q7) since the increasing contribution of the explained part tend to be compensated with the decreasing contribution of unexplained factors. Interestingly, however the data also reveal that the explained (unexplained) portion of the gap steadily increases (decreases) over the quantiles, revealing that what really matters to deal with the obesity epidemic among overweight women is to focus the attention on regional disparities in endowments. Note that the contribution of differences in observable characteristics is always statistically significant and reaches its highest values at the 8th and 9th deciles that correspond to high levels of overweight or pre-obesity statuses among women.

The pattern regarding the separate contribution of groups of covariates is in line with what we obtained from the decomposition of average differentials. It seems worth noticing the significant increase in the role played by SES across the distribution of BMI. More specifically, the positive contribution of observable characteristics is mainly driven by schooling, while income and employment status also contribute to the explained part of the difference. We show that schooling is the main contributor throughout the BMI distribution whereas, on the contrary, food habits (mainly meat consumption) are detrimental for women in the North. While women in the North are worse off with respect to their lifestyle habits (eating habits and physical activity), they have much more advantaged endowments with respect to their SES status and, as a result, they exhibit lower BMI values. Overall, the unexplained factors are significant at the bottom tail, while most of the difference at the higher parts of the distribution is attributed to explained factors. The evidence from the RIF de-

[16] Table 4.9 in the Appendix contains the detailed RIF decomposition results for the top three deciles of the unconditional distribution of BMI for women as they correspond to overweight and obesity statuses. Detailed results for all the deciles are available upon request.

composition suggests that policies aimed at enhancing women's human capital in the South of Spain could reduce the prevalence of overweight and obesity problems and favour convergence to the relatively lower values observed in northern Spanish regions.

Figure 4.2: RIFR Decomposition

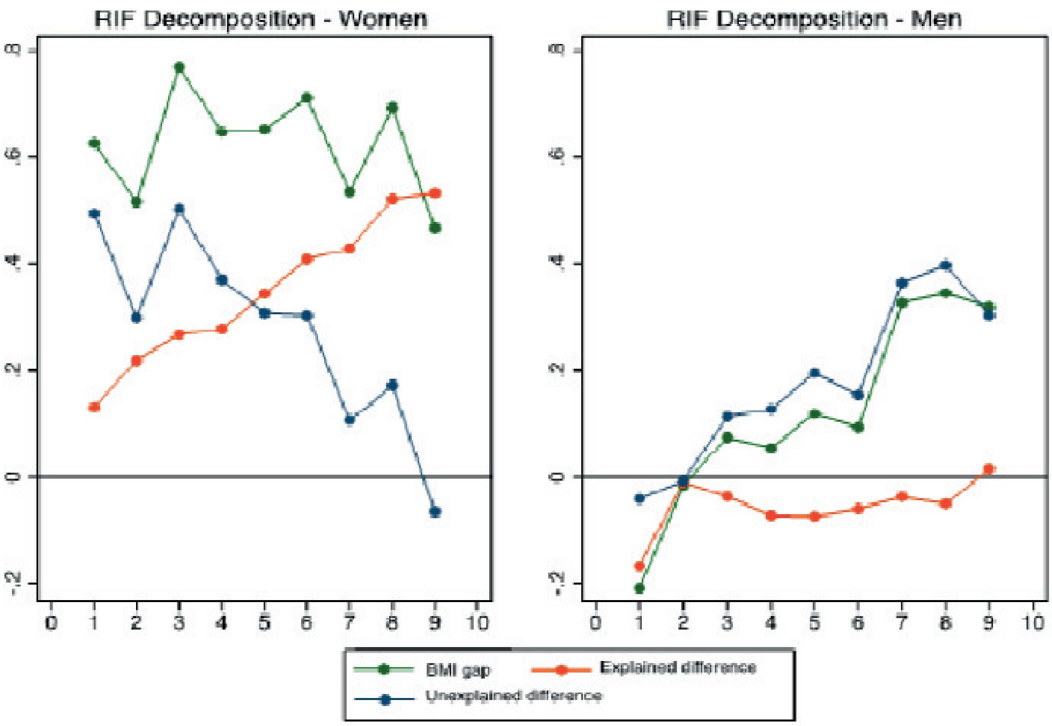


Table 4.7: Quantile Decomposition (Women)

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
South	20.24	21.46	22.49	23.44	24.44	25.61	26.94	28.74	31.60
z-stat	169.01	179.32	179.39	174.66	163.51	152.46	145.59	124.65	91.31
North	19.73	20.83	21.83	22.73	23.81	24.93	26.39	28.04	31.28
z-stat	194.59	187.74	184.38	180.20	169.86	154.74	146.05	126.03	108.23
BMI Difference (south-north)	**0.50**	**0.63**	**0.66**	**0.71**	**0.63**	**0.68**	**0.55**	**0.70**	**0.32**
z-stat	3.22	3.87	3.85	3.86	3.10	2.92	2.13	2.19	0.71
Explained difference	**0.11**	**0.15**	**0.24**	**0.30**	**0.36**	**0.50**	**0.59**	**0.71**	**0.77**
z-stat	1.84	2.22	3.13	3.57	3.81	4.78	5.21	5.25	4.27
Sociodemographics	0.01	0.01	0.00	-0.03	-0.06	-0.05	-0.04	-0.05	-0.09
z-stat	0.56	0.18	0.07	-0.60	-1.38	-1.14	-1.01	-0.99	-1.53
SES	0.13	0.18	0.27	0.37	0.52	0.63	0.75	0.87	0.99
z-stat	2.82	3.65	5.19	6.50	7.53	7.92	8.10	7.85	6.83
Lifestyle habits	-0.03	-0.03	-0.03	-0.05	-0.10	-0.09	-0.12	-0.11	-0.13
z-stat	-0.93	-1.10	-1.08	-1.41	-2.48	-2.03	-2.50	-1.89	-1.61
Unexplained difference	**0.39**	**0.48**	**0.42**	**0.41**	**0.27**	**0.18**	**-0.04**	**-0.01**	**-0.45**
z-stat	2.41	2.88	2.46	2.26	1.37	0.79	-0.14	-0.03	-0.99
Sociodemographics	0.33	0.08	-0.01	0.01	0.17	0.26	0.06	0.10	-0.29
z-stat	1.52	0.35	-0.04	0.05	0.65	0.86	0.19	0.24	-0.48
SES	0.01	-0.08	-0.41	-0.28	-0.19	-1.18	-0.36	0.28	0.15
z-stat	0.03	-0.15	-0.71	-0.45	-0.28	-1.47	-0.40	0.25	0.09
Lifestyle habits	-0.95	-0.76	-0.45	-0.00	-0.40	-0.34	-1.09	-1.19	-0.98
z-stat	-1.99	-1.71	-0.99	-0.00	-0.75	-0.56	-1.66	-1.44	-0.85
Constant	0.10	1.24	1.29	0.68	0.69	1.43	1.35	0.79	0.67
z-stat	1.36	1.67	1.65	0.81	0.74	1.32	1.12	0.52	0.31

4.5.2 Robustness checks

In this section we consider the robustness of the previous findings with respect to three main issues that might affect our estimations: the presence of missing values in the family income variable, the potential bias in BMI due to the self-reported nature of the variables height and weight in the EHIS survey and the grouping of regions that we adopted in this work.

Regarding the first issue, so far we considered that the relatively high proportion (around 20%) of missing values in net family income is at random, and the corresponding observations were excluded. In order to deal with the potential selectivity bias due to the non-randomness of non-reporting in the household income variable, which is reported in intervals, we repeat the estimation including all the observations, plus an additional dummy variable for missing family income. Moreover, we also replace missing values in income categories with predicted values obtained from an ordered probit model based on demographics, SES status and other information of the head of the household and spouse (Allison, 2001). As it can be appreciated in columns (1) and (2) of Table 4.11 in the Appendix, the overall results from the OB decomposition is mostly unaffected by imputing missing values of family income using the two selected techniques. We only observe a small decrease in the contribution of endowments for females.

Second, we adjust self-declared weight, height and the subsequent computation of BMI to deal with the misreporting of such information by adopting the procedure proposed by Gil and Mora (2011). Also under this alternative scenario, the results from the decomposition of average BMI differentials is virtually unaffected, as shown in column (3) of Table E.

Third, we analyse the results obtained under alternative groupings of Spain's regions. In column (4) of Table 4.11 we adopt the Eurostat NUTS-2 classification rather than our ad-hoc classification. This implies adding the region of Aragon in the group of northern regions and Extremadura is excluded from the group of southern regions. Finally, in column (5) we deviate from a purely geographical criterion and we rank regions according to the distribution of GDP per capita in year 2014 (Spanish Regional Accounts, Base 2010, National Statistics Institute - INE). We split regions into three parts of equal size, according to the distribution of GDP per capita and select the top and the bottom groups, corresponding to high-and-low income regions respectively. [17] Interestingly, using the NUTS-2 definition of regions does not alter our results, whereas as expected splitting the regions according to aggregate income per capita increases BMI differentials, which are now significant for both genders, and the unexplained part now becomes preponderant with respect to the explained part.

[17] High-income regions are: Madrid, Basque Country, Navarra, Catalonia and Rioja; whereas low-income regions are: Andalucia, Extremadura, Castilla-La Mancha, Murcia and Galicia. The group of regions belonging to the central part of the distribution are excluded. Notice that the Islands and Ceuta and Melilla territories are excluded from the analysis (see footnote 6).

4.6 Conclusion

This paper investigates the conditioning factors behind the North–South BMI divide in Spain. We use decomposition analysis that enables us to disentangle the contribution of each covariate and the corresponding coefficients to this difference. Starting with the OB decomposition, we reveal that the mean BMI gap between the South and North of Spain is mostly driven by differences between women residing in the two areas of the country. A large and significant part of this regional average gap in BMI (73%) is due to differences in endowments related to SES status (basically years of education), whereas differences in returns to such characteristics play a minor and insignificant role in accounting for the observed BMI differential. [18] Indeed, in view of the epidemic of obesity as a global public health concern, policymakers are mostly interested in designing effective policies against the overweight and obese. Hence, we proceed with the distributional analysis and the corresponding decomposition, since the findings at the upper tail of the BMI distribution are the ones actually capturing overweight and obesity problems. Interestingly, we evidence that differences in SES endowments and particularly schooling explain a very significant part of the women's North to South differential (accounting for up to 85% of the gap at the 8th decile) at the top of the BMI distribution. Notice these findings prove to be quite robust to alternative scenarios dealing with missing information, BMI bias and regional grouping.

Therefore, a significant part of the cross-regional BMI gap can be mitigated by implementing the right policies focused on improving human capital. Given that the education gradient in obesity seems to be much stronger in women than in men (as in Devaux et al. (2011)), efforts aimed at improving (years of) schooling for women in the South would substantially mitigate differences in overweight and obesity between the two groups of regions. Such a policy intervention would additionally reduce differences in obesity-related diseases and/or improve health in general, inasmuch as obesity constitutes a key risk factor for many chronic conditions and health complications. However, it must also be stressed that even equalizing the female endowments across the two groups of regions, there would still be a certain differential in BMI that penalizes southern Spanish regions in terms of the prevalence of overweight and obesity problems.

[18] This evidence is reassuring for the usefulness of our results for policymaking, since we are unable to solve the potential caveats due to selection across regions and unobserved heterogeneity. Dealing with this issue will be object of future research, when more detailed data will be made available.

This is in line with the evidence from existing related research, suggesting that regional inequalities in education are responsible for regional health inequalities (Safaei, 2014; Ergin and Kunst, 2015). Ballas et al. (2012) report that inequalities in education between regions observed in several EU countries tend to reinforce inequalities between income, wealth, social status and health, contributing to persistent inter-regional disparities. How educational inequalities translate into income, employment and health disparities through a complex set of mechanisms is a research question beyond the aims of this work. Albeit in this specific paper we do not provide causal evidence, the results exhibit a very strong conditional correlation between education and BMI, being the endowment of the former variable responsible for a substantial share of the gap in BMI between women residing in different Spanish regions. Indeed, this is consistent with the causal evidence obtained by Brunello et al. (2013) for several European countries, indicating that exogenous increases in schooling generate a protective effect for females (but no such causal impact is found for males). Therefore, the causal effect of education in mitigating regional disparities in BMI, overweight and obesity and consequently other health related variables should be further investigated in future research.

4.7 Appendix

Table 4.8: Description of dependent and independent variables

Variables	Description
Dependent Variable	
BMI	weight in kg divided by height in meters squared $((kg)/(m^2))$
Independent Variables	
Demographics	
Age: 18-35	1 when aged 18-35 , 0 otherwise
Age: 36-45	1 when aged 36-45, 0 otherwise
Age: 46-55	1 when aged 46-55, 0 otherwise
Age: 55-65	1 when aged 55 65, 0 otherwise
Male	1 when male, 0 otherwise
Married	1 when married, 0 otherwise
Kids	number of children in the household
SES	
Schooling	years of schooling (derived by the education level)
Income1	net family income lower than 970 euros/month
Income2	net family income ranges from 970 to 1400 euros/month
Income3	net family income ranges from 1401 to 2040 euros/month
Income4	net family income ranges from 2041 to 3280 euros/month
Income5	net family income is higher than 3280 euros/month
Working	1 if working, 0 otherwise
Lifestyle variables	
Sedentary job	1 if working in a sedentary job, 0 otherwise
Weekly sport activities	1 if doing a physical activity many times per week, 0 otherwise
Daily smoker	1 if daily smoker, 0 otherwise
Weekly alcohol consumption (index)	daily alcohol consumption (in grams)
Meat	1 if consumes meat more than 4 times per week, 0 otherwise
Fruits	1 if consumes fruit more than 4 times per week, 0 otherwise
Vegetables	1 if consumes vegetables more than 4 times per week, 0 otherwise
Legumes	1 if consumes legumes more than 4 times per week, 0 otherwise

Table 4.9: Detailed RIFR Decomposition (Women)

	Q7		Q8		Q9	
	mean BMI	**z-stat**	**mean BMI**	**z-stat**	**mean BMI**	**z-stat**
South	26.94	145.59	28.74	124.65	31.60	91.31
North	26.38	146.05	28.04	126.03	31.28	108.23
	Coef.	**z-stat**	**Coef.**	**z-stat**	**Coef.**	**z-stat**
BMI Difference (south-north)	0.55	2.13	0.70	2.19	0.32	0.71
Explained difference	0.59	5.21	0.71	5.25	0.76	4.27
Sociodemographics						
Age: 18-35	-0.01	-0.54	-0.01	-0.54	-0.01	-0.54
Age:36-45	-0.00	-0.13	-0.00	-0.32	-0.00	-0.31
Age:46-55	-0.00	-0.14	-0.00	-0.14	-0.00	-0.16
Age: 55-65	-0.01	-0.72	-0.01	-0.72	-0.01	-0.71
Married	0.03	1.67	0.05	1.88	0.01	0.56
Kids	-0.05	-2.07	-0.07	-2.35	-0.07	-1.97
SES						
Schooling	0.38	5.52	0.44	5.30	0.54	4.77
Income1	0.15	3.27	0.14	2.69	0.18	2.39
Income2	0.04	2.32	0.05	2.28	0.05	1.73
Income3	-0.00	-0.07	0.00	0.07	0.01	0.49
Income4	0.05	1.50	0.09	2.44	0.11	2.34
Income5	0.06	3.02	0.05	2.44	0.04	1.85
Working	0.07	1.68	0.09	1.60	0.06	0.82
Lifestyle habits						
Sedentary job	-0.03	-1.72	-0.04	-1.80	-0.08	-1.88
Weekly sport activities	-0.00	-0.64	-0.00	-0.63	-0.00	-0.53
Daily smoker	-0.03	-1.84	-0.02	-1.34	-0.02	-1.11
Weekly alcohol consumption	0.02	0.79	0.04	1.63	0.10	2.58
Meat	-0.06	-1.83	-0.07	-1.83	-0.09	-1.65
Fruits	-0.02	-1.18	-0.01	-0.71	-0.04	-1.34
Vegetables	-0.00	-0.49	-0.00	-0.48	0.00	0.27
Legumes	0.00	0.56	0.00	0.56	0.00	0.22
Unexplained difference	-0.04	-0.14	-0.01	-0.03	-0.44	-0.99
Sociodemographics						
Age: 18-35	0.01	0.14	-0.00	-0.03	-0.37	-2.00
Age:36-45	-0.10	-0.73	-0.10	-0.61	-0.14	-0.61
Age:46-55	-0.19	-1.70	-0.16	-1.09	0.19	0.90
Age: 55-65	0.23	1.86	0.22	1.39	0.29	1.24
Married	0.04	0.11	0.07	0.16	-0.55	-0.86
Kids	0.07	0.38	0.08	0.32	0.29	0.92
SES						
Schooling	-0.01	-0.02	0.22	0.22	0.73	0.50
Income1	-0.28	-1.90	-0.00	-0.03	0.05	0.19
Income2	-0.03	-0.29	-0.17	-1.15	-0.20	-0.90
Income3	-0.14	-1.18	-0.14	-0.94	0.16	0.83
Income4	0.16	1.98	0.10	1.02	0.04	0.37
Income5	0.05	1.00	0.04	0.68	-0.00	-0.14
Working	-0.09	-0.32	0.24	0.64	-0.63	-1.23
Lifestyle habits						
Sedentary job	0.06	0.38	-0.01	-0.03	0.09	0.28
Weekly sport activities	0.04	0.56	-0.07	-0.76	-0.08	-0.64
Daily smoker	0.01	0.07	-0.00	-0.02	0.13	0.50
Weekly alcohol consumption (index)	-0.15	-1.24	-0.13	-0.94	-0.13	-0.78
Meat	-0.10	-0.58	-0.00	-0.03	0.23	0.77
Fruits	-0.91	-1.80	-1.25	-2.00	-1.18	-1.40
Vegetables	-0.09	-0.22	0.21	0.39	-0.01	-0.01
Legumes	0.04	0.59	0.06	0.81	-0.03	-0.22
Constant	1.35	1.12	0.79	0.52	0.67	0.31

Table 4.10: Detailed RIFR Decomposition (Men)

	Q7		Q8		Q9	
	mean BMI	**z-stat**	**mean BMI**	**z-stat**	**mean BMI**	**z-stat**
South	28.38	195.17	29.67	154.85	31.95	121.48
North	28.08	192.33	29.59	162.86	31.75	133.39
	Coef.	**z-stat**	**Coef.**	**z-stat**	**Coef.**	**z-stat**
BMI Difference (south-north)	0.30	1.46	0.08	0.29	0.20	0.57
Explained difference	0.02	0.18	-0.00	-0.01	0.10	0.67
Sociodemographics						
Age: 18-35	-0.08	-3.58	-0.08	-3.31	-0.10	-3.23
Age:36-45	-0.00	-0.20	-0.00	-0.17	-0.00	-0.20
Age:46-55	0.01	0.88	0.00	0.74	0.00	0.58
Age: 55-65	-0.04	-2.07	-0.06	-2.31	-0.06	-1.80
Married	0.03	1.58	0.03	1.49	0.00	0.27
Kids	-0.04	-1.97	-0.05	-2.19	-0.04	-1.33
SES						
Schooling	0.15	3.81	0.17	3.46	0.22	3.32
Income1	-0.02	-0.62	-0.02	-0.42	0.05	0.66
Income2	0.02	1.65	0.02	1.40	0.01	0.87
Income3	-0.00	-0.16	0.00	0.39	0.01	0.82
Income4	0.05	1.94	0.04	1.20	0.08	1.80
Income5	-0.00	-0.40	-0.00	-0.30	-0.01	-0.48
Working	0.02	1.66	0.03	1.64	0.03	1.39
Lifestyle habits						
Sedentary job	-0.00	-0.55	-0.01	-0.56	-0.01	-0.55
Weekly sport activities	-0.03	-1.52	-0.04	-1.53	-0.04	-1.52
Daily smoker	-0.01	-1.26	-0.01	-1.16	-0.03	-1.34
Weekly alcohol consumption	-0.01	-0.76	-0.02	-1.08	0.01	0.47
Meat	-0.01	-0.73	0.00	0.32	-0.00	-0.06
Fruits	-0.00	-0.26	-0.00	-0.55	-0.02	-1.01
Vegetables	0.00	0.63	0.00	0.71	0.01	0.76
Legumes	-0.00	-0.26	-0.00	-0.42	-0.01	-0.83
Unexplained difference	0.28	1.34	0.08	0.29	0.10	0.28
Sociodemographics						
Age: 18-35	0.11	1.26	0.16	1.54	0.11	0.77
Age:36-45	0.00	0.04	-0.08	-0.63	-0.17	-0.91
Age:46-55	-0.07	-0.77	-0.13	-1.10	-0.28	-1.76
Age: 55-65	-0.04	-0.46	0.02	0.13	0.26	1.44
Married	0.51	1.67	0.13	0.34	-0.06	-0.11
Kids	0.29	1.99	0.21	1.22	0.44	1.90
SES						
Schooling	0.44	0.70	0.29	0.37	0.15	0.14
Income1	-0.09	-0.86	-0.11	-0.81	-0.11	-0.55
Income2	0.14	1.38	0.20	1.50	0.23	1.34
Income3	-0.01	-0.06	0.04	0.36	0.11	0.67
Income4	-0.05	-0.70	-0.07	-0.77	-0.06	-0.54
Income5	0.01	0.30	0.00	0.01	-0.03	-0.43
Working	-0.64	-2.03	-0.66	-1.62	-0.95	-1.71
Lifestyle habits						
Sedentary job	0.17	1.21	0.17	0.90	0.01	0.03
Weekly sport activities	-0.02	-0.30	-0.04	-0.45	0.09	0.85
Daily smoker	-0.14	-0.97	-0.20	-1.06	-0.01	-0.05
Weekly alcohol consumption (index)	0.34	2.35	0.28	1.45	0.12	0.51
Meat	0.01	0.08	0.06	0.33	0.02	0.07
Fruits	0.16	0.51	0.28	0.68	-0.33	-0.56
Vegetables	0.42	1.61	0.56	1.66	0.99	2.14
Legumes	-0.03	-0.60	-0.00	-0.00	0.03	0.30
Constant	-1.21	-1.30	-1.05	-0.85	-0.44	-0.25

Table 4.11: Robustness Checks

	Complete Case		**(1)**		**(2)**		**(3)**		**(4)**		**(5)**	
Women												
Overall decomposition	mean	z-stat	mean	z-stat	mean	z-stat	mean	z-stat	mean	z-stat	mean	z-stat
South regions	25.29	203.18	25.27	218.77	25.27	218.77	26.26	205.15	25.28	183.53	25.32	226.85
North regions	24.74	198.12	24.69	227.96	24.69	227.96	25.7	256.98	24.74	198.12	24.05	227.48
Difference	**0.55**	**3.11**	**0.58**	**3.63**	**0.58**	**3.63**	**0.56**	**3.09**	**0.54**	**2.9**	**1.28**	**8.31**
Explained diff.	0.40	4.68	0.36	4.64	0.39	4.94	0.39	4.34	0.37	4.1	0.55	6.88
Unexplained diff.	0.15	0.87	0.21	1.36	0.19	1.21	0.17	0.98	0.17	0.94	0.72	4.69
Constant	0.66	0.74	0.67	0.89	0.59	0.75	0.7	0.78	0.97	1.03	2.27	2.85
No. of observations	2860		3534		3534		2860		2591		3580	
Men												
Overall decomposition	mean	z-stat	mean	z-stat	mean	z-stat	mean	z-stat	mean	z-stat	mean	z-stat
South regions	26.76	235.69	26.67	250.68	26.67	250.68	27.03	255.34	26.75	205.82	26.83	275.24
North regions	26.65	257.39	26.61	291.74	26.61	291.74	26.95	271.97	26.65	257.39	26.28	292.07
Difference	**0.11**	**0.72**	**0.06**	**0.43**	**0.06**	**0.43**	**0.07**	**0.51**	**0.1**	**0.63**	**0.54**	**4.09**
Explained diff.	-0.04	-0.58	-0.06	-0.97	-0.08	-1.26	-0.09	-1.25	-0.08	-1.01	0.23	3.38
Unexplained diff.	0.15	0.99	0.13	0.88	0.14	1.01	0.16	1.09	0.18	1.11	0.31	2.37
Constant	-0.94	-1.31	-1.61	-2.56	-1.3	-1.97	-0.98	-1.46	-0.62	0.82	0.28	0.44
No of observations	2749		3344		3344		2749		2465		3570	

Note: estimations obtained with the Complete Case approach are the baseline results discussed in Section 4.1.

Results in Columns (1) to (5) correspond in the five robustness checks discussed in section 4.2.

5 Conclusion

This dissertation deals with three different yet connected critical aspects of the analysis of obesity in Spain and is aimed to provide a holistic picture of the key determinants of this epidemic disease as well as to elicit valuable information with particular policy relevance on the specific matter.

As global data shows that SES-related inequality in obesity is still a reality, our investigation in the second chapter aims to observe the development of obesity and income-related inequality in obesity in Spain over the past years, as well as the basic factors responsible for this inequality. We evidence that overall, obesity is more concentrated among the worse-off in Spain, as the Concentration Index is always negative and statistically significant, a finding that calls for a need for policy and practice to focus on inequalities in obesity and develop interventions to reduce the gap between rich and poor individuals. In more detail, we conclude that even though our measures of obesity (prevalence, depth and severity) have an increasing trend, income-related inequalities in all three measures decrease. When we differentiate by gender, we notice that this result is mainly driven by women, while when decomposing these inequalities for the last available survey (i.e. the 2017 wave) we observe that the basic contributing factors to the total income-inequality in obesity status, depth and severity is basically income and education. When comparing our results of 2017 versus the ones of the 1997 wave, we conclude that inequalities where much larger for women than men in 1997, as well as that income-related inequalities in 1997 resulted from the confounding effects of demographics and SES, rather than mainly from SES attributes, which is the case for 2017. These more pronounced differences in inequalities for women between these two time periods, may be attributed to changes in the social patterning of obesity. That is, the gender difference could be an aftermath of the stronger emphasis on thinness in women in our contemporary society, where men value a larger and more muscular shape. In addition, the fact that the more severe cases of obesity are becoming more and more equally distributed according to income over the years could mirror deeper changes in the basic risk factors of obesity. Health outcomes such as obesity, result from the balancing of the forces that express a country s development stage and the conditions of its population. Individual s choices within this context could be the response to the so-called obesogenic environment.

With a view to account for the interplay between individual and regional attributes in shaping individual obesity risk, we proceed to the third chapter. Using a multilevel framework, we conclude that both group and individual effects play a key role in understanding BMI and obesity. Our ML estimates confirm the expected individual-level and regional effects on BMI and obesity. Furthermore, we provide evidence that our proxies of the social environment (criminality and green spaces) have a positive and statistically significant effect on female BMI and the prevalence of obesity. The findings of this chapter are providing important information to public health authorities, since we report that environmental and regional characteristics influence individual BMI and obesity. This means that local governments and local communities can play an important role in implementing specific policies, such as promoting environments that encourage and support healthy lifestyles.

This regional approach set up the main motivation of the fourth chapter. As important differences in terms of BMI between the North and the South of Spain are observed, this third chapter seeks to understand the main conditioning factors behind this North to South BMI divide in Spain. Starting with a simple OLS, our results show that the conditional correlation between observable determinants of BMI differs in the two groups of regions (North vs South) and by gender. We proceed with the Oaxaca-Blinder decomposition analysis that enables us to disentangle the contribution of each covariate and the corresponding coefficients to this difference. We reveal that the mean BMI gap between the South and North of Spain is mostly driven by differences between women residing in the two areas of the country as well as that a large and significant part of this regional average gap in BMI (73%) is due to differences in endowments related to SES status (basically years of education). Since the findings at the upper tail of the BMI distribution are the ones actually capturing overweight and obesity problems, we proceed with the distributional analysis and the corresponding decomposition and evidence that differences in SES endowments and particularly schooling explain a very significant part of the women s North to South differential. Overall, our results indicate that SES differentials (mainly educational attainment) between women residing in the North versus their counterparts living in the South of the country are producing remarkable differences in BMI across these regions, both at the mean and at the top of the BMI distribution. Results exhibit a very strong conditional correlation between education and BMI, being the endowment of the former variable responsible for a substantial share of the gap in BMI between women residing in different Spanish regions. Efforts aimed at improving (years of) schooling for women in the South would substantially mitigate differences in overweight and obesity between the two groups of regions, a policy intervention that would additionally reduce differences in obesity-related diseases and/or improve health in general, as obesity constitutes a key risk factor for many chronic conditions Therefore, the causal effect of education in mitigating regional disparities in BMI, overweight and obesity and even other health-related variables should be further investigated in future research when appropriate data for such purposes are available.

Bibliography

Allison, P. (2001). *Missing Data. Series: Quantitative Applications in the Social Sciences 136.* Sage Publications, Inc.

Anderson, N. B. (1995). Behavioral and sociocultural perspectives on ethnicity and health: Introduction to the special issue. *Health Psychology*, 14:589–591.

Antonovsky, A. (1967). Social class, life expectancy and overall mortality. *The Milbank Memorial Fund Quarterly*, 45:31–73.

Bailey, T. C. and Gatrell, A. C. (1995). Interactive spatial data analysis. *John Wiley and Sons, New York, NY*.

Ballas, D., Lupton, R., Dale, R., Kavroudakis, D., Hennig, B., Yiagopoulou, V., Dale, R., and Dorling, D. (2012). Mind the gap. education inequality across eu regions. *NESSE Network for Experts for the European Commission.*

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

Bilger, M., Kruger, E., and Finkelstein, E. (2017). Measuring socioeconomic inequality in obesity: Looking beyond the obesity threshold. *Health Economics*, 26(8):1052–1066.

Black, J. and Macinko, J. (2009). The changing distribution and determinants of obesity in the neighborhoods of new york city, 2003-2007. *American Journal of Epidemiology*, 171:765–775.

Blinder, A. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human Resources.*, 8(4):436–455.

Borg, S., Persson, U., Odegaard, K., Berglund, G., Nilsson, J., and Nilsson, P. (2005). Obesity, survival, and hospital costs-findings from a screening project in Sweden. *Value Health*, pages 562– 71.

Bostrom, G. and Diderichsen, F. (1997). Socioeconomic differentials in misclassification of height, weight and body mass index based on questionnaires data. *International Journal of Epidemiology*, 26(4):860–866.

Box, G. and Cox, D. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B*, (Methodological) 26(2):211–252.

Brion, M., Lawlor, D., Matijasevich, A., Horta, B., Anselmi, L., Araújo, C., Menezes, A., Victora, C., and Smith, G. (2011). What are the causal effects of breastfeeding on iq, obesity and blood pressure? evidence from comparing high-income with middle-income cohorts. *International Journal of Epidemiology*, 40:670–680.

Brunello, G., Fabbri, D., and Fort, M. (2013). The causal effect of education on body mass: Evidence from europe. *Journal of Labor Economics*, 31.

Caliendo, M. and Lee, W. (2013). Fat chance! Obesity and the transition from unemployment to employment. *Economics and Human Biology*, 11:121–133.

Carrieri, V. and Jones, A. (2016). The income-health relationship beyond the mean, new evidence from biomarkers. *Health Economics*.

Cawley, J. (2000). An instrumental variables approach to measuring the effect of body weight on employment disability. *Health Services Research*, 35 (5):1159–1179.

Cawley, J. (2004). The impact of obesity on wages. *Journal of Human Resources*, 39(2):451–474.

Cawley, J. and Burkhauser, R. (2008). Beyond bmi: the value of more accurate measures of fatness and obesity in social science research. *Journal of Health Economics*, 27 (2):519–529.

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

Cawley, J. H., Moran, J., and Simon, K. I. (2010). The impact of income on the weight of elderly americans.

Chou, S., Grossman, M., and Saffer, H. (2004). An economic analysis of adult obesity: results from the behavioral risk factor surveillance system. *Journal of Health Economics*, 23:565–587.

Clarke, P., Gerdtham, U.-G., Johannesson, M., Bingefors, K., and Smith, L. (2002). On the measurement of relative and absolute income-related health inequality. *Social Science and Medicine.*, 55:1923–1928.

Costa-Font, J., Fabbri, D., and Gil, J. (2009). Decomposing body mass index gaps between mediterranean countries: a counterfactual quantile regression analysis. *Economics and Human Biology.*, 7:351–365.

Costa-Font, J., Fabbri, D., and Gil, J. (2010a). Decomposing cross-country differences in levels of obesity and overweight: does the social environment matter? *Social Science and Medicine*, 70:1185–1193.

Costa-Font, J., Fabbri, D., and Gil, J. (2010b). Decomposing cross-country differences in levels of obesity and overweight: does the social environment matter? *Social Science and Medicine*, 70:1185–1193.

Costa-Font, J. and Gil, J. (2008). What lies behind socio-economic inequalities in obesity in Spain: a decomposition approach. *Food Policy*, 33:61–73.

Costa-Font, J., Hernandez-Quevedo, C., and Jimenez-Rubio, D. (2014). Income inequalities in unhealthy life styles in England and Spain. *Economics and Human Biology*, 13:66–75.

Costa-Font, J. and Pons-Novell, J. (2007). Public health expenditures and spatial interactions in a decentralized national health system. *Health Economics*, 16:291–306.

Courtemanche, C., Pinkston, J., and Stewart, J. (2015). Adjusting body mass index for measurement error with invalid validation data. *Economics and Human Biology*, 19:275–293.

Cubbin, C., Sundquist, K., Ahlen, H., Johansson, S., Winkleby, M., and Sundquist, J. (2006). Neighborhood deprivation and cardiovascular disease risk factors: protective and harmful effects. *Public Health*, 34:228–37.

Cutler, D. M., Huang, W., and Lleras-Muney, A. (2015). When does education matter? the protective effect of education for cohorts graduating in bad times. *Social Science and Medicine*, 127(C):63–73.

Cutler, D. M., Lleras-Muney, A., and Vogl, T. (2012). Socioeconomic status and health: Dimensions and mechanisms. *The Oxford Handbook of Health Economics*.

Dadvand, P., Villanueva, C. M., Font-Ribera, L., Martinez, D., Basagana, X., Belmonte, J., Vrijheid, M., Grazuleviciene, R., Kogevinas, M., and Nieuwenhuijsen, M. J. (2014). Risks and benefits of green spaces for children: A cross-sectional study of associations with sedentary behavior, obesity, asthma and allergy. *Environmental Health Perspectives.*, 122:12.

Dare, S., Mackay, D., and Pell, J. (2015). Relationship between smoking and obesity: A cross-sectional study of 499,504 middle-aged adults in the UK general population. *Plos One*.

Davillas, A. and Benzeval, M. (2016). Alternative measures to bmi: exploring incomerelated inequalities in adiposity in great britain. *Social Science & Medicine*, 166:223–232.

Davillas, A., Jones, A. M., and Benzeval, M. (2017). The income-health gradient: evidence from self-reported health and biomarkers using longitudinal data on income. *In Panel Data Econometrics Volume II: Empirical Applications, Elsevier, In press*.

Devaux, M. and Sassi, F. (2013). Social inequalities in obesity and overweight in 11 OECD countries. *European Journal Of Public Health*, 23:464–469.

Devaux, M., Sassi, F., Church, J., Cecchini, M., and Borgonovi, F. (2011). Exploring the relationship between education and obesity. *OECD Journal: Economic Studies.*, 1:1–40.

DiNardo, J., Fortin, N., and Lemieux, T. (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica.*, 64:1001–1044.

Drewnowski, A. (2003). Fat and sugar: an economic analysis. *Journal of Nutrition*, 133:838S–840S.

Drewnowski, A. and Specter, S. (2004). Poverty and obesity: the role of energy density and energy costs. *The American Journal of Clinical Nutrition*, 79:6–16.

Duan, N., Manning, W., Morris, C., and Newhouse, J. (1983). A comparison of alternative models for the demand for medical care. *Journal of Business and Economic Statistics.*, 1(2):115–126.

Dutton, D. and McLaren, L. (2016). How important are determinants of obesity measured at the individual level for explaining geographic variation in body mass index distributions? observational evidence from canada using quantile regression and blinder-oaxaca decomposition. *Journal of Epidemiology and Community Health*, 70:367–373.

Dutton, D. J. and McLaren, L. (2011). Explained and unexplained regional variation in canadian obesity prevalence. *Obesity*, 19(7):1460–1468.

Eberth, B. and Gerdtham, U. (2008). What socioeconomic factors explain inequalities in obesity in England and Scotland? Scottish Economic Society Conference.

Egger, G. and Swinburn, B. (1997). An ecological approach to the obesity pandemic. *BMJ*, 315:477–80.

Eid, J., Overman, H., Puga, D., and Turner, M. (2008). Fat city: The relationship between urban sprawl and obesity. *Journal of Urban Economics*, 63(2):385–404.

Ellis, A. and Fry, R. (2010). Regional health inequalities in england. *Regional Trends.*, 42:60–79.

Ergin, I., Hassoy, H., and A., K. (2012). Socio-economic inequalities in overweight among adults in Turkey: a regional evaluation. *Public Health Nutrition*, 15(1):58–66.

Ergin, I. and Kunst, A. (2015). Regional inequalities in self-rated health and disability in younger and older generations in turkey: the contribution of wealth and education. *BMC Public Health*, 15:987.

Erreygers, G. (2009a). Correcting the concentration index. *Journal of Health Economics*, 28:504–515.

Erreygers, G., Clarke, P., and Van Ourti, T. (2012). Mirror, mirror, on the wall, who in this land is fairest of all? Distributional sensitivity in the measurement of socioeconomic inequality of health. *Journal of Health Economics*, 31(1):257–70.

Erreygers, G. and Van Ourti, T. (2011). Measuring socioeconomic inequality in health, health care and health financing by means of rank-dependent indices: a recipe for good practice. *Journal of Health Economics*, 30 (4):685–694.

Ferretti, F. and Mariani, M. (2017). Gender discrimination, gender disparities in obesity and human development. *Heliyon*, 14;3(3).

Finkelstein, E., Fiebelkorn, I., and Wang, G. (2004). State level estimates of annual medical expenditures attributable to obesity. *Obesity Research*, 12:18–24.

Finkelstein, E., Fiebelkorn, I., and Wang, G. (2005). The costs of obesity among full-time employees. *American Journal of Health Promotion.*, 20:45–51.

Finkelstein, E., Trogdon, J., Cohen, J., and Dietz, W. (2009). Annual medical spending attributable to obesity: payer-and service specific estimates. *Health Affairs*, 28(5):w822–w831.

Firpo, S., Fortin, N. M., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica.*, 77(3):953–973.

Flowerdew, R., Manley, D., and Sabel, C. (2008). Neighborhood Effects on health: does it matter where you draw the boundaries? *Social Science and Medicine*, 66:1241–1255.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition methods in economics. handbook of labor economics. 4A:1–102. Chapter 1.

Fortin, N. M. (2008). The gender wage gap among young adults in the united states. the importance of money versus people. *Journal of Human Resources*, 43(4).

Foster, J., Greer, J., and Thorbecke, E. (1984). A class of decomposable poverty measures. *Econometrica*, 52:761–776.

Frankenberg, E., Ho, J. Y., and Thomas, D. (2016). Biological health risks and economic development. *The Oxford Handbook of Economics and Human Biology.*, page 454.

Freedman, D., Khan, L., Serdula, M., Galuska, D., and Dietz, W. (2002). Trends and correlates of class 3 obesity in the United States from 1990 through 2000. *Journal of the American Medical Association*, 288:1758–1761.

Freeman, R. (1980). Unionism and the dispersion of wages. *Industrial and Labor Relations Review.*, 34:3–23.

Freeman, R. B. (1984). Longitudinal analysis of the effect of trade unions. *Journal of Labor Economics.*, 2:1–26.

Garcia-Villar, J. and Quintana-Domeque, C. (2009). Income and body mass index in Europe. *Economics and Human Biology*, 7:73–83.

Gil, J. and Mora, T. (2011). The determinants of misreporting weight and height: The role of social norms. *Economics and Human Biology*, 9:78–91.

Goettler, A., Grosse, A., and Sonntag, D. (2017). Productivity loss due to overweight and obesity: a systematic review of indirect costs. *BMJ Open.*, 7(10).

Gutierrez-Fisac, G., Rodriguez Artalejo, F., Guallar-Castillon, P., Banegas Banegas, J., and del Rey Calero, J. (1999). Determinants of geographical variations in body mass index (bmi) and obesity in Spain. *International Journal of Obesity and related metabolic disorders.*, 23(4):342–7.

Hacking, J., Muller, S., and Buchan, J. (2011). Trends in mortality from 1965 to 2008 across the english north-south divide: comparative observational study,. *BMJ*, pages 342–508.

Hajizadeh, M., Campbell, M., and Sarma, S. (2014). Socioeconomic inequalities in adult obesity risk in canada: trends and decomposition analyses. *The European Journal of Health Economics.*, 15:203–221.

Heger, D. (2016). Decomposing differences in health and inequality using quasiobjective health indices. *RUHR Economic Papers No. 607*.

Humphries, K. and van Doorslaer, E. (2000). Income-related health inequality in canada. *Social Science and Medicine*, 50:663–71.

Jann, B. (2008). The blinder-oaxaca decomposition for linear regression models. *The Stata Journal.*, 4:453–479.

Janssen, I., Katzmarzyk, P., and Ross, R. (2004). Waist circumference and not body mass index explains obesity-related health risk. *The American Journal of Clinical Nutrition.*, 79(3):379–384.

Johnston, D., Propper, C., and Shields, M. (2009). Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. *Journal of Health Economics.*, 28(3):540–552.

Jolliffe, D. (2004). Continuous and robust measures of the overweight epidemic from 1971-2000. *Demography*, 41(2):303–314.

Jolliffe, D. (2011). Overweight and poor? On the relationship between income and the Body Mass Index. *Economics and Human Biology.*, 9 (2011):342–355.

Juhn, C., Murphy, K., and Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 101:410–441.

Kakwani, N. (1977). Measurement of tax progressivity. An international comparison. *Economic Journal.*, 87(345):71–80.

Kakwani, N., Wagstaff, A., and van Doorslaer, E. (1997). Socioeconomic inequalities in health: measurement, computation and statistical inference. *Journal of Econometrics.*, 77(1):87–103.

Kim, Y. (2016). The long-run effect of education on obesity in the us. *Economics and Human Biology*, 21:100–109.

Kinge, J. (2016). Body mass index and employment status: a new look. *Economics and Human Biology*, 22::117–125.

Kosteniuk, J. and Dickinson, H. (2003). Tracing the social gradient in the health of canadians: primary and secondary determinants. *Social Science and Medicine*, 57:263–76.

Kreft, I. and De Leeuw, J. (1998). *Introducing Multilevel Modeling*.

Kuczmarski, M., Kuczmarski, R., and Najjar, M. (2001). Effects of age on validity of self-reported height, weight and body mass index: findings from the third nhanes 1988-1994. *Journal of the American Dietetic Association*, 101(1):28–34.

Lakdawalla, D. and Philipson, T. (2009). The growth of obesity and technological change: a theoretical and empirical investigation. *Economics and Human Biology*, 7(3):283–293.

Leyland, A. and Groenewegen, P. (2003). Multilevel modeling and public health policy. *Scandinavian Journal of Public Health*, 31:267–274.

Lhila, A. and Long, S. (2012). What is driving the black-white difference in low birth weight in the us? *Health Economics.*, 21:301–315.

Lindeboom, M., Lundborg, P., and van der Klaauw, B. (2010). Assessing the impact of obesity on labor market outcomes. *Economics and Human Biology*, 8(3):309–319.

Link, B., Northridge, M., Phelan, J., and Ganz, M. (1998). Social epidemiology and the fundamental cause concept: On the structuring of effective cancer screens by socioeconomic status. *The Milbank Quarterly*, 76:375–402.

Link, B. and Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior*, pages 80–94.

Ljungvall, A. and Gerdtham, U. (2010). More equal but heavier: a longitudinal analysis of income-related obesity inequalities in an adult swedish cohort. *Social Science & Medicine.*, 70(2):221–231.

Machado, J. and Mata, J. (2005). Counterfactual decomposition of changes in wwage deistributions using quantile regression. *Journal of Applied Econometrics*, 20 (4):445–465.

Madden, D. (2006). Body Mass Index and the measurement of Obesity. *Health, Econometrics and Data Group (HEDG) Working Papers 06/11, Department of Economics, University of York.*

Malnick, S. and Knobler, H. (2006). The medical complications of obesity. *Quarterly Journal of Medicine: An International Journal of Medicine*, 99(9):565–579.

Manning, W. and Mullahy, J. (2001). Estimating log models: to transform or not to transform? *Journal of Health Economics*, 20(4):461–494.

Mark Austin, A. F. and Spine, M. (2002). The effects of neighborhood conditions on perceptions of safety. *Journal Of Criminal Justice*, 30:417–427.

Markwick, A., Vaughan, L., and Ansari, Z. (2013). Opposing socioeconomic gradients in overweight and obese adults. *Australian and New Zealand Journal of Public Health*, 37(1):32–8.

Martin, A., Ogilvie, D., and Suhrcke, M. (2014). Evaluating causal relationships between urban built environment characteristics and obesity: a methodological review of observational studies. *International Journal of Behavioral Nutrition and Physical Activity*, 11:142.

McLaren, L. (2007). Socioeconomic status and obesity. *Epidemiologic Reviews*, 29 (1):29–48.

Merino Mantosa, M. and Urbanos-Garrido, R. (2016). Disentangling effects of socioeconomic status on obesity : A cross-sectional sstudy of the spanish adult population. *Economics and Human Biology*, 22:216–224.

Merino Ventosa, M. and Urbanos-Garrido, R. (2016). Disentangling effects of socioeconomic status on obesity: A cross-sectional study of the Spanish adult population. *Economics and Human Biology*, 22:216–224.

Monteiro, C., Benicio, M., Conde, W., and Popkin, B. (2000). Shifting obesity trends in brazil. *European Journal of Clinical Nutrition*, 54:342–346.

Monteiro, C., Moura, E., Conde, W., and Popkin, B. (2004). Socioeconomic status and obesity in adult populations of developing countries: a review. *Bulletin of the World Health Organization*, 82:940–946.

Moore, C. and Cunningham, S. (2012). Social position, psychological stress, and obesity: A systematic review. *Journal of the Academy of nutricion and dietics*, 112:518–526.

Mora, T., Gil, J., and Sicras-Mainar, A. (2015). The influence of obesity and overweight on medical costs: a panel data perspective. *European Journal Of Health Economics*, 16(2):161–73.

Morris, S. (2006). Body mass index and occupational attainment. *Journal of Health Economics*, 25(2):347–364.

Muller-Riemenschneider, F., Reinhold, T., Berghofer, A., and Willich, S. (2008). Health-economic burden of obesity in Europe. *European Journal Of Epidemiology.*, 23:499–509.

Nakamura, K., Okamura, T., Kanda, H., Hayakawa, T., Okayama, A., and Ueshima, H. (2007). Health promotion research committee of the Shiga National Health Insurance Organizations. Medical costs of obese Japanese: a 10-year follow-up study of National Health Insurance in Shiga, Japan. *European Journal Of Public Health.*, 17(5):424–429.

Nayga, R. M. (1999). Sociodemographic factors associated with obesity in the USA. *Journal of Consumer Studies and Home Economics.*, 23:161–164.

Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society.*, Series A (General) 135(3):370–384.

Nikolaou, A. and Nikolaou, D. (2008). Income-related inequality in the distribution of obesity among europeans. *Journal of Public Health.*, 16(6):403–411.

Oaxaca, R. (1973). Male-female wage differentials in urban labour markets. *International Economic Review.*, 14(3):693–709.

O'Donnell, O., O'Neill, S., Van Ourti, T., and Walsh, B. (2016). conindex: Estimation of concentration indices. *The Stata Journal*, 16(1):112–138.

O'Donnell, O., van Doorslaer, E., Wagstaff, A., and Lindelow, M. (2008). Analyzing health equity using household survey data: A guide to techniques and their implementation. *World Bank Publications, Washington.*, 434.

OECD (2014). Obesity and the economics of prevention: Fit not fat. *Key Facts -Spain, Update 2014*.

OECD (2015). Health at a glance.

OECD (2017). Obesity update. *OECD*.

O'Neill, J. E. and O'Neill, M. (2006). What do wage differentials tell about labor market discrimination? *The Economics of Immigration and Social Diversity, Emerald Group Publishing Limited.*, 24.

Park, R. (1966). Estimation with heteroscedastic error terms. *Econometrica (pre-1986)*, 34(4):888.

Plantinga, A. J. and Bernell, S. (2015). Can urban planning reduce obesity? the role of selfselection in explaining the link between weight and urban sprawl. *Review of Agricultural Economics*, 29:557–563.

Pohlmeier, W. and Ulrich, V. (1995). An econometric model of the two-part decisionmaking process in the demand for health care. *The Journal of Human Resources*, 30:339–361.

Powdthavee, N. (2010). Does education reduce the risk of hypertension? Estimating the biomarker effect of compulsory schooling in England. *Journal of Human Capital.*, 4(2):173–202.

ProspectiveStudiesCollaboration (2009). Body mass index and cause-specific mortality in 900.000 adults: collaborative analyses of 57 prospective studies. *The Lancet*, 373(9669):1083–1096.

Raebel, M., Malone, D., Conner, D., Xu, S., Porter, J., and Lanty, F. (2004). Health services use and health care costs of obese and non-obese individuals. *Archives of internal medicine*, 164:2135–2140.

Raftopoulou, A. (2017). Geographic determinants of individual obesity risk in spain: A multilevel approach. *Economics and Human Biology*, 24:185–193.

Reimers, C. W. (1983). Labor market discrimination against hispanic and black men. *Review of Economics and Statistics.*, 65:570–579.

Rodriguez-Caro, A., Vallejo-Torres, L., and Lopez-Valcarcel, B. (2016). Unconditional quantile regressions to determine the social gradient of obesity in Spain 1993-2014. *International Journal for Equity in Health.*, 15:175.

Rose, V. (1998). National heart, lung, and blood institute releases new guidelines for the treatment of hypertension. *American Family Physician.*, 57(2):362–4, 366.

Ross, C. and ling Wu, C. (1995). The links between education and health. *American Sociological Association*, 60:719–745.

Roux, A. D. (2000). Multilevel analysis in public health research. *Public Health*, 21:171–192.

Safaei, J. (2014). Socioeconomic health disparities in canadian regions. *Open Medicine Journal*, 1:90–95.

Sampson, R. J., Morenoff, J. D., and Gannon-Rowley, T. (2002). Assessing neighborhood effects: Social processes and new directions in research.'. *Annual Review of Sociology*, 28:443–478.

Sander, B. and Bergemann, R. (2003). Economic burden of obesity and its complications in Germany. *European Journal of Health Economics*, 4:248–253.

Sassi, F., Devaux, M., Cecchini, M., and Rusticelli, E. (2009). The obesity epidemic: Analysis of past and projected future trends in selected OECD countries. *OECD Health Working Papers*, (45).

Sen, B. (2014). Using the oaxaca-blinder decomposition as an empirical tool to analyze racial disparities in obesity. *Obesity*, 22(7):1750–1755.

Sobal, J. (1991). Obesity and socioeconomic status: a framework for examining relationships between physical and social variables. *Medical Anthropology*, 13:231–248.

Sobal, J. and Stunkard, A. (1989). Socioeconomic status and obesity: A review of the literature. *Psychological Bulletin*, 105:260–275.

Specchia, M., Veneziano, M., Cadeddu, C., Ferriero, A., Mancuso, A., Ianuale, C., Parente, P., Capri, S., and Ricciardi, W. (2015). Economic impact of adult obesity on health systems: A systematic review. *European Journal of Public Health.*, 25:255–262.

Subramanian, S. V. and Kawachi, I. (2004). Income inequality and health: What have we learned so far? *Epidemiologic Review*, 26.

Teachman, J. and Crowder, K. (2002). Multilevel models in family research: Some conceptual and methodological issues. *Journal of Marriage and Family*, 64:280–294.

Tremmel, M., Gerdtham, U., Nilsson, P., and Saha, S. (2017). Economic burden of obesity: a systematic literature review. *International Journal of Environmental Research and Public Health*, 14(4):435.

Valdes, S. e. a. (2014). Prevalence of obesity, diabetes and other cardiovascular risk factors in andalusia (southern spain). comparison with national prevalence data. *Revista Espanola de Cardiologia (English Edition)*, 67:442–448.

van Baal, P., Polder, J., de Wit G.A., Hoogenveen, R., and Feenstra, T. (2008). Lifetime medical costs of obesity: prevention no cure for increasing health expenditure. *PloS Med*, 5(2):e29.

van Doorslaer, E. and Koolman, X. (2004). Explaining the differences in income-related health inequalities across european countries. *Health Economics.*, 13:609–628.

Vazquez-Sanchez, R. and Lopez-Alemany, J. (2002). Los costes de la obesidad alcanzan el 7 porcentaje del gasto sanitario. *Revista Espanola de Economia de la Salud.*, 1(3).

Von Lengerke, T., Reitmeier, P., and John, J. (2006). Direct medical costs of (severe) obesity: a bottom-up assessment of over vs. normal weight adults in the KORA-study region (Augsburg, Germany). *Gesundheitswesen*, 68:110–115.

Wagstaff, A. (2005). The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. *Health Economics*, 14:429–432.

Wagstaff, A. and van Doorslaer, E. (2000). Income inequality and health: what does the literature tell us? *Annual Review of Public Health*, 21:543–567.

Wagstaff, A., van Doorslaer, E., and Watanabe, N. (2003). On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. *Journal of Econometrics.*, 112(1):207–223.

Wagstaff, W., Paci, P., and van Doorslaer, E. (1991). On the measurement of inequalities in health. *Social Science and Medicine.*, 33:545–557.

Wang, Y. and Beydoun, M. (2007). The obesity epidemic in the united states: gender, age, socioeconomic, racial/ethnic, and geographic characteristics: a systematic review and meta-regression analysis. *Epidemiology Reviews*, 29:6–28.

WHO (1995). Global database on body mass index.

WHO (2015). Obesity and overweight.

WHO (2017). Global action plan for the prevention and control of noncommunicable diseases 2013-2020. *Global Action Plan*.

Willett, W., Dietz, W., and Colditz, G. (1999). Guidelines for healthy weight. *New England Journal of Medicine.*, 341:427–434.

Wolf, A. and Colditz, G. (1998). Current estimates of the economics costs of obesity in the United States. *Obesity Research.*, 6:97–106.

Wolfenstetter, S. (2012). Future direct and indirect costs of obesity and the influence of gaining weight: Results from the MONICA/KORA cohort studies, 1995-2005. *Economics and Human Biology.*, 10:127–138.

Zabarsky, G., Beek, C., Hagman, E., Pierpont, B., Caprio, S., and Weiss, R. (2018). Impact of severe obesity on cardiovascular risk factors in youth. *The Journal of Pediatrics.*, 192:105–114.

Zhang, Q. and Wang, Y. (2004). Socioeconomic inequality of obesity in the United States: do gender, age, and ethnicity matter? *Social Science and Medicine.*, 58(6):1171–1180.

Zhao, Z., Kaestner, R., and Xu, X. (2014). Spatial mobility and environmental effects on obesity. *Economics and Human Biology*, 14:128–140.