

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

Antonio Di Paolo

Universitat de Barcelona

Joan Gil (✉ joangil@ub.edu)

Universitat de Barcelona <https://orcid.org/0000-0003-4159-379X>

Athina Raftopoulou

EURAC Centro di Biomedicina

Research

Keywords: Regional BMI Differentials, Obesity, Decomposition analysis, Spain

Posted Date: March 23rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-18335/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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

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 that have been experiencing high trends in the prevalence of overweight and obesity during the last decades. As a response to these alarming trends, the Spanish Ministry of Health developed a national strategy for the prevention and control of obesity in 2005 called the NAOS strategy.¹ The strategy included informative campaigns, agreements with public and private institutions, voluntary working agreements, educational programs and supporting health promotion initiatives (Ballesteros-Arribas et al., 2007). However, due to its nature as a (mainly) school based initiative, sectors including the built environment, transportation and the food environment received less attention (Franco et. al., 2010). This could explain the fact that according to the OECD (2014) adult obesity rates in Spain were higher than the OECD average, while based on the latest data, roughly 18% of adults are obese and 55% are overweight (including obese) in Spain (INE, 2019), doubling the rates shown at the end of the 90s.

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

¹ Among other initiatives, the NAOS strategy stood out due to its aim to stress the importance of adopting a lifelong perspective in the prevention and control of obesity (by fostering healthy eating and positively promoting taking up healthy habits) and because of its setting, that consists of a set of actions directed at the entire population.

than others (Gutiérrez-Fisac et al., 1999; Valdes et al., 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 (Global Action Plan, 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. Indeed, highlighting the factors that contribute to differences in BMI across regions could be helpful for a better understanding of the mechanisms behind the regional gaps in productivity, wages and overall economic performance that are observed in Spain as well as in other countries (OECD, 2018).

In this paper, we decompose regional differentials in BMI between northern and southern Spanish regions.² 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

² As long as the definition of northern and southern regions is inevitably subject to some degree of arbitrariness, we analyse the sensitivity of the results obtained under different grouping of regions in the robustness checks section.

effective health and food policies to combat the obesity epidemic. 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). 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)³ 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 (OECD, 2010).

The decomposition approach has been widely used to analyse regional wage gaps. Pereira and Galego (2011) investigated regional differences in wages for Portugal using both the OB decomposition and the Juhn, Murphy and Pierce (JMP) decomposition, which allows for the analysis along the outcome's distribution (Juhn et al., 1993). In subsequent papers, the same authors applied the Machado and Mata (2005, MM) decomposition method (Pereira and Galego, 2014) and the RIF-regression decomposition to analyse the drivers of regional differentials along the conditional and the unconditional distribution (Galego and Pereira, 2014) respectively. For the Spanish case, Motellón et al. (2011) and López-Bazo and Motellón (2012) implemented, respectively, the JMP and the OB approaches to study distributional and average differences in hourly wages across Spanish regions. More recently, Murillo-Huertas et al. (2020) also analysed the regional wage gap in Spain, but implementing the more powerful decomposition technique based on RIF-regressions that provides evidence along the unconditional distribution of wages and enables a detailed decomposition (besides other desirable properties explained in section 3). Finally, the RIF-regression decomposition has been also used by Herrera-Idárraga et al. (2016), to investigate the role of informality in explaining the regional wage disparities along the unconditional distribution for the case of Colombia.

The application of decomposition techniques to explain differences in BMI is more limited. Indeed, only a few papers applied decomposition methods to investigate

³ The WHO defines obesity as a BMI ≥ 30 kg/m², while severe obesity corresponds to BMI ≥ 40 kg/m².

geographical differences in BMI.⁴ Specifically, Costa-Font et al. (2009) applied the MM decomposition to gauge the drivers of cross-country disparities in BMI at different quantiles among Mediterranean countries, while Costa-Font et al. (2010) implemented an extension of the OB decomposition to a non-linear model to analyse differences in overweight and obesity between Italy and Spain. More recently, Dutton and McLaren (2016) used the RIF-regression decomposition to examine the importance of individual-level characteristics for explaining geographic variation in BMI distributions for the case of Canada. However, they only focused on aggregated differences, without disentangling the specific contribution of each regressor. With this paper, we are the first in applying OB and RIF-regression decomposition to analyse average and distributional regional differences in BMI for a European country. Moreover, we also add to the literature by decomposing regional differences in BMI along its unconditional distribution into the contribution of endowments and returns to each observable characteristics separately, through the detailed decomposition. This places our study a step ahead with respect to previous evidence, since our analysis enables highlighting the relevance of specific key factors for the design of interventions targeting overweight and obese individuals (such as age, education or lifestyle and food habits) in accounting for the regional disparities in BMI. Nevertheless, the reader must bear in mind that the evidence obtained from decomposition techniques is useful to understand the contribution of endowments and returns of each specific factor in “accounting” for the differences in BMI across regions, but should not be directly interpreted in terms of the underlying causal mechanisms (Fortin et al., 2011). However, the results from decomposition analysis are still relevant to draw policy implications, especially when the analysis highlights the prevalent relevance of some specific factor. This would motivate a deeper analysis of the impact of that factor using the appropriate techniques to identify the corresponding causal relationship.

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 (0.546 points, z-stat 2.96 for females relative to 0.157 points, z-stat 0.99 for males). Around 74% of the cross-regional gap in BMI among women is accounted by differences in observable characteristics.

⁴ It seems also worth mentioning that Sen (2014) applied the OB decomposition to investigate ethnic disparities in BMI for the case of US.

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. Overall, differences in the endowment of human capital represent the main factor behind the north-south gap in BMI and obesity in Spain.

2. 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 survey⁵ with valid information on the relevant variables.⁶ We also discard observations from the Balearic and Canary Islands, as well as the autonomous cities of Ceuta and Melilla.⁷

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., Boström and Diderichsen, 1997; Kuczmarski et al., 2001; Gil and Mora, 2011).

⁵ 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).

⁶ The exception here is the variable family income, which is missing for a non-trivial proportion of the sample (roughly 20%). We examine the results obtained using imputation methods in the robustness checks section.

⁷ Ceuta and Melilla were excluded because these regions represent Spanish enclaves located in northern Africa. We also removed observations from the Balearic and Canary Islands, since individuals residing in these Spanish islands might differ with respect to mainland inhabitants along many (unobservable) dimensions that are likely to be correlated with health outcomes.

Mostly following the existing literature on BMI, we divide the conditioning factors into three main groups, namely 1) sociodemographic variables, 2) socioeconomic status (SES), and 3) lifestyle variables (see Appendix: Table A). 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.⁸

Figure 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 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 0.36 units in mean BMI between the South (26.04 kg/m²) and the North (25.68 kg/m²).

[Figure 1 and Table 1 around here]

2.1 Descriptive statistics

⁸ Specifically, we consider 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.

Tables 2 and 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 2, we document a large and significant difference in mean weight level of around 1.61 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/m² (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 (11.99 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 36%) and less fruit. Differences in the consumption of vegetables and legumes among women are not statistically significant between the two groups of regions. Table 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.

[Tables 2 and 3 around here]

3. Empirical Methodology

3.1 Average BMI differentials between groups of regions.

The starting point of our empirical methodology is a simple OLS regression⁹ that 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,

$$bmi_i^S = \alpha^S + \beta^S X_i + u_i^S \quad (1a)$$

$$bmi_i^N = \alpha^N + \beta^N X_i + u_i^N \quad (1b)$$

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:

$$bmi_i = \alpha + \beta' X_i + \gamma_N I(N = 1) + \gamma_S I(S = 1) + u_i \quad (2)$$

$$\text{subject to } \gamma_S + \gamma_N = 0$$

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

⁹ OLS results are reported in Table B in the Appendix and are not discussed for brevity reasons. The computation of the Variance Inflation Factor for the estimated models indicates no multicollinearity problems.

$$\begin{aligned}
\overline{bmi}^S - \overline{bmi}^N &= (\overline{X}^S - \overline{X}^N)\hat{\beta} + (\hat{\gamma}_S - \hat{\gamma}_N) = \\
&= (\overline{X}^S - \overline{X}^N)\hat{\beta} + \underbrace{[(\overline{X}^S(\hat{\beta}_S - \hat{\beta}) + (\hat{\alpha}_S - \hat{\alpha}))]}_{\hat{\gamma}_S} - \underbrace{[(\overline{X}^N(\hat{\beta}_N - \hat{\beta}) + (\hat{\alpha}_N - \hat{\alpha}))]}_{\hat{\gamma}_N}
\end{aligned} \tag{3}$$

The term $(\overline{X}^S - \overline{X}^N)\hat{\beta}$ represents the composition effect (i.e. share of average BMI gap due to differences in observable characteristics), whereas the term $(\hat{\gamma}_S - \hat{\gamma}_N) = (\overline{X}^S(\hat{\beta}_S - \hat{\beta}) + (\hat{\alpha}_S - \hat{\alpha})) - (\overline{X}^N(\hat{\beta}_N - \hat{\beta}) + (\hat{\alpha}_N - \hat{\alpha}))$ 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).¹⁰

3.2 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 (especially 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). 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 τ th quantile (q_τ) of BMI corresponds to,

$$RIF(bmi_i, q_\tau) = q_\tau + IF(bmi_i, q_\tau) = q_\tau + \frac{\tau - I(bmi_i \leq q_\tau)}{f_{bmi_i}(q_\tau)} \tag{4}$$

¹⁰ 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).

where $f_{bmi_i}(q_\tau)$ is the unconditional density of BMI evaluated at the τ th quantile and $I(\cdot)$ an indicator function. By replacing the unknown elements of equation (4) by their sample estimators it is possible to obtain an estimate of the RIF, which is,

$$\widehat{RIF}(bmi_i, q_\tau) = \hat{q}_\tau + \widehat{IF}(bmi_i, q_\tau) = \hat{q}_\tau + \frac{\tau - I(bmi_i \leq \hat{q}_\tau)}{\hat{f}_{bmi_i}(\hat{q}_\tau)} \quad (5)$$

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}_τ) 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, it is 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 described by equation (3) to the RIF, that is:

$$\hat{q}_{S\tau} - \hat{q}_{N\tau} = \overline{RIF}(bmi^S, \hat{q}_\tau) - \overline{RIF}(bmi^N, \hat{q}_\tau) = (\overline{X^S} - \overline{X^N})\hat{\beta}_\tau + \left[\left(\overline{X^S}(\hat{\beta}_{S\tau} - \hat{\beta}_\tau) + (\hat{\alpha}_{S\tau} - \hat{\alpha}_\tau) \right) - \left(\overline{X^N}(\hat{\beta}_{N\tau} - \hat{\beta}_\tau) + (\hat{\alpha}_{N\tau} - \hat{\alpha}_\tau) \right) \right] \quad (6)$$

Here $\hat{\beta}_\tau$ corresponds to the non-discriminatory BMI structure (estimated from a pooled RIF regression) at quantile τ estimated in a similar fashion as equation (2). Similar to equation (3), the term $(\overline{X^S} - \overline{X^N})\hat{\beta}_\tau$ represents the composition effect and the term

$\left(\overline{X^S}(\hat{\beta}_{S\tau} - \hat{\beta}_\tau) + (\hat{\alpha}_{S\tau} - \hat{\alpha}_\tau)\right) - \left(\overline{X^N}(\hat{\beta}_{N\tau} - \hat{\beta}_\tau) + (\hat{\alpha}_{N\tau} - \hat{\alpha}_\tau)\right)$ captures the unexplained component of BMI differential evaluated at the τ -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. Results

4.1 OB decomposition results

Tables 4 and 5 present the aggregated and detailed OB decomposition results respectively, differentiating by gender. The decomposition analysis shown in Table 4 evidences that up to 74% (0.402 BMI units) of the overall South to North mean BMI gap for women (0.546 BMI units) is due to differences in endowments, whereas the remaining 26% (0.144 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 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 5, as a whole, the unexplained part or returns to certain characteristics, which accounts for 26% 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 (0.157 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).

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–South gap for males and females over the entire unconditional distribution of BMI.

[Tables 4 and 5 around here]

4.2 RIF decomposition results

Figure 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 6 and Table C of the Appendix¹¹ respectively). As shown in Table 6, 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 tends to be compensated with the decreasing contribution of unexplained factors. Interestingly, 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 living in the North. While women in the North are worse off with

¹¹ Table C 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.

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 decomposition 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 2 and Table 6 around here]

4.3 Robustness checks

In this section we consider the robustness of the previous findings with respect to three main issues that might affect our estimations, namely 1) the presence of missing values in the family income variable, 2) the potential bias in BMI due to the self-reported nature of the variables height and weight in the *EHIS* survey and 3) 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).¹² As it can be appreciated in columns (1) and (2) of Table E in the Appendix, the overall results from the OB decomposition is mostly unaffected by

¹² We also ran the OLS regressions for BMI (by groups of regions and gender) after replacing missing values of family income using Multiple Imputation techniques based on either ordinal models or interval regression (given the nature of the family income variable). The results, available upon request, are virtually the same than those shown in Table B of the Appendix, in terms of both the magnitude and significance of the dummies for family income.

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 E 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. Once again, the overall results from the OB decomposition are unaffected by the alternative grouping of regions. Overall, the sensitivity checks presented in this section points to the robustness of our results.

5. 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 average gap in BMI between the South and North of Spain is mostly driven by differences in characteristics between women residing in the two areas of the country. A large and significant part of this regional average gap in BMI (74%) 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. Indeed, in view of the epidemic of obesity as a global public health concern, policy-makers 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, in detriment of women residing in the South, explain again a very significant part of the North to South differential at the top of the BMI distribution. Moreover, the relative weight of the importance of observable socioeconomic factors increases along the quintiles, mostly driven by schooling. This suggests that difference in education attainments contributes substantially in explaining the higher prevalence of obesity among women in the South. Notice these findings prove to be quite robust to alternative scenarios dealing with missing information, BMI bias and regional grouping.

Therefore, we show that a significant part of the cross-regional BMI gap can be mitigated by implementing regional policies focused on improving human capital. One alternative could be rising the amount of compulsory schooling, since this exogenous increase of the amount of human capital has been found to be an effective policy to reduce BMI (Brunello et al., 2013). Moreover, the increase in compulsory education might reduce BMI and obesity indirectly, due to its effect on health-related behaviours and habits (Brunello et al. 2016). However, Spanish regional authorities do not have the competences to change the years of compulsory schooling autonomously, since these kind of policies are determined at the state level. Therefore, regional policy-makers should design policies aimed at reducing school dropout and improving education quality, also through the introduction/improvement of health education programs during the first stages of the education process. Moreover, 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 increasing the endowment of schooling in the South would be especially beneficial in mitigating differences in overweight and obesity between the two groups of regions. Such policy interventions would additionally reduce differences in obesity-related diseases and improve health in general, inasmuch obesity constitutes a key risk factor for many chronic conditions and health complications. However, it must also be stressed that even equalizing the schooling 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.

It is worth mentioning that the decomposition framework applied in this paper is not aimed at estimating causal relationships, but rather at disentangling the contribution of

observable characteristics and their corresponding coefficients to regional differentials in BMI in an “accounting” sense (Fortin et al., 2011). Therefore, the estimates regarding years of education (as well as other covariates) should not be interpreted as causal evidence, since we are not able to address the potential endogeneity of schooling in the BMI regressions with the available data. However, the results are strong and robust enough not only in order to motivate additional research, based on methods that enable causal inference, but also to derive policy recommendations.

Indeed, the evidence reported in this paper is in line with the results from existing related research, suggesting that regional inequalities in education are responsible for regional health inequalities (Safaei, 2014; Ergin and Kunst, 2015). Indeed, 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 individuals (especially women) residing in different Spanish regions. Therefore, investigating the causal effect of education in mitigating regional disparities in BMI, overweight and obesity and consequently other health related variables should be the subject of future research.

References

- Allison PD. Missing Data. Series: Quantitative Applications in the Social Sciences 136. Sage Publications Inc.; 2001.
- Ballas D, Lupton R, Dale R, Kavrouidakis D, Hennig B, Yiagopoulou V, et al. Mind the Gap – education inequality across EU regions. NESSE Network for Experts for the European Commission, 2012.
- Ballesteros-Arribas JM, Dal-Ré Saavedra M, Pérez-Farinos N, Villar Villalba C. The Spanish strategy for nutrition, physical activity and the prevention of obesity (NAOS Strategy). *Revista Espanola de Salud Publica*. 2007;81: 443-449.
- Blinder AS. Wage discrimination: reduced form and structural estimates. *Journal of Human Resources*. 1973; 8(4): 436-455.
- Boström G, Diderichsen F. Socioeconomic differentials in misclassification of height, weight and body mass index based on questionnaires data. *International Journal of Epidemiology*. 1997; 26(4): 860-866.
- Brunello G, Fabbri D, Fort M. The causal effect of education on body mass: Evidence from Europe. *Journal of Labor Economics*. 2013; 31(1): 195-223.
- Brunello G, Fort M, Schneeweis N, Winter-Ebmer R. The causal effect of education on health: What is the role of health behaviors? *Health Economics*. 2016; 25(3): 314-336.
- Cawley J. The impact of obesity on wages. *Journal of Human Resources*. 2004; 39(2): 451-474.
- Costa-Font J, Fabbri D, Gil J. Decomposing body mass index gaps between Mediterranean countries: a counterfactual quantile regression analysis. *Economics & Human Biology*. 2009; 7: 351-365.

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

Devaux M, Sassi F, Church J, Cecchini M, Borghonovi F. Exploring the relationship between education and obesity. *OECD Journal: Economic Studies*. 2011; 1: 1-40.

Dutton DJ, McLaren L. 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*. 2016; 70: 367-373.

Ellis A, Fry R. Regional health inequalities in England. *Regional Trends*. 2010; 42: 60-79.

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

Finkelstein EA, Trogon JG, Cohen JW, Dietz W. Annual medical spending attributable to obesity: payer-and service specific estimates. *Health Affairs*. 2009; 28(5): w822-w831.

Firpo S, Fortin NM, Lemieux T. Unconditional quantile regressions. *Econometrica*. 2009; 77(3): 953-973.

Fortin NM. The gender wage gap among young adults in the United States: the importance of money versus people. *Journal of Human Resources*. 2008; 43(4): 884-918.

Fortin NM, Lemieux T, Firpo S. Decomposition Methods in Economics. *Handbook of Labor Economics*, O. Ashenfelter and D. Card (Eds.), Chapter 1, Volume 4A: 1-102. North Holland, 2011.

Franco M, Sanz B, Otero L, Domínguez-Vila A, Caballero B. Prevention of childhood obesity in Spain: a focus on policies outside the health sector. *SESPAS report 2010, Gaceta Sanitaria*. 2010; 24: 49-55.

Galego A, Pereira J. Decomposition of regional wage differences along the wage distribution in Portugal: the importance of covariates. *Environment and Planning A*. 2014. 46(10): 2514-2532.

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

Gutierrez-Fisac G, Rodriguez-Artalejo F, Guallar-Castillon P, Banegas Banegas J, del Rey Calero J. Determinants of geographical variations in body mass index (BMI) and obesity in Spain. *International Journal of Obesity and Related Metabolic Disorders*. 1999; 23(4): 342–347.

Hacking J, Muller S, Buchan J. Trends in mortality from 1965 to 2008 across the English North-South divide: comparative observational study. *BMJ*. 2011; 342-508.

Herrera-Idárraga P, López-Bazo E, Motellón E. Regional wage gaps, education and informality in an emerging country: The case of Colombia. *Spatial Economic Analysis*. 2016; 11(4): 432-456.

INE – Spanish National Statistics Institute (2019) Encuesta Nacional de Salud 2017 (“<http://www.ine.es/en/welcome.shtml>”).

Jann B. The Blinder-Oaxaca decomposition for linear regression models. *The Stata Journal*. 2008; 4: 453-479.

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

Kinge, J.M. Body mass index and employment status: a new look. *Economics & Human Biology*. 2016; 22: 117-125.

Kuczmarski MF, Kuczmarski RJ, Najjar M. 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*. 2001; 101(1): 28-34.

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

López-Bazo E, Motellón E. Human capital and regional wage gaps. *Regional Studies*. 2012; 46(10): 1347-1365.

Machado JAF, Mata J. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*. 2005; 20: 445-465.

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

Motellón E, López-Bazo E, El-Attar M. Regional heterogeneity in wage distributions: evidence from Spain. *Journal of Regional Science*. 2011; 51(3): 558-584.

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

Murillo-Huertas IP, Ramos R, Simón H. (2020). Revisiting interregional wage differentials: new evidence from Spain with matched employer-employee data. *Journal of Regional Science*. 2020; 60(2): 296-347.

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

OECD. *Obesity and the Economics of Prevention: Fit not Fat*. OECD Publishing, Paris. 2010. <https://doi.org/10.1787/9789264084865-en>.

OECD. Obesity and The Economics Of Prevention: Fit Not Fat, Key Facts – Spain, Update 2014. <https://www.oecd.org/spain/Obesity-Update-2014-SPAIN.pdf>

OECD. OECD Regions and Cities at a Glance 2018, OECD Publishing, Paris. 2018. https://doi.org/10.1787/reg_cit_glance-2018-en

Pereira J, Galego A. Regional wage differentials in Portugal: Static and dynamic approaches. *Papers in Regional Science*. 2011; 90(3): 529-548.

Pereira J, Galego A. Inter-regional wage differentials in Portugal: an analysis across the wage distribution. *Regional Studies*. 2014; 48(9): 1529-1546.

Raftopoulou A. Geographic determinants of individual obesity risk in Spain: A multilevel Approach. *Economics & Human Biology*. 2016; 24: 185–193.

Safaei J. Socioeconomic Health Disparities in Canadian Regions. *Open Medicine Journal*. 2014; 1: 90-95.

Sen B. Using the Oaxaca-Blinder decomposition as an empirical tool to analyze racial disparities in Obesity. *Obesity*. 2014; 22(7): 1750-1755.

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

Valdes S. et al. Prevalence of obesity, diabetes and other cardiovascular risk factors in Andalusia (southern Spain). Comparison with national prevalence data. *Revista Española de Cardiología (English Edition)*. 2014; 67: 442-448.

WHO. Global Action Plan for the Prevention and Control of Noncommunicable Diseases 2013-2020. 2017. <http://www.who.int/nmh/publications/ncd-action-plan/en/>.

Figures

Figure 1: Average BMI per region and by gender

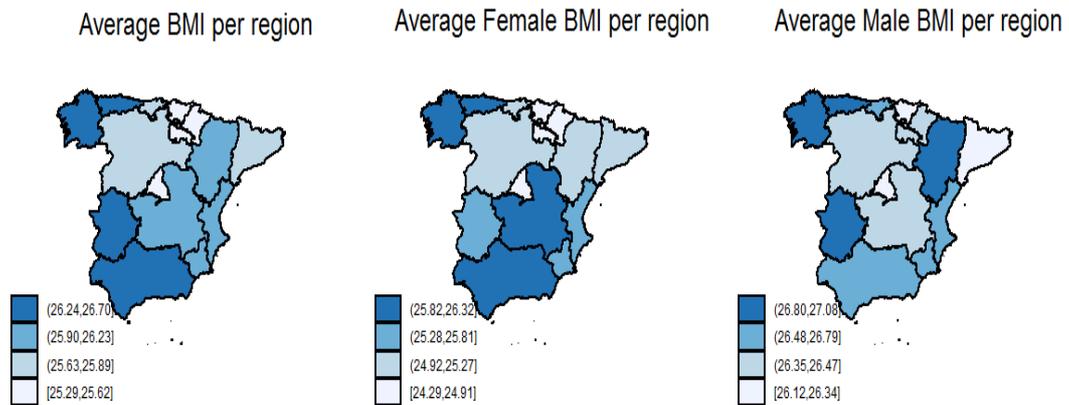
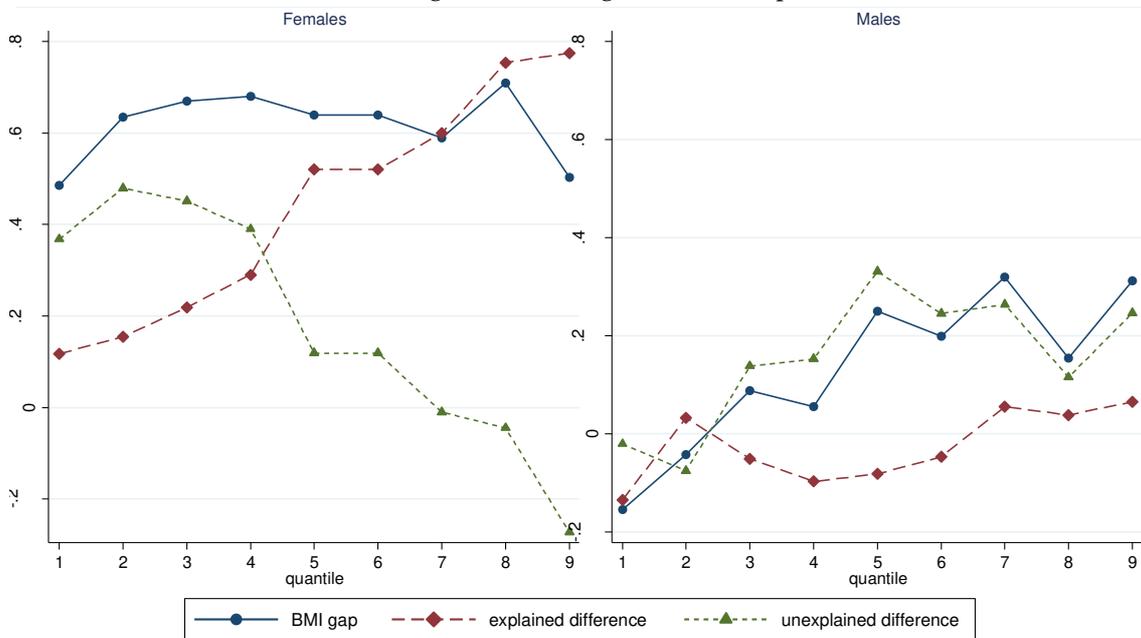


Figure 2: RIF-Regression Decomposition



Tables

Table 1: Groups of regions

	Sample size	%	Mean BMI	S.D. BMI
South				
Andalucía	1475	58.62	26.131	4.547
Extremadura	538	21.38	26.089	4.412
Murcia	503	19.99	25.744	4.323
Total	2516	100	26.044	4.475
North				
Asturias	497	18.49	26.107	4.665
Cantabria	313	11.64	25.814	4.525
Galicia	547	20.35	26.531	4.839
Navarra	412	15.33	25.247	4.131
Basque Country	637	23.70	25.188	4.423
Rioja	282	10.49	24.915	3.901
Total	2688	100	25.684	4.507

Table 2: Descriptive statistics by groups of regions for Women

Variables	South		North		Diff. South-North
	Mean	<i>s.d.</i>	Mean	<i>s.d.</i>	
Height	162.01	6.40	161.79	6.33	0.22
Weight	66.25	12.09	64.63	12.33	1.61***
BMI	25.28	4.62	24.73	4.79	0.55***
Sociodemographic characteristics					
Age: 18-35	0.23	0.42	0.22	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	0.23	0.41	0.24	0.43	-0.01
Household composition					
Married	0.63	0.48	0.57	0.49	0.06***
Kids	0.59	0.83	0.47	0.74	0.12***
Socioeconomic status					
Schooling	10.48	4.53	11.99	4.27	-1.51***
Income1	0.32	0.44	0.16	0.33	0.16***
Income2	0.26	0.41	0.21	0.36	0.05***
Income3	0.25	0.41	0.29	0.41	-0.04**
Income4	0.12	0.29	0.24	0.38	-0.12***
Income5	0.05	0.20	0.10	0.26	-0.05***
Working	0.46	0.49	0.62	0.48	-0.16***
Lifestyle variables					
Sedentary job	0.28	0.44	0.32	0.46	-0.04**
Weekly sport activities	0.10	0.30	0.09	0.27	0.01
Daily smoker	0.28	0.44	0.23	0.42	0.05***
Weekly alcohol consumption	2.28	5.15	3.35	6.16	-1.07***
Food habits variables					
Meat	0.25	0.43	0.36	0.47	-0.11***
Fruit	0.76	0.42	0.80	0.39	-0.04***
Vegetables	0.73	0.43	0.74	0.44	-0.01
Legumes	0.06	0.24	0.05	0.22	0.01
Number of observations	1277		1372		

Note: *** denotes significance at 1%; ** denotes significance at 5%; * denotes significance at 10%.

Table 3: Descriptive statistics by groups of regions for Men

Variables	South		North		Diff. south-north
	Mean	s.d.	mean	s.d.	
Height	174.04	7.13	174.20	6.85	-0.16
Weight	81.22	13.21	80.99	12.61	0.23
BMI	26.83	4.17	26.67	3.83	0.16
Sociodemographic characteristics					
Age: 18-35	0.25	0.43	0.19	0.39	0.07***
Age: 36-45	0.28	0.44	0.28	0.44	-0.00
Age: 46-55	0.28	0.44	0.26	0.44	0.02
Age: 55-65	0.19	0.39	0.28	0.44	-0.09***
Household composition					
Married	0.62	0.48	0.59	0.49	0.03*
Kids	0.56	0.82	0.42	0.72	0.14***
Socioeconomic status					
Schooling	9.99	4.10	11.24	4.01	-1.25***
Income1	0.30	0.43	0.13	0.30	0.17***
Income2	0.27	0.42	0.22	0.37	0.05***
Income3	0.25	0.41	0.30	0.42	-0.05***
Income4	0.13	0.31	0.25	0.39	-0.12***
Income5	0.06	0.22	0.10	0.27	-0.04***
Working	0.60	0.48	0.66	0.47	-0.06***
Lifestyle variables					
Sedentary job	0.31	0.46	0.31	0.46	-0.00
Weekly sport activities	0.16	0.36	0.14	0.33	0.02
Daily smoker	0.34	0.47	0.30	0.46	0.04*
Weekly alcohol consumption	8.91	13.38	10.09	14.02	-1.18**
Food habits variables					
Meat	0.30	0.45	0.40	0.48	-0.10***
Fruit	0.71	0.45	0.68	0.46	0.03
Vegetables	0.62	0.48	0.60	0.49	0.02
Legumes	0.06	0.23	0.08	0.27	-0.02**
Number of observations	1239		1316		

Note: *** denotes significance at 1%; ** denotes significance at 5%; * denotes significance at 10%.

Table 4: Oaxaca Decomposition (Aggregated Results)

	Women		Men	
	Mean BMI	z-stat	Mean BMI	z-stat
Overall decomposition				
South regions	25.281	195.19	26.832	226.56
North regions	24.735	187.86	26.675	251.65
BMI Difference (south-north)	0.546	2.96	0.157	0.99
<i>Explained difference</i>	0.402	4.37	0.002	0.03
Sociodemographic characteristics	-0.054	-1.42	-0.118	-2.92
SES	0.548	7.64	0.201	4.02
Lifestyle	-0.091	-2.33	-0.081	-2.58
<i>Unexplained difference</i>	0.144	0.80	0.155	0.97
Sociodemographic characteristics	0.145	0.58	0.593	2.72
SES	0.188	0.75	-0.136	-0.25
Lifestyle	-0.931	-1.90	0.615	1.57
Constant	0.743	0.78	-0.916	-1.23

Table 5: Oaxaca Decomposition (Detailed Results)

	Women				Men			
	Mean BMI		z-stat		Mean BMI		z-stat	
Overall Decomposition								
South regions	25.281		195.19		26.832		226.56	
North regions	24.735		187.86		26.675		251.65	
BMI Difference (south-north)	0.546		2.96		0.16		0.99	
	Explained		Unexplained		Explained		Unexplained	
	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat	Coef.	z-stat
	0.402	4.37	0.144	0.80	-0.002	-0.03	0.155	0.97
Sociodemographic characteristics	-0.054	-1.42	0.145	0.58	-0.118	-2.92	0.593	2.72
Age: 18-35	-0.010	-0.62	-0.093	-1.20	-0.079	-3.67	0.112	1.74
Age: 36-45	-0.002	-0.36	-0.112	-1.15	-0.001	-0.16	-0.055	-0.66
Age: 46-55	0.000	0.05	-0.013	-0.16	0.008	1.13	-0.099	-1.41
Age: 55-65	-0.016	-1.03	0.201	2.17	-0.051	-2.92	0.014	0.18
Married	0.023	1.60	-0.068	-0.26	0.023	1.58	0.399	1.66
Kids	-0.051	-2.52	0.231	1.74	-0.019	-1.12	0.222	2.01
SES	0.548	7.64	0.188	0.75	0.201	4.02	-0.136	-0.25
Schooling	0.347	5.89	0.326	0.49	0.129	3.96	0.262	0.54
Income1	0.053	1.61	-0.124	-1.21	-0.007	-0.22	-0.044	-0.54
Income2	0.023	1.95	-0.018	-0.21	0.017	1.76	0.184	2.45
Income3	-0.004	-0.71	-0.049	-0.58	-0.004	-0.59	-0.021	-0.28
Income4	0.033	1.50	0.056	0.93	0.053	2.68	-0.063	-1.21
Income5	0.030	2.25	0.032	0.94	-0.002	-0.18	-0.009	-0.26
Working	0.065	2.00	-0.035	-0.17	0.015	1.19	-0.444	-1.80
Lifestyle variables	-0.091	-2.33	-0.931	-1.90	-0.081	-2.58	0.615	1.57
Sedentary job	-0.033	-1.90	-0.008	-0.06	-0.002	-0.24	0.139	1.26
Weekly sport activities	-0.004	-0.58	0.032	0.65	-0.021	1.23	-0.064	-1.25
Daily smoker	-0.030	-2.05	0.088	0.84	-0.027	-1.59	-0.031	-0.27
Weekly alcohol consumption	0.018	1.18	-0.156	-2.03	-0.008	-1.02	0.118	1.09
Meat	-0.033	-1.50	-0.047	-0.39	-0.021	-1.21	0.016	0.14
Fruits	-0.006	-1.59	-0.677	-1.90	-0.002	-0.34	-0.108	-0.43
Vegetables	-0.001	-0.30	-0.216	-0.71	0.003	0.62	0.550	2.79
Legumes	-0.002	-0.22	0.053	1.32	-0.004	0.58	-0.005	-0.11
Constant			0.743	0.78			-0.916	-1.23

Table 6: Quantile Decomposition (Women)

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9
South	20.222	21.457	22.461	23.444	24.408	25.561	26.908	28.712	31.694
z-stat	162.99	172.93	172.85	169.22	163.51	147.38	140.36	119.76	85.75
North	19.736	20.823	21.791	22.764	23.749	24.922	26.319	28.003	31.192
z-stat	185.98	179.61	176.04	173.66	169.86	150.18	139.75	119.57	100.79
BMI Difference (south-north)	0.485	0.634	0.670	0.680	0.660	0.640	0.589	0.709	0.503
z-stat	2.97	3.73	3.73	3.57	3.10	2.66	2.19	2.12	1.04
Explained difference	0.117	0.155	0.219	0.290	0.360	0.521	0.599	0.754	0.775
z-stat	1.77	2.09	2.68	3.24	3.81	4.68	4.91	5.12	3.90
Sociodemographics	0.012	0.001	-0.007	-0.043	-0.072	-0.065	-0.056	-0.058	-0.121
z-stat	0.40	0.03	-0.16	-0.94	-1.38	-1.39	-1.34	-1.10	-1.91
SES	0.132	0.176	0.255	0.367	0.510	0.652	0.754	0.909	1.046
z-stat	2.71	3.35	4.60	5.97	7.53	7.69	7.61	7.53	6.49
Lifestyle habits	-0.026	-0.022	-0.029	-0.033	-0.078	-0.066	-0.099	-0.097	-0.150
z-stat	-0.77	-0.71	-0.84	-0.92	-2.48	-1.43	-1.81	-1.50	-1.61
Unexplained difference	0.368	0.479	0.451	0.390	0.300	0.119	-0.010	-0.044	-0.272
z-stat	2.14	2.75	2.49	2.06	1.37	0.50	-0.04	-0.13	-0.56
Sociodemographics	0.350	0.061	0.002	-0.058	0.212	0.269	0.123	0.232	-0.225
z-stat	1.49	0.25	0.01	-0.22	0.65	0.84	0.34	0.51	-0.34
SES	0.060	0.067	-0.404	-0.211	-0.347	-1.150	-0.199	0.757	0.145
z-stat	0.11	-0.12	-0.67	-0.33	-0.28	-1.39	-0.21	0.64	0.08
Lifestyle habits	-1.111	-1.019	-0.731	-0.233	-0.606	-0.478	-1.440	-1.328	-1.096
z-stat	-2.21	-2.19	-1.53	-0.46	-0.75	-0.75	-2.06	-1.51	-0.87
Constant	1.068	1.370	1.584	0.892	1.041	1.478	1.506	0.294	0.903
z-stat	1.39	1.74	1.93	1.01	0.74	1.31	1.19	0.18	0.39

Appendix

Table A: Description of dependent and independent variables

Variables	Description
<i>Dependent Variable</i>	
BMI	weight in kg divided by height in meters squared ((kg) / [height (m)] ²)
<i>Independent Variables</i>	
<i>Demographics</i>	
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
<i>SES</i>	
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 €/month
Income3	net family income ranges from 1401 to 2040 €/month
Income4	net family income ranges from 2041 to 3280 €/month
Income5	net family income is higher than 3280 €/month
Working	1 if working, 0 otherwise
<i>Lifestyle variables</i>	
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 B: BMI Estimations: OLS Results

	Women		Men	
	South	North	South	North
Constant	26.631*** (0.584)	26.825*** (0.718)	26.254*** (0.487)	26.904*** (0.525)
Sociodemographic characteristics				
Age: 18-35 (reference category)	<i>reference category</i>			
Age: 36-45	0.654* (0.385)	0.640* (0.358)	1.018*** (0.333)	1.731*** (0.327)
Age: 46-55	1.655*** (0.378)	1.289*** (0.383)	1.107*** (0.341)	1.995*** (0.316)
Age: 55-65	2.580*** (0.431)	1.298** (0.474)	1.527*** (0.400)	1.984*** (0.346)
Household composition				
Married	0.363 (0.307)	0.477 (0.305)	1.058*** (0.305)	0.404 (0.255)
Kids	-0.198 (0.180)	-0.638*** (0.177)	0.050 (0.171)	-0.419*** (0.155)
Socioeconomic status				
Schooling	-0.219*** (0.038)	-0.247*** (0.044)	-0.089* (0.033)	-0.114*** (0.031)
Income1 (reference category)	<i>reference category</i>			
Income2	0.290 (0.341)	-0.164 (0.448)	0.791* (0.337)	-0.226 (0.411)
Income3	-0.107 (0.350)	-0.444 (0.461)	0.139 (0.373)	0.028 (0.410)
Income4	-0.174 (0.486)	-1.027** (0.480)	-0.513 (0.417)	-0.403 (0.431)
Income5	-0.491 (0.539)	-1.456** (0.601)	0.163 (0.515)	0.001 (0.537)
Working	-0.430 (0.273)	-0.365 (0.284)	-0.617** (0.284)	0.084 (0.268)
Lifestyle variables				
Sedentary job	0.831** (0.297)	0.858** (0.313)	0.600* (0.264)	0.162 (0.230)
Weekly sport activities	-0.435 (0.392)	-0.775** (0.353)	-1.415*** (0.251)	-0.979*** (0.243)
Daily smoker	-0.527* (0.288)	-0.875*** (0.298)	-0.905*** (0.253)	-0.807** (0.246)
Weekly alcohol consumption	-0.049* (0.029)	0.007 (0.018)	0.013 (0.008)	-0.000 (0.007)
Food habits variables				
Meat	0.212 (0.303)	0.371 (0.266)	0.196 (0.263)	0.159 (0.207)
Fruits	-0.271 (0.324)	0.590* (0.320)	-0.165 (0.270)	0.111 (0.238)
Vegetables	0.099 (0.288)	0.392 (0.302)	0.562* (0.238)	-0.337 (0.220)
Legumes	-0.367 (0.461)	-1.258** (0.494)	-0.137 (0.512)	0.208 (0.384)
R-squared	0.160	0.143	0.116	0.095
Number of Observations	1277	1372	1239	1316

Note: robust standard errors in parenthesis. *** denotes significance at 1%; ** denotes significance at 5%; * denotes significance at 10%.

Table C: Detailed RIFR Decomposition (Women)

	Q7		Q8		Q9	
	mean BMI	<i>z-stat</i>	mean BMI	<i>z-stat</i>	mean BMI	<i>z-stat</i>
South	26.908	140.36	28.712	119.76	31.694	85.75
North	26.319	139.75	28.003	119.57	31.192	100.89
	Coef.	<i>z-stat</i>	Coef.	<i>z-stat</i>	Coef.	<i>z-stat</i>
BMI Difference (south-north)	0.589	2.19	0.709	2.12	0.503	1.04
Explained difference	0.599	4.91	0.754	5.12	0.775	3.90
Sociodemographics	-0.056	-1.34	-0.058	-1.1	-0.121	-1.91
Age: 18-35	-0.010	-0.61	-0.008	-0.60	-0.014	-0.61
Age:36-45	0.000	-0.19	-0.004	-0.36	-0.003	-0.34
Age:46-55	0.000	0.05	0.000	0.05	0.000	0.05
Age: 55-65	-0.015	-0.99	-0.021	-1.00	-0.020	-0.94
Married	0.031	1.53	0.048	1.74	0.007	0.22
Kids	-0.062	-2.20	-0.073	-2.17	-0.092	-2.12
SES	0.754	7.61	0.909	7.53	1.046	6.49
Schooling	0.414	5.57	0.476	5.25	0.566	4.54
Income1	0.119	2.46	0.146	2.41	0.182	2.11
Income2	0.037	2.05	0.052	2.14	0.054	1.73
Income3	-0.007	-0.07	-0.004	-0.37	0.006	0.43
Income4	0.039	1.15	0.102	2.50	0.111	2.10
Income5	0.062	2.90	0.057	2.42	0.053	1.94
Working	0.090	1.80	0.080	1.32	0.075	0.90
Lifestyle habits	-0.099	-1.81	-0.097	-1.5	-0.150	-1.61
Sedentary job	-0.039	-1.84	-0.051	-1.85	-0.087	-1.91
Weekly sport activities	-0.004	-0.54	-0.004	-0.52	-0.005	-0.52
Daily smoker	-0.038	-1.91	-0.030	-1.47	-0.038	-1.38
Weekly alcohol consumption	0.025	1.03	0.041	1.35	0.100	2.35
Meat	-0.033	-1.06	-0.046	-1.19	-0.090	-1.57
Fruits	-0.006	-0.44	-0.002	-0.13	-0.028	-1.07
Vegetables	-0.002	-0.30	-0.003	-0.30	0.001	0.21
Legumes	-0.002	-0.22	-0.003	-0.22	-0.002	-0.21

Table C (continued): Detailed RIFR Decomposition (Women)

	Q7		Q8		Q9	
Unexplained difference	-0.010	-0.04	-0.044	-0.13	-0.272	-0.56
Sociodemographics	0.123	0.34	0.232	0.51	-0.225	-0.34
Age: 18-35	0.000	0.00	-0.064	-0.48	-0.390	-2.06
Age:36-45	-0.086	-0.61	-0.047	-0.27	-0.064	-0.26
Age:46-55	-0.247	-2.01	-0.174	-1.14	0.184	0.81
Age: 55-65	0.289	2.14	0.260	1.50	0.293	1.13
Married	0.107	0.29	0.247	0.52	-0.474	-0.68
Kids	0.059	0.29	0.012	0.05	0.226	0.71
SES	-0.199	-0.21	0.757	0.64	0.145	0.08
Schooling	0.018	-0.02	0.664	0.61	0.830	0.52
Income1	-0.251	-1.73	0.018	0.10	0.029	0.12
Income2	-0.020	-0.16	-0.175	-1.08	-0.176	-0.73
Income3	-0.156	-1.23	-0.149	-0.96	0.120	0.56
Income4	0.149	1.65	0.088	0.84	0.033	0.25
Income5	0.056	1.10	0.040	0.66	0.006	0.07
Working	0.006	0.02	0.271	0.69	-0.697	-1.26
Lifestyle habits	-1.440	-2.06	-1.328	-1.51	-1.096	-0.87
Sedentary job	0.035	0.20	0.037	0.16	0.096	0.28
Weekly sport activities	0.058	0.75	-0.051	-0.54	-0.105	-0.82
Daily smoker	0.021	0.13	0.028	0.15	0.219	0.81
Weekly alcohol consumption	-0.135	-1.09	-0.158	-1.11	-0.119	-0.64
Meat	-0.170	-1.01	-0.061	-0.29	0.244	0.78
Fruits	-0.921	-1.72	-1.263	-1.89	-1.212	-1.31
Vegetables	-0.383	-0.85	0.003	0.00	-0.268	-0.31
Legumes	0.055	0.93	0.137	1.90	0.049	0.11
Constant	1.506	1.19	0.294	0.18	0.903	2.33

Table D: Detailed RIFR Decomposition (Men)

	Q7		Q8		Q9	
	mean BMI	<i>z-stat</i>	mean BMI	<i>z-stat</i>	mean BMI	<i>z-stat</i>
South	28.397	185.44	29.763	151.51	31.958	117.98
North	28.077	188.21	29.608	161.44	31.646	132.72
	Coef.	<i>z-stat</i>	Coef.	<i>z-stat</i>	Coef.	<i>z-stat</i>
BMI Difference (south-north)	0.319	1.49	0.149	0.57	0.320	0.87
Explained difference	0.056	0.57	0.038	0.32	0.066	0.42
Sociodemographics	-0.138	-2.97	-0.163	-3.02	-0.209	-3.16
Age: 18-35	-0.074	-3.34	-0.070	-3.01	-0.093	-3.04
Age:36-45	0.000	-0.16	0.000	0.02	-0.001	-0.16
Age:46-55	0.008	1.06	0.004	0.70	0.003	0.42
Age: 55-65	-0.046	-2.10	-0.073	-2.46	-0.078	-2.03
Married	0.025	1.51	0.023	1.34	0.001	0.03
Kids	-0.051	-2.08	-0.047	-1.64	-0.040	-1.06
SES	0.269	4.04	0.280	3.39	0.392	3.47
Schooling	0.168	3.82	0.163	3.10	0.216	3.03
Income1	-0.013	-0.32	0.002	0.05	0.040	0.52
Income2	0.028	1.96	0.026	1.63	0.022	1.13
Income3	0.000	0.02	0.007	0.59	0.014	0.92
Income4	0.062	2.20	0.050	1.47	0.074	1.69
Income5	-0.001	-0.05	0.000	0.01	-0.008	-0.36
Working	0.025	1.49	0.031	1.45	0.033	1.17
Lifestyle habits	-0.075	-1.98	-0.079	-1.59	-0.118	-1.77
Sedentary job	-0.002	-0.24	-0.004	-0.24	-0.004	-0.24
Weekly sport activities	-0.026	-1.22	-0.033	-1.22	-0.035	-1.22
Daily smoker	-0.018	-1.39	-0.023	-1.39	-0.040	-1.49
Weekly alcohol consumption	-0.007	-0.74	-0.011	-0.80	0.006	0.37
Meat	-0.017	-0.74	0.002	0.06	-0.006	-0.15
Fruits	-0.002	-0.31	-0.007	-0.73	-0.032	-1.28
Vegetables	0.004	0.68	0.004	0.52	0.008	0.70
Legumes	-0.006	-0.69	-0.007	-0.66	-0.015	-0.86

Table D (continued): Detailed RIFR Decomposition (Men)

	Q7		Q8		Q9	
Unexplained difference	0.264	<i>1.20</i>	0.116	<i>0.41</i>	0.247	<i>0.66</i>
Sociodemographics	0.690	2.42	0.209	0.57	0.244	0.48
Age: 18-35	0.107	<i>1.25</i>	0.155	<i>1.48</i>	0.133	<i>0.98</i>
Age:36-45	0.037	<i>0.34</i>	-0.049	<i>-0.36</i>	-0.169	<i>-0.90</i>
Age:46-55	-0.091	<i>-0.93</i>	-0.148	<i>-1.19</i>	-0.272	<i>-1.61</i>
Age: 55-65	-0.063	<i>-0.58</i>	0.005	<i>0.03</i>	0.233	<i>1.20</i>
Married	0.440	<i>1.39</i>	0.090	<i>0.22</i>	-0.015	<i>-0.03</i>
Kids	0.260	<i>1.72</i>	0.156	<i>0.84</i>	0.334	<i>1.33</i>
SES	-0.307	-0.42	-0.316	-0.34	-0.696	-0.54
Schooling	0.305	<i>0.47</i>	0.254	<i>0.31</i>	0.044	<i>0.04</i>
Income1	-0.144	<i>-1.36</i>	-0.147	<i>-1.08</i>	-0.158	<i>-0.82</i>
Income2	0.186	<i>1.78</i>	0.222	<i>1.65</i>	0.288	<i>1.59</i>
Income3	0.003	<i>0.03</i>	0.020	<i>0.16</i>	0.118	<i>0.68</i>
Income4	-0.080	<i>-1.05</i>	-0.090	<i>-0.94</i>	-0.101	<i>-0.84</i>
Income5	0.033	<i>0.66</i>	0.024	<i>0.38</i>	-0.016	<i>-0.19</i>
Working	-0.609	<i>-1.89</i>	-0.599	<i>-1.46</i>	-0.871	<i>-1.56</i>
Lifestyle habits	1.176	2.24	1.077	1.61	1.324	1.46
Sedentary job	0.190	<i>1.30</i>	0.158	<i>0.83</i>	0.001	<i>0.00</i>
Weekly sport activities	-0.051	<i>-0.66</i>	-0.091	<i>-1.05</i>	0.027	<i>0.27</i>
Daily smoker	-0.127	<i>-0.84</i>	-0.197	<i>-1.02</i>	0.017	<i>0.07</i>
Weekly alcohol consumption	0.360	<i>2.36</i>	0.186	<i>0.94</i>	0.156	<i>0.60</i>
Meat	-0.037	<i>-0.24</i>	0.109	<i>0.56</i>	0.084	<i>0.32</i>
Fruits	0.383	<i>1.14</i>	0.375	<i>0.87</i>	-0.060	<i>-0.10</i>
Vegetables	0.482	<i>1.79</i>	0.542	<i>1.59</i>	1.079	<i>2.30</i>
Legumes	-0.023	<i>-0.40</i>	-0.003	<i>-0.04</i>	0.020	<i>0.18</i>
Constant	-1.295	<i>-1.34</i>	-0.855	<i>-1.25</i>	-0.626	<i>-0.35</i>

Table E: Robustness Checks

	Complete Case		(1)	(2)	(3)	(4)
Women						
overall decomposition	mean	z-stat	mean	z-stat	mean	z-stat
south regions	25.281	195.19	25.227	210.79	25.227	210.79
north regions	24.735	187.86	24.682	216.27	24.682	216.27
difference	0.546	2.96	0.545	3.30	0.545	3.30
explained diff.	0.402	4.37	0.350	4.16	0.365	4.33
unexplained diff.	0.144	0.80	0.196	1.19	0.181	1.10
no. of observations	2649		3259		2649	
Men						
overall decomposition	mean	z-stat	mean	z-stat	mean	z-stat
south regions	26.832	226.56	26.758	241.13	26.758	241.13
north regions	26.675	251.65	26.630	284.63	26.630	284.63
difference	0.157	0.99	0.128	0.88	0.128	0.88
explained diff.	0.002	0.03	-0.041	-0.58	-0.046	-0.66
unexplained diff.	0.155	0.97	0.170	1.16	0.174	1.19
no of observations	2555		3098		2555	

Note: estimations obtained with the Complete Case approach are the baseline results reported in Table 4. Results in Column (1) are obtained using also observations with missing information on family income and including a dummy for missing values in this variable. Results in Column (2) are obtained using also observations with missing information on family income and imputing missing values. Results in Column (3) are based on the corrected BMI variable, using the methodology proposed by Gil and Mora (2011). Results in Column (4) are based on the NUTS-2 grouping of regions.

Figures

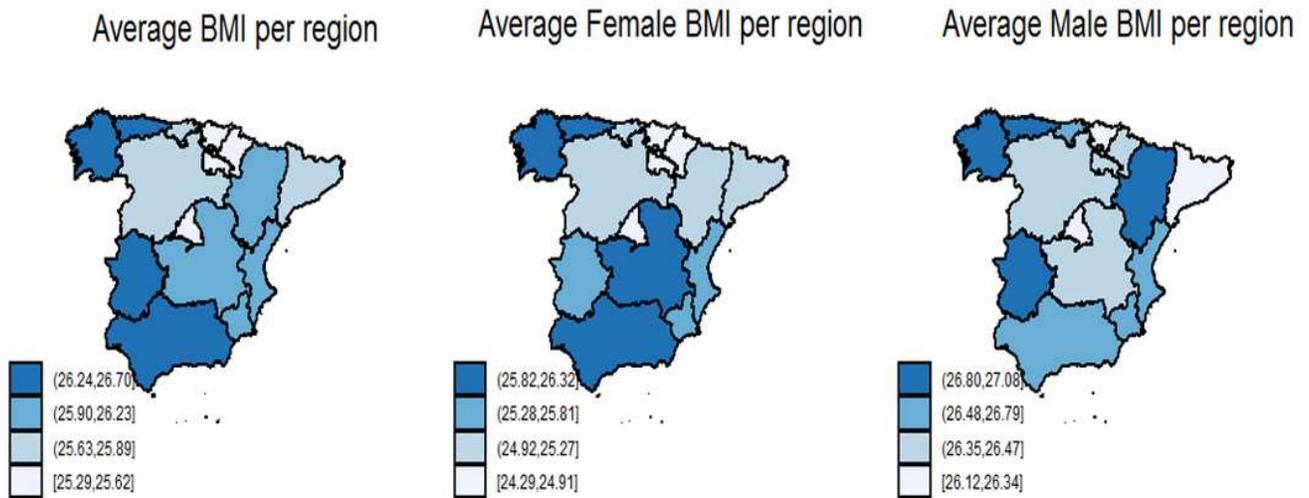


Figure 1

Average BMI per region and by gender

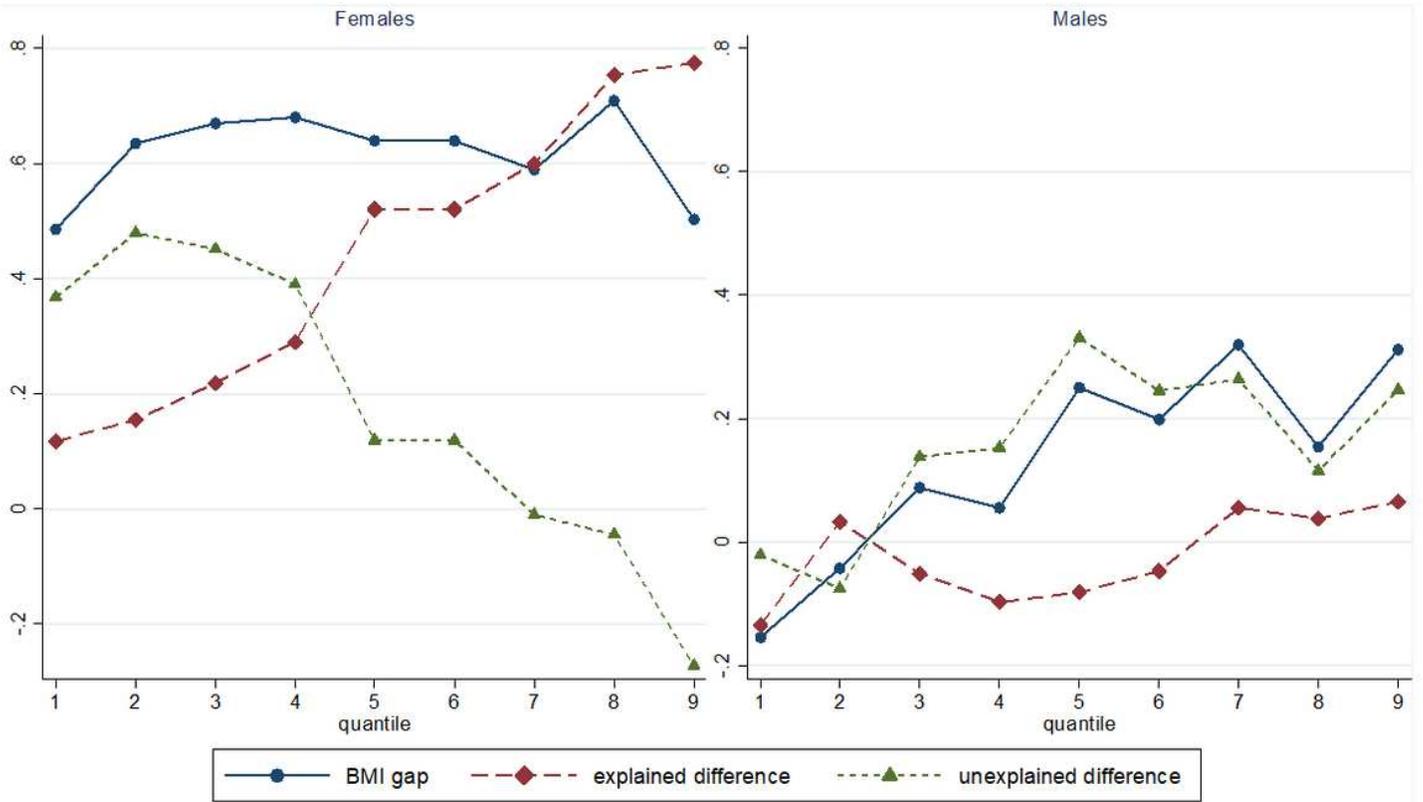


Figure 2

RIF-Regression Decomposition