



DISCUSSION PAPER SERIES

IZA DP No. 5366

## Overweight and Poor? On the Relationship between Income and the Body Mass Index

Dean Jolliffe

December 2010

Forschungsinstitut  
zur Zukunft der Arbeit  
Institute for the Study  
of Labor

# **Overweight and Poor? On the Relationship between Income and the Body Mass Index**

**Dean Jolliffe**

*IZA and NPC*

Discussion Paper No. 5366  
December 2010

IZA

P.O. Box 7240  
53072 Bonn  
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: [iza@iza.org](mailto:iza@iza.org)

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

## ABSTRACT

### Overweight and Poor? On the Relationship between Income and the Body Mass Index<sup>\*</sup>

Contrary to conventional wisdom, NHANES data indicate that the poor have never had a statistically significant higher prevalence of overweight status at any time in the last 35 years. Despite this empirical evidence, the view that the poor are less healthy in terms of excess accumulation of fat persists. This paper provides evidence that conventional wisdom is reflecting important differences in the relationship between income and the body mass index. The first finding is based on distribution-sensitive measures of overweight which indicates that the severity of overweight has been higher for the poor than the nonpoor throughout the last 35 years. The second finding is from a newly introduced estimator, unconditional quantile regression (UQR), which provides a measure of the income-gradient in BMI at different points on the unconditional BMI distribution. The UQR estimator indicates that the strongest relationship between income and BMI is observed at the tails of the distribution. There is a strong negative income gradient in BMI at the obesity threshold and some evidence of a positive gradient at the underweight threshold. Both of these UQR estimates imply that for those at the tails of the BMI distribution, increases in income are correlated with healthier BMI values.

JEL Classification: I1, I18, I32

Keywords: overweight, obesity, body mass index, unconditional quantile regression, Foster-Greer-Thorbecke poverty measures, NHANES

Corresponding author:

Dean Jolliffe  
E-mail: [Jolliffe@alumni.princeton.edu](mailto:Jolliffe@alumni.princeton.edu)

---

<sup>\*</sup> Jolliffe is a Research Fellow with the Institute for the Study of Labor (IZA) in Bonn and a Research Affiliate with the National Poverty Center at the Ford School of Public Policy, University of Michigan. The author wishes to thank for useful comments: John Cawley, Sheldon Danziger, John DiNardo, Lisa Mancino, Chad Meyerhoefer, Mark Prell, Laura Tiehen, Steve Vosti, Muzhe Yang, Jim Ziliak, session participants at the Western Economic Association International conference, and seminar participants at the University of Kentucky, Lehigh University, National Poverty Center, and the Agency for Healthcare Research and Quality. The author also wishes to thank Sergio Firpo, Nicole Fortin and Thomas Lemieux for sharing their code for the UQR estimator. The author assumes responsibility for the contents of this paper and any errors it may contain.

## 1. Introduction

Mortality and morbidity rates for many health outcomes are inversely related to income (Deaton and Paxson, 1999; Deaton, 2001). Deaton (2002) estimates that people in the U.S. with family income less than \$5,000 (in 1980 dollars) have a life expectancy that is around 25 percent lower than those with income above \$50,000. He also notes that the negative income gradient in health has a long history, first documented in France in the 1820s. Perhaps the most prominent U.S. public health concern in the last decade has been the rapid rise of overweight. Ogden et al. (2006) estimate that 66 percent of U.S. adults are overweight based on data from 2003 and 2004. The potential health consequences from being overweight or obese include being at increased risk of morbidity from hypertension, stroke, type 2 diabetes, osteoarthritis, respiratory problems, and breast, prostate, and colon cancers.<sup>1</sup> A reasonable and common assertion then is that the poor suffer significantly higher rates of overweight.

There are important policy implications linked to correctly understanding this relationship. In both the popular press and academic research, there is the argument that the growth of fast food and energy-dense food has been an important cause of the overweight epidemic in the U.S. and that this has disproportionately affected poor people. Drewnowski and Specter (2004) argue that limited economic resources may shift dietary choices toward a diet that provides maximum calories at the least cost. An implication of this line of research is that the poor cannot afford healthy diets.<sup>2</sup> In contrast, Lakdawalla, Philipson and Bhattacharya (2006) argue that technological change has lowered food prices, made work more sedentary, and led to the increase in BMI. They also note that their model of weight transition based on earned and

---

<sup>1</sup> See National Heart, Lung, and Blood Institute (1998, Chapters 1 and 2) for a more complete list of health problems associated with being overweight and for citations for each of the listed health problems.

<sup>2</sup> In popular press, Critser (2003) similarly argues that cheap fats and sugars are the primary cause of overweight and notes that "... one fact stuck out above all others ...In late-twentieth-century America, it was the poor, the underserved, and the underrepresented who were most at risk from excess fat" (p.109).

unearned income explains why “...richer people are thinner than poorer people...” (p. 253).<sup>3</sup> A different line of reasoning suggests that Federal food assistance programs are exacerbating the overweight epidemic. In the Washington Post, Besharov (2002) argues that programs such as food stamps are increasing the food budgets of the poor, who are already over-consuming. In Congressional testimony he states that “Today, as many as 70 percent of low-income adults are overweight, about 10 percent more than the nonpoor” (2003).

The aim of this paper is to first assess the accuracy of the common portrayal of income and overweight prevalence (section 2), and then in section 3, to measure the income gradients in BMI. Section 2 documents what is now reasonably well known to obesity researchers, that there is no correlation between the prevalence of overweight (BMI > 25) and poverty status. A key new finding in this section though, is that distribution-sensitive measures reveal that the severity of overweight is much greater for the poor than the nonpoor. This finding indicates that the BMI distributions for the poor and nonpoor have very different shapes, and this has important implications for the regression models used in section 3 to measure the income gradient in BMI.

There has been research that examines the relationship between income and BMI which has focused on the nonlinearity between income and BMI, that is the correlation between income and BMI differs at different income levels (see for example, Lakdawalla and Philipson, 2002, 2009). This paper focuses on a different issue by measuring whether the correlation between income and BMI varies at different points on the BMI distribution. While both estimates are of interest, the latter is more consistent with the literature examining income-health gradient. The findings indicate that there is a positive income-gradient in BMI at the low-end of the BMI distribution, and a negative income-gradient at the obesity threshold (BMI=30).

---

<sup>3</sup> Cutler, Glaeser and Shapiro (2002) suggest that the BMI increase is primarily due to declines in the pecuniary and non-pecuniary price of food; but they note that income is not an important variable in their models predicting the probability of obesity.

## 2. Overweight and Poverty Status

The official estimates of overweight and obesity come from the National Health and Nutrition Examination Survey (NHANES), which is conducted by the National Center for Health Statistics of the Centers for Disease Control. The NHANES samples are representative of the U.S. civilian, non-institutionalized population and observations are selected following a stratified, multi-stage design. I use seven rounds of the NHANES data: 1971-1974 (NHANES I), 1976-1980 (NHANES II), 1988-1994 (NHANES III), 1999-2000, 2001-2002, 2003-2004 and the 2005-2006 files. Body weight and height measures were obtained by trained health technicians, and effective sample sizes of those persons between 20 and 75 years of age range from 7,592 in the pooled 1999-2002 two-year cycles to 12,901 from NHANES I.

Current medical research indicates that excess accumulation of body fat, as a percent of total body weight, is the primary source of health concerns associated with being overweight. Federal guidelines use the body mass index (BMI), which is body weight in kilograms divided by the square of height in meters, as an approximation for measuring body fat. In 1998, the U.S. Federal Government adopted the recommendations of the World Health Organization Expert Committee (1995) and defined a person as overweight if they had a BMI greater than or equal to 25, and obese as greater than or equal to 30.<sup>4</sup>

### *2.1 Prevalence of overweight by poverty status*

Table 1 lists rates of overweight and obesity from 1971 to 2006 by poor and non-poor categories. Throughout this paper, I categorize an individual as poor if their income is less than 130 percent

---

<sup>4</sup> Throughout this paper, overweight describes people with BMI greater than 25, including those who are obese. This is done primarily to facilitate the extension to the distribution-sensitive measures, but it is also consistent with the initial definition adopted by the U.S. government (see for example, National Center for Health Statistics, 2001). Currently, the term overweight is more typically defined as BMI between 25 and 29.9, which the CDC had previously referred to as pre-obese (see for example, Flegal et al. 1998).

of the poverty line. I use this income cutoff primarily because it matches the gross-income eligibility criterion for the Supplemental Nutrition Assistance Program,<sup>5</sup> the largest of the Federal food assistance programs.<sup>6</sup> In terms of BMI, Table 1 indicates that there has historically been no relationship at all between being poor and being overweight. Between 1971 and 2006, there are no statistically significant differences in the rates of overweight between the poor and nonpoor. While the estimated difference is not statistically significant, the point estimates from the pooled 2003-06 sample indicate that the overweight rate for the poor is actually 2.8 percentage points lower than for the nonpoor.

[INSERT TABLE 1 APPROXIMATELY HERE]

The findings change if we ignore overweight and just consider obesity. Panel B of Table 1 now provides at least some historical evidence that is more consistent with the common assertion in the popular press. Between 1971 and 2002, the poor did have higher rates of obesity and the difference in the rates ranged from 5.1 to 6.5 percentage points higher than the nonpoor. The more current point estimates from 2003-06 though, show essentially no difference in the obesity rate between the poor and nonpoor.<sup>7</sup>

From the simple comparison of poverty status with the overweight and obese outcomes, the

---

<sup>5</sup> Using 130 percent of the poverty line for the poor-nonpoor split also has the advantage of a greater sample of poor observations (relative to using the poverty line) and thereby increases the power of poor – nonpoor comparisons.

<sup>6</sup> In 2008, federal spending on the Supplemental Nutrition Assistance Program (formerly called the Food Stamp Program) was \$37.5 billion, accounting for 62 percent of total spending on food assistance programs. Monthly participation rates topped 31 million people in the fall of 2008 (USDA, 2009).

<sup>7</sup> As a robustness check of the statement that there are no statistically significant poor-nonpoor differences in overweight and obesity prevalence from 2003-06, I consider several definitions of poverty based on the following: Define someone as poor if  $\text{income} < \alpha * (\text{poverty line})$ . For all of the analysis in this paper,  $\alpha = 1.3$ . For the robustness check, I consider 11 values of  $\alpha$  ranging from 0.5 to 1.5, in increments of 0.1. Over all 11 variants of this definition, in no case do the poor have a statistically significantly higher prevalence of overweight or obesity. In one case, the opposite is true. If poverty is defined as having less than 110% of the poverty line, then the nonpoor have a statistically significant higher level of overweight.

evidence is mixed. If overweight (BMI > 25) is the relevant health indicator, then there never has been an association between this measure and poverty. If on the other hand, obesity (BMI > 30) is the relevant measure of health, then the poor have historically been less well off, but this relationship appears to no longer exist. Part of the reason for this mixed message is due to an attempt to simplify a complex relationship. Discussing overweight and obesity in terms of prevalence rates requires that the continuous BMI measure is converted into discrete outcomes, and this loses information. Discrete outcomes for overweight and obese have the important advantage that they are easy for the public to understand, but they also have disadvantages.

In particular, research indicates that the risks of health problems associated with being overweight are increasing in BMI (Willett, Dietz and Colditz, 1999 and Freedman et al., 2002). For example, the risk of heart failure increases 5 percent in adult men and 7 percent in adult women with a unit increase in BMI (Kurth et al., 2002). Similarly, a one-unit increase in BMI is associated with a 6 percent increase in the relative risks of total, ischemic and hemorrhagic stroke for men (Kenchiah et al., 2002). Treating BMI as a discrete outcome largely ignores the research indicating that someone whose BMI is twice the overweight threshold is at higher risk of negative health outcomes than someone whose BMI is just slightly greater than the threshold. As one example, Filardo et al. (2007) assert that categorizing BMI into discrete outcomes not only loses information, but also results in critically biased estimates of the association between BMI and post-CABG (coronary artery bypass graft surgery) morbidity and mortality.

## *2.2 Alternate measures of overweight and poverty status*

To avoid this loss of information, I extend on Jolliffe (2004) and use distribution-sensitive measures of overweight and obesity for the poor and nonpoor. These measures draw from the poverty literature, and were introduced by Foster, Greer and Thorbecke (1984, hereafter referred



to as FGT). Modifying the FGT index, one can express a class of overweight indices,  $OW_\alpha$ , as:

$$OW_\alpha = 1/n \sum_i I(BMI_i \geq f) [(BMI_i - f)/f]^\alpha \quad (1)$$

where  $n$  is the sample size,  $i$  subscripts the individual,  $f$  is the cutoff point identifying who is overweight, and  $I$  is an indicator function which takes the value of one if the statement is true and zero otherwise.<sup>8</sup>

When  $\alpha=0$ ,  $OW_0$  is the proportion of the population that is overweight, or the overweight *prevalence*. When  $\alpha=1$ ,  $OW_1$ , can be described as revealing the *depth* of the problem. A useful interpretation of  $OW_1$  is to recognize that it is equal to the product of the prevalence rate and the average value of excess BMI of the overweight (expressed as a fraction of the overweight cutoff point). When  $\alpha=2$ , the resulting measure is the average of the squared values of the individual overweight-gaps (i.e. the proportionate difference between BMI and the threshold), and is sensitive to (mean-preserving) changes in the BMI distribution of the overweight. Using the poverty semantics,  $OW_2$  can be described as reflecting the *severity* of overweight. The merit of these measures can be illustrated by considering an overweight person who gains weight. This weight gain has no effect on the overweight prevalence ( $OW_0$ ), but the health of this person has changed and this is reflected in changes to  $OW_1$  and  $OW_2$ .

[INSERT FIGURE 1 APPROXIMATELY HERE]

The first indication that there have been important changes in the distribution of BMI that would be masked by prevalence measures, but revealed by the FGT measures, can be seen in

---

<sup>8</sup> The indices and sampling variance are detailed in Jolliffe and Semykina (1999).

Figure 1. This figure compares two distributions of BMI, one for the early 1970s and the other from the pooled samples of 1999-2006. The most striking change is the significant shift to the right of BMI over time. It is clear from this figure that significantly more of the 1999-2006 BMI distribution lies to the right of the cutoff at 25. This shift is reflected in the increased prevalence of overweight from 47 percent in the 1970s to the more recent estimate of 66 percent. What is also clear from the figure is that BMI in 1999-2006 is significantly less-peaked indicating greater spread in the tails. The changing shape of the density functions indicates that the variance in BMI is now greater. This change is largely hidden in the prevalence measures, but will be revealed in the depth and severity measures.

Table 2 presents the three overweight indices by income status, starting with estimates from 1971-1974 NHANES and ending with the pooled NHANES 2003-2006. While the  $OW_0$  measure replicates the finding above, that the poor have never had a greater prevalence of overweight; the depth and severity measures are more consistent with the portrayal in the popular press. Throughout the 1970s and 1980s, both the depth and severity measures were statistically significantly greater for the poor than the nonpoor. While more recently the poor-nonpoor differences have diminished, the overweight severity measure for the poor has continued to be statistically significantly higher up through the current 2003-06 estimates.<sup>9</sup>

The distribution-sensitive measures indicate an important similarity in the change in BMI distribution over time for both the poor and nonpoor. Between the early 1970s (NHANES I) and the current estimates from 2003-06, the rate of increase in each of the overweight measures is greater for larger values of  $\alpha$ . That is  $OW_2$  has increased by more than  $OW_1$ , and similarly  $OW_1$  by more than  $OW_0$ , for both the poor and nonpoor. Noting that a greater  $\alpha$  means  $OW_\alpha$  is more

---

<sup>9</sup> In terms of poor-nonpoor differences, it is also worth noting the difference in growth of the measures. For example, the severity measure for the poor has increased by 135 percent, while this measure for the nonpoor has more than tripled. While the overweight severity is greater for the nonpoor, if the trend continues, one would expect poor-nonpoor convergence in this measure as well at some future point.

sensitive to changes at the high-end of the BMI distribution, this pattern shows that the increases in the prevalence rates fail to reveal that an important component of the change over the last three decades has been a large shift out of the right tail of the BMI distributions for both the poor and nonpoor. For example, while median BMI has increased less than 3 units between NHANES I and the recent estimates from 2003-2006 (increasing from 24.6 to 27.4), the 95<sup>th</sup> percentile has increased by 6.6 units in this period (increasing from 34.1 to 40.7).<sup>10</sup>

[INSERT TABLE 2 APPROXIMATELY HERE]

While there is no poor-nonpoor difference in prevalence of overweight, Table 2 shows that the overweight poor are heavier on average, and this may in part explain the perception that the poor are more overweight. The ratio  $OW_1/OW_0$  provides a measure of the extent to which BMI of the overweight surpasses the overweight threshold of 25. Using the 2003-2006 estimates, this ratio indicates that the overweight poor are 29 percent overweight (i.e. their BMI is on average 29 percent greater than 25), while this estimate for the nonpoor is 26 percent. Figure 2 plots the most recent BMI density functions for the poor and nonpoor and graphically illustrates this point. There is more mass in the density function for the overweight nonpoor near the threshold of 25 (the nonpoor density function lies above the poor density for BMI between about 25 and 33). Similarly, the density function for the poor lies above the nonpoor at the extreme values of BMI, between about 40 and 50.

[INSERT FIGURE 2 APPROXIMATELY HERE]

---

<sup>10</sup> To put this increase in terms of kilograms, consider an average-height male at 1.77 meters (or 5 feet, 10 inches). An increase of BMI by 6.6 means a weight gain of 20.9 kilograms (or 46 pounds).

Gender adds an important dimension to the correlation between poverty and BMI which indicates very different associations for men and women, and the distribution-sensitive measures again provide a much more detailed profile of overweight by sex and income (see Table 3). For example, over the last 35 years, nonpoor men had much higher prevalence of overweight than poor men, but there were largely no poor-nonpoor differences in the severity measures for men.<sup>11</sup> For men, the severity measure dampens the difference observed in the prevalence measure. For women, the opposite is true. Poor women had higher prevalence of overweight in all years, and the severity measures amplify this difference. For example in 2003-06, the overweight prevalence for poor women is about 10 percent greater than for nonpoor women, while the overweight severity for poor women is 40 percent greater than for nonpoor women.

As another example of how measures which account for the distributional differences alter the portrayal, note that that in 2003-06 three fourths of nonpoor men, and 60 percent of nonpoor women, were overweight. Without accounting for distributional differences, one might infer that being overweight is a larger problem for nonpoor men than nonpoor women. The severity measure reveals that this would overly simplify the picture.  $OW_2$  for nonpoor women is more than 30 percent higher than that for nonpoor men, indicating that of the overweight nonpoor, women are overweight by much greater amounts relative to men.<sup>12</sup>

[INSERT TABLE 3 APPROXIMATELY HERE]

### **3. BMI and Income, using continuous measures**

An important motivation for the alternative measures of overweight considered above is the

---

<sup>11</sup> The 1999-2002 severity measure for poor men was greater ( $p < 0.1$ ) than that for the nonpoor.

<sup>12</sup> From the 2003-06 pooled estimates, nonpoor, overweight women exceed the overweight threshold by 30 percent. Nonpoor, overweight men exceed this threshold by 23 percent on average.

argument that the dichotomous (prevalence of overweight) measure fails to reveal information about the changing distribution of those who are overweight. Similarly, treating income as dichotomous (poor and not poor) could very well also be failing to reveal important aspects of the relationship between income and BMI. Case, Lubotsky and Paxson (2002, p. 1308) note that the income gradient in health status “is evident throughout the income distribution.” For example, the decision to treat all of the poor as the same might hide important differences between those who are in severe poverty compared to those whose income is closer to the poverty line. In this section, I avoid converting either the BMI or income measures into dichotomous outcomes, and consider the relation between continuous measures of each.

### *3.1 Correlation between Income and BMI, the conditional mean*

There is a large literature on the income gradient in health outcomes which fairly uniformly documents positive correlation between bad health outcomes and decreases in income. See for example, Pappas et al. (1993), Sorlie, Backlund and Keller (1995), Deaton and Paxson (1999), Deaton (2001). Essentially all of this analysis is based on estimating the correlation between the probability of a negative health outcome and income. BMI as a health outcome has an important complicating factor relative to many other health outcomes (or at least for how these outcomes are typically measured). At high levels of BMI, decreases in BMI indicate health improvements; but at low levels of BMI, increases in BMI indicate health improvements (the Centers for Disease Control and Prevention, CDC, consider 18.5 as the threshold for underweight status).

Given the current epidemic of overweight, it’s reasonable to assume that negative correlation between income and BMI indicates that higher income is associated with better BMI outcomes. But for measurement purposes, it is important to recognize that for the underweight, negative correlation indicates that deteriorating BMI outcomes are associated with increases in income.

An OLS regression of BMI on income and other controls,  $X$ , provides an estimate of the partial derivative of BMI (conditional on  $X$ ) with respect to income, or:

$$\partial E(\text{BMI} \mid \text{Income}, X) / \partial \text{Income} \quad (2)$$

The OLS estimator measures the change in the mean of BMI from a change in income. A nonparametric (or lowess, or spline) estimator would allow the estimated partial derivative to vary at different levels of income, but it would continue to estimate the change in the mean of BMI from a change in income. Lakdawalla and Philipson (2002, 2009) provide a theoretical explanation for when we would expect the correlation between BMI and income to differ by income. They propose that the demand for food and the demand for an ideal body weight can have differing effects at different levels of income. At low levels of income, an increase in income might mean increased demand for food and weight, which work together to increase BMI. At high levels of income, an ideal weight might mean weight reduction, offsetting the demand for more food, and potentially resulting in decreased BMI. Lakdawalla and Philipson also estimate their model, using income splines, and indeed validate that the correlation between BMI and income changes at different points on the income distribution.

### *3.2 Correlation between Income and BMI, at different points on the BMI distribution*

Estimates such as these fail to allow for the possibility that income could have very different effects on BMI at different points on the distribution of BMI. This is the same point made by Chamberlain (1994) who compares OLS and quantile estimators to measure the wage premium from union participation. The OLS estimates indicate that union participation has a positive effect on mean wages. The quantile estimates allow one to see that the premium is much larger

for low (conditional) wage earners than for high (conditional) wage earners, and the OLS estimate falls between the two.

In contrast to the union example, where the union effect is diminished at higher points on the conditional wage distribution, one might expect the income effect on BMI to potentially reverse signs. In other words, if there is an income gradient in BMI which indicates a positive relationship between income and improvements in BMI, then there should be positive correlation at low levels of BMI and negative correlation at high levels. The OLS estimator is unable to reflect this diversity. The obvious alternative estimate of the marginal effect of income on BMI, which would allow for variation in the slopes, would be to use the quantile estimator:

$$\partial Q_{BMI}(\tau | Income, X) / \partial Income \quad (3)$$

Following the notation of Koenker (2005),  $Q_{BMI}$  is the conditional quantile function of BMI and  $\tau$  represents quantiles of the conditional BMI distribution. The expression in Equation (3) is the conditional quantile marginal effect (CQME), estimated by the quantile estimator.

It is not always the case that the quantile estimator will necessarily provide qualitatively different information from OLS, but Koenker and Basset (1982) show that in the presence of a heteroscedastic error distribution, the quantile estimator will typically differ from the OLS estimator.<sup>13</sup> The analysis in section 2 provides evidence that there are important and significant differences in the BMI distribution of the poor and nonpoor, which suggests that the error structure from a regression of BMI on income is likely to be heteroscedastic.

---

<sup>13</sup> More accurately, they show that variation in the quantile estimator at different points of the conditional distribution can be used as a basis to test for heteroscedasticity. An implication of their test is that if the error distribution is homoscedastic, the quantile estimator will provide similar information at different points on the conditional distribution (including the conditional mean).

Stifel and Averett (2009) examine the correlation between children's BMI and several explanatory variables such as mother's characteristics and household composition, and argue that OLS fails to reveal important correlations. They use the quantile estimator to measure these correlations at five points (the 5<sup>th</sup>, 25<sup>th</sup>, 50<sup>th</sup>, 85<sup>th</sup>, and 90<sup>th</sup> percentiles) on the conditional distribution of BMI. They find that "OLS models mask some of the correlates of childhood BMI at both the upper and lower tails of the weight distribution" (p. 396).

With the quantile estimator, marginal effects are typically compared at fixed points on the conditional distribution, such as the five points used by Stifel and Averett.<sup>14</sup> Because I examine the income gradient in BMI using data from different points in time, this approach of fixing the quantile will produce estimates that are difficult to interpret. For example, the quantile estimator evaluated at the conditional median on data from 1988 and later would estimate the relationship between income and BMI at some point above the overweight threshold. But for the earlier years, the conditional median (for the specifications considered in this paper) BMI is below the overweight threshold, and the public health literature is fairly silent as to whether we believe health is positively or negatively affected for BMI changes between 18.5 and 25.

Rather than fixing the quantile, the more relevant concern is to estimate the marginal effect at a fixed value of BMI. The current medical literature designates primarily three values of BMI as key thresholds – 18.5, which defines underweight; 25, which defines overweight; and 30, which defines obesity.<sup>15</sup> As these are the thresholds for defining this public health concern, I propose that they are the policy-relevant points at which to measure whether there is an income gradient in BMI. This choice also helps to place these findings in the context of related literature documenting that there is a negative income gradient in poor health outcomes. Modifying

---

<sup>14</sup> For other examples of estimating at a fixed point on the conditional distribution, see Chamberlain, 1994, Nguyen et al. (2007), Patrinos and Sakellariou (2006).

<sup>15</sup> Obesity is sometimes further decomposed into different categories, including morbid (or class 3) obesity defined as a BMI greater than 40.



equation (3) slightly, I express the CQME evaluated at each of the  $\tau^{\text{th}}$  quantile of the BMI distribution corresponding to BMI values of 18.5, 25, and 30. For example, if the prevalence of obesity (threshold BMI=30) in the selected sample year is 34 percent, the CQME is estimated at the 66<sup>th</sup> quantile of the conditional distribution.

$$\partial Q_{BMI}(\tau_{BMI=18.5,25,30} | Income, X) / \partial Income \quad (4)$$

The advantage of the quantile estimator over OLS, as noted above, is that it can be used to estimate marginal effects at the tails of the conditional distribution, allowing for the concern that the sign of the marginal effect might switch. The disadvantage of this estimator is that policy makers typically aren't interested in the CQME, but rather want to know about the effect of the explanatory variable on the *unconditional* distribution of the relevant statistic. In the case of the quantile estimator, the nonlinearity of the estimator means that the CQME is not equal to the unconditional quantile marginal effect (UQME). The parameter estimate from the quantile estimator (3) will describe the change in conditional BMI at the  $\tau^{\text{th}}$  quantile; but it does not measure change in BMI at that  $\tau^{\text{th}}$  quantile. Equation (5) formalizes this statement. For example, if  $\mathbf{X}$  includes controls for education, the  $\tau^{\text{th}}$  quantile of BMI conditional on low education status will correspond to a different BMI level than for the CQME evaluated at the same quantile for a person with more education.

$$\partial Q_{BMI}(\tau | Income, X) / \partial Income \neq \partial Q_{BMI}(\tau) / \partial Income \quad (5)$$

An important advantage of OLS is that the distinction between the conditional and unconditional distributions is not a concern. The OLS estimator is a consistent estimator of the

marginal effect of some explanatory variable on both the conditional and unconditional mean of the dependent variable. This latter characteristic comes from the linearity of OLS, and the law of total expectations (also known as the law of iterated expectations). If  $X$  and  $Y$  are random variables, and the  $E(Y) < \infty$  (i.e.  $Y$  is integrable), then the expected value of  $Y$  is equal to the conditional expected value of  $Y$  given  $X$  (i.e.  $E(Y) = E(E(Y|X))$ ). This means that the OLS estimator,  $\beta_{OLS}$ , estimates the marginal effect of  $X$  on the mean of  $Y$ .

Firpo, Fortin and Lemieux (2009, hereafter referred to as FFL) propose a new estimator, the Unconditional Quantile Regression (UQR) estimator, which has attractive characteristics of both the OLS and the quantile estimators. The UQR estimator allows marginal effects to be estimated at different points on the distribution, like the quantile estimator; and also has the characteristic that it respects the law of total expectations, like OLS. The implication of this statement is that it estimates the Unconditional Quantile Marginal Effect (UQME), meaning it allows the estimation of the marginal effect of income on BMI (at each of the points on the BMI distribution corresponding to the public-health thresholds), while also conditioning on  $X$ .

The UQR estimator is based on influence functions, which were introduced by Hampel (1988) as a tool in robust estimation techniques. Consider some distributional statistic,  $v(F_Y)$ , such as the median, inter-quantile range, or any quantile. The influence function,  $IF(Y; v, F_Y)$ , represents the influence of an individual observation on the distributional statistic,  $v(F_Y)$ . A key innovation of FFL is that they center the influence function by adding  $v(F_Y)$  to it, and call this a re-centered influence function (RIF). By design then, the  $E(RIF(Y; v, F_Y)) = v(F_Y)$ , that is the expectation of the RIF at the  $\tau^{\text{th}}$  quantile is the value of the  $\tau^{\text{th}}$  quantile. Using notation from FFL, they define  $m_{\tau}(\mathbf{X}) = E(RIF(Y; \tau, F_Y)|\mathbf{X})$  as the unconditional quantile regression model (which is one case from their family of RIF regression models). The parameter estimates from the RIF regression model then provide estimates of the Unconditional Quantile Marginal Effect

(UQME), or in our example the marginal effect on BMI at the  $\tau^{\text{th}}$  quantile from small shift in a covariate (all else constant).<sup>16</sup>

### *3.3 Correlation between Income and BMI, Results*

Table 4 reports the regression coefficients from the OLS and UQR estimators of BMI on income, with controls for age, square of age, and indicator variables for race, and education levels. Table 5 replicates the estimates without the control variables.<sup>17</sup> In all of the analysis, income is measured relative to the poverty line and scaled to one (e.g. a value of two indicates that income is twice the poverty line). Appendix Figure 1 compares the income data from the pooled 1999-2006 NHANES with the income estimates from March Supplements of the Current Population Survey, both from 1999 and 2006. The overall shapes of the two distributions are similar, with similar central tendencies as well, though due most likely to sample size differences, the CPS density is smoother. For all regression estimates, the exam sample weights are used, allowing inferences to be drawn to the reference population (U.S. civilian, non-institutionalized population). The estimated variance-covariance matrix is adjusted to correct clustering.<sup>18</sup>

For the regression estimates, I pool together four cross sections of the 2-year cycle NHANES data, resulting in a sample from 1999 to 2006. While this crosses a long time period, it is useful to note that the early NHANES also had lengthy periods of field work (e.g. NHANES III is from

---

<sup>16</sup> FFL provide an estimation method based on transforming the dependent variable into the re-centered influence function and then using OLS estimation. FFL show that this approach yields a consistent estimator of the average marginal effect,  $E[d \Pr[Y > \tau | X]/dX]$ , if  $\Pr[Y > \tau | X=x]$  is linear in  $x$ .

<sup>17</sup> The Table 5 estimates are provided to map more directly with the discussion in the popular press, though they are perhaps less nuanced in their interpretation. The conditional correlation estimates in Table 4 are presented as providing a stronger case for the existence of an income gradient, existing over many subpopulations (e.g. by race, education, age groupings).

<sup>18</sup> The standard errors are corrected for the sample weights and clustering, but not for stratification. Given the general result that clustering reduces efficiency and stratification, if anything, increases efficiency, the reported standard errors are interpreted as slightly downward biased. Correcting for stratification will produce marginally smaller standard errors.

1988-94). The primary reason for this pooling though is that it produces sample sizes that are roughly similar across each of the decades. For example, in the Table 4 specification for the female subsample, the sample sizes range from 5,889 (NHANES II) to 7,604 (NHANES I) observations. The pooled 1999-2006 lies within this range. One purpose of presenting the regression estimates over the four decades is to examine the similarities and patterns over time. When sample sizes differ substantially, one must consider the extent to which observed differences across samples are driven by differences in power of the estimator. In using the same estimator, same specification, and similar sample sizes, one need be significantly less concerned about power differences across estimates.<sup>19</sup>

An additional reason for pooling the four 2-year cycles (1999-2006) is because we are estimating marginal effects in some cases on the extremes of the tails of the distribution. In particular, the prevalence of underweight in the U.S. has been approximately two percent over the last three decades. Estimating the UQME of income on BMI at the second percentile for one of the two-year cycles would mean estimating this quantile with about 3,500 observations (for the full sample, about half this when considering the sex-specific subsamples). By pooling the last four cycles of data, the sample size increases to 14,134.

[INSERT TABLES 4 AND 5 APPROXIMATELY HERE]

Panel A in Table 4 reports the income gradient in BMI for all adults between the ages of 20 and 75 from 1971-74 to 1999-2006 (conditioning on age, race, and education). The approach of using the OLS estimator shows a negative and statistically significant income gradient in BMI in

---

<sup>19</sup> Alternatively, it is also of interest to know what is happening recently, and for this purpose, pooling from 1999 to 2006 is limiting. To address this concern, tables are available from the author with the 1999-2006 sample separated into two separate samples – 1999-2002 and 2003-06. The qualitative story presented in Tables 4 and 5 is similar, though the point estimates differ somewhat.

all decades. The magnitude of the point estimate ranges from -0.09 to -0.22.<sup>20</sup> Recalling that income is measured in poverty line units (where an income of 1 means income equal to the poverty line), the 1999-2006 OLS estimates suggest a unit increase in income is associated with a reduction of BMI by approximately 0.22 during the last two decades.<sup>21</sup>

The UQR estimates in Panel A though, reveal that the OLS estimates are averaging over important differences across the distribution. At the underweight threshold, there is a *positive* and statistically significant income gradient in BMI in three of the four decades examined (the insignificant estimates are from the 1971-74 period). The 1999-2006 estimates indicate that a one unit increase in income is associated with an *increase* in BMI of 0.11 when evaluated at the underweight threshold. It is useful to note that this gradient is opposite in sign to the OLS estimate, but also consistent with the prior that presumed health improvements (increases in BMI at the underweight threshold) are associated with increases in income. The data also indicate that there is a negative income gradient in BMI at the overweight threshold in two of the four decades examined (1971-74 and 1988-94), But, the estimates from 1999-2006 indicate a relatively flat gradient. (Table 5 shows that the correlation coefficient is a precisely estimated zero.

Finally, at the obese threshold, there is a negative and statistically significant income gradient in BMI in all decades examined. The 1999-2006 estimates indicate that a one unit increase in income is associated with a decrease in BMI of 0.40 when evaluated at the quantile corresponding to the obesity threshold (BMI=30). It is useful to note that this gradient is essentially twice the magnitude of the OLS estimate, and during the 1970s is more than four times the magnitude of the corresponding OLS estimates.

An important measurement issue in these findings then is that because the OLS estimator

---

<sup>20</sup> For all estimates, the p-values are less than 0.1. When considering the simple bivariate regressions in Table 5, the point estimates fall between -0.14 and -0.17 with all p-values less than 0.01.

<sup>21</sup> The 2003-2006 pooled estimate (available from the author) suggests that this has dropped to 0.12.

assumes a constant gradient at all points on the conditional distribution, it mis-estimates the sign of the effect for the underweight and significantly underestimates the magnitude of the gradient for the obese. The findings for the estimates without the controls for age, race and education (Table 5) are quite similar. The OLS estimates suggest an income gradient that is about half the size of the UQR estimator when evaluated at the obesity threshold, and the OLS estimate are the wrong sign when estimating the effect at the quantile corresponding to the underweight threshold (BMI=18.5). There is little evidence of any statistically significant correlation between income and BMI (for either specification) when evaluated at the overweight threshold based on the 1999-2006 pooled data.

Given the significant gender differences in the relationship between BMI and income observed in Table 3, it is not that surprising that Tables 4 and 5 show differences in the income gradient by gender. For males, the OLS estimates suggest that there is a positive relationship between BMI and income. Without the control variables (Table 4), the correlation is weak; but in the model with controls (Table 5), the positive correlation is strong (p-values < 0.01) in three of the four periods. As with the full sample results, the OLS estimates are masking quite a bit of variation at different points in the BMI distribution. The UQR estimates show that the positive correlation is much larger than the OLS estimates when examined at BMI=25. Whereas the OLS estimate for 1999-2006 indicate an increase of 0.16 for a one unit increase in income, the UQR estimate evaluated at the overweight threshold is almost three times larger (0.45). The more recent UQR estimates further indicate that there is essentially no income gradient in BMI at either the underweight or obese threshold for men. This result is consistent with the findings from the alternative measures that there were large differences for men in terms of the prevalence of overweight, but essentially no differences in terms of the severity.

For women, the OLS – UQR comparison is more an issue of magnitude of the estimates.

Across all years, and for both models with and without controls, the OLS estimates of the income gradient are negative and statistically significant (p-values < 0.01). This is also true of the UQR estimates evaluated at both the overweight and obese thresholds (p-values < 0.01). This indicates a very robust finding consistent with the traditional income gradient in health (i.e. increases in income associated with better health outcomes). The primary difference is that the UQR estimates evaluated at the obesity threshold are much larger than the OLS estimates. For example, the obesity UQR estimate is about twice the size of the OLS estimate in the 1970s, and about 45 percent greater than OLS in the most recent estimates.<sup>22</sup>

#### **4. Conclusion**

Understanding the correlation between income and BMI is important for policies aimed at the nutritional intake of poor persons as well as policies aimed to reduce the prevalence of overweight and obesity. Current portrayals of the relationship between income and BMI in the popular press, policy briefings, and in some limited cases, academic writing, suggest that the poor have much higher rates of overweight and obesity. The basic descriptive statistics do not support this assertion. NHANES data from 2003-06 indicate no statistically significant differences in the prevalence of overweight or of obesity for the poor and nonpoor.

This paper examines continuous measures of overweight, and thereby helps to explain, in part, the conventional wisdom that the poor are more overweight. The choice to use these measures is based primarily on a desire to reflect research indicating that the severity (or probability) of negative health outcomes associated with being overweight are increasing in BMI. If one considers the overweight severity measures, which accounts for both the mass and spread of the BMI distribution above the overweight threshold, then the conventional wisdom

---

<sup>22</sup> The OLS estimates are about the same magnitude as the UQR estimates evaluated at BMI=25.

holds – the overweight severity for the poor is statistically greater than that for the nonpoor. The depth index helps to explain this by revealing that the overweight and poor exceed the overweight threshold (BMI=25) by 29 percent, while the nonpoor and overweight exceed this threshold by 26 percent.

The regression analysis provides further evidence that the relationship between BMI and income is more closely linked to conventional wisdom than is suggested by the cross-tabulations. While the paper shows that the OLS estimate of the marginal effect is negative at the conditional mean (consistent with the conventional wisdom), the UQR estimates show that OLS is masking important variation. In particular, OLS gets the sign wrong for the underweight, and significantly underestimates the magnitude of the negative correlation for the obese. When considering the adult population using recent data, the OLS estimate from regressing BMI on income indicates that a one unit increase in income (e.g. increasing income from the poverty line to twice this level) reduces the average BMI by .22 points. The UQR estimates reveal that at the underweight threshold (BMI=18.5), a one unit increase in income *increases* BMI by .11 points and at the obese threshold (BMI=30), the same increase in income *reduces* BMI by .40 points.

This paper provides evidence that the cross tabulation of overweight and poor provides a very incomplete picture of the relationship between income and BMI by indicating essentially no association between poverty and overweight and obesity status. Making policy decisions based on this though would lead to poor choices. The distribution-sensitive measures of overweight indicate that the overweight poor are the most overweight. The UQR estimates further suggest that there is a positive income gradient in BMI for the underweight and a negative gradient for the obese (much larger in magnitude than that estimated by OLS), which matches the standard health and wealth gradient, and conventional wisdom.



## REFERENCES

- Besharov, Douglas. "We're Feeding the Poor as if They're Starving." *Washington Post*, p. B02, December 8, 2002. [Online:] [www.welfareacademy.org/pubs/overfeedingthepoor.pdf](http://www.welfareacademy.org/pubs/overfeedingthepoor.pdf) [June 27, 2006]
- Besharov, Douglas. "Growing Overweight and Obesity in America: The Potential Role of Federal Nutrition Programs." Testimony Prepared for the Committee on Agriculture, Nutrition, and Forestry, U.S. Senate, April 3, 2003. [Online:] [www.welfareacademy.org/pubs/testimony-040303.pdf](http://www.welfareacademy.org/pubs/testimony-040303.pdf) [January 18, 2007]
- Case, Anne, Lubotsky, Darren and Paxson, Christina. "Economic Status and Health in Childhood: The Origins of the Gradient." *American Economic Review*, December 2002, 92(5): 1308-1334.
- Chamberlain, Gary. "Quantile Regression, Censoring, and the Structure of Wages" in *Advances in Econometrics*, Sims Christopher (ed.). New York: Elsevier, 1994.
- Critser, Greg. *Fat Land: How Americans Became the Fattest People in the World*, New York: Houghton Mifflin, 2003.
- Cutler, David, Glaeser, Edward and Shapiro, Jesse. "Why Have Americans Become More Obese?" *The Journal of Economic Perspectives*, Summer 2003, 17(3): 93-118.
- Deaton, Angus. "Relative Deprivation, Inequality, and Mortality." National Bureau of Economic Research Working Paper no. 8099, 2001.
- Deaton, Angus. "Policy Implications of the Gradient of Health and Wealth." *Health Affairs*, April 2002, 21(2)13-30.
- Deaton, Angus and Paxson, Christina. "Mortality, Education, Income, and Inequality among American Cohorts." National Bureau of Economic Research Working Paper no. 7140, 1999.
- Drewnowski, Adam and Specter, SE. "Poverty and obesity: the role of energy density and energy costs." *The American Journal of Clinical Nutrition*, January 2004, 79(1): 6-16.
- Filardo, Giovanni, Hamilton, Cody, Hamman, Baron and Grayburn, Paul. "Obesity and Stroke After Cardiac Surgery: The Impact of Grouping Body Mass Index." *The Annals of Thoracic Surgery*, 2007, 84(3): 720-722.
- Firpo, Sergio, Fortin, Nicole, and Lemieux, Thomas. "Unconditional Quantile Regression." *Econometrica*, 2009, 77(3): 953-73.
- Firpo, Sergio, Fortin, Nicole, and Lemieux, Thomas. "Decomposing Wage Distributions using Recentered Influence Function Regressions," 2009, University of British Columbia.

- Flegal KM, Carroll MD, Kuczmarski RJ, Johnson CL. "Overweight and Obesity in the United States: Prevalence and Trends, 1960-1994." *International Journal of Obesity*, January 1998, 22(1): 39-47.
- Foster, James; Greer, Joel and Thorbecke, Erik. "A Class of Decomposable Poverty Measures." *Econometrica*, May 1984, 52(3): 761-765.
- Freedman, David, Khan, Laura, Serdula, Mary, Galuska, Deborah and Dietz, William. "Trends and Correlates of Class 3 Obesity in the United States From 1990 Through 2000." *Journal of the American Medical Association*, 2002, 288(14): 1758-1761.
- Hampel, Frank; Ronchetti, Elvezio; Rousseeuw, Peter; and Stahel, Werner. 1988. *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley Series in Probability and Statistics, 1988.
- Jolliffe, Dean. "Continuous and Robust Measures of the Overweight Epidemic from 1971-2000." *Demography*, May 2004, 41(2): 303-314.
- Jolliffe, Dean and Semykina, Anastassia. "Robust Standard Errors for the Foster-Greer-Thorbecke Class of Poverty Indices: SEPOV." 1999. *Stata Technical Bulletin*, STB-51.
- Kenchaiah, Satish, Evans, Jane, Levy, Daniel, Wilson, Peter, Benjamin, Emelia, Larson, Martin, Kannel, William and Vasan, Ramachandran. "Obesity and the Risk of Heart Failure." *New England Journal of Medicine*, August 2002, 347(5): 305-313.
- Kish, Leslie, *Survey Sampling*, New York: John Wiley & Sons, 1965.
- Koenker, Roger. *Regression Quantiles*, New York: Cambridge University Press, 2005.
- Koenker, Roger and Bassett, Gilbert Jr. "Robust Tests for Heteroscedasticity Based on Regression Quantiles." *Econometrica*, January 1982, 50 (1): 43-61.
- Kurth T, Gaziano M, Berger K, Kase C, Rexrode K, Cook N, Buring J, Manson J. "Body Mass Index and the Risk of Stroke in Men." *Archives of Internal Medicine*, December 2002, 162(22): 2557-2562.
- Lakdawalla, Darius and Philipson, Tomas. "The Growth of Obesity and Technological Change: A Theoretical and Empirical Examination." National Bureau of Economic Research, NBER Working Paper no. 8946, 2002.
- Lakdawalla, Darius and Philipson, Tomas. "The Growth of Obesity and Technological Change." *Economics and Human Biology* 2009, 7(3): 283-293.
- Lakdawalla, Darius, Philipson, Tomas and Bhattacharya, Jay. "Welfare-Enhancing Technological Change and the Growth of Obesity." *American Economic Review: AEA Papers and Proceedings*, 2006, 95(2): 253-257.

- National Center for Health Statistics. *Health: United States 2001 with Urban and Rural Health Chartbook*. Hyattsville, MD: Center for Disease Control and Prevention, 2001. [Online:] <http://www.cdc.gov/nchs/data/hus/hus01cht.pdf> [September 22, 2010].
- National Heart, Lung, and Blood Institute. *Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults: The Evidence Report*, NIH Publication No. 98-4083, Rockville, MD: National Institutes of Health, September 1998.
- Nguyen, Binh, Albrecht, James, Vroman, Susan and Westbrook, M. Daniel. "A Quantile Regression Decomposition of Urban-rural Inequality in Vietnam." *Journal of Development Economics*, 2007, 83(2): 466-490.
- Ogden, Cynthia, Carroll, Margaret, Curtin, Lester, McDowell, Margaret, Tabak, Carolyn, Flegal, Katherine. "Prevalence of Overweight and Obesity in the United States, 1999-2004." *Journal of the American Medical Association*, April 2006, 295(13): 1549-55.
- Pappas, Gregory, Susan Queen, Wilbur Hadden and Gail Fisher. "The Increasing Disparity in Mortality Between Socioeconomic Groups in the United States, 1960 and 1986," *New England Journal of Medicine*, 1993, 329(2): 103-109.
- Patrinos, Harry and Sakellariou, Chris. "Economic Volatility and Returns to Education in Venezuela: 1992-2002." *Applied Economics*, September 2006, 38(17): 1991-2005.
- Scott, A.J. and Holt, D. "The Effect of Two-Stage Sampling on Ordinary Least Squares Methods." *Journal of American Statistical Association*, 1982, 77(380): 848-854.
- Solon, Gary. "Intergenerational Income Mobility in the United States." *American Economic Review*, June 1992, 82(3): 393-408.
- Sorlie, Paul D., Eric Backlund and Jacob Keller. "US Mortality by Economic, Demographic and Social Characteristics: The National Longitudinal Mortality Study," *American Journal of Public Health*, 1995, 85(7): 949-956.
- Stifel, David C., and Averett, Susan L. "Childhood Overweight in the United States: A Quantile Regression Approach," *Economics and Human Biology*, 2009, 7(3): 387-397.
- USDA Economic Research Service. *The Food Assistance Landscape: FY 2008 Annual Report*. ERS Economic Information Bulletin no. 6-6, Washington DC: USDA, April 2009.
- WHO Expert Committee. *Physical Status: The Use and Interpretation of Anthropometry*, WHO Technical Report Series, No. 854, Geneva: World Health Organization, 1995.
- Willett W.C., Dietz W.H., Colditz G.A. "Guidelines for Healthy Weight." *New England Journal of Medicine*, August 1999, 341(6): 427-434.

Table 1: Overweight and Obese by Income

$OW_{\alpha}$ Indices of Overweight	1971-74	1976 -80	1988 -94	1999 -02	2003 -06	% Change 1971-2006
<i>Panel A: Overweight, BMI &gt; 25</i>						
Overweight, Poor	48.3 (1.54)	46.5 (1.40)	55.7 (1.40)	64.4 (1.41)	64.7 (1.53)	28%
Overweight, Not Poor	46.5 (0.77)	45.9 (0.84)	53.9 (1.09)	65.4 (1.03)	67.5 (1.09)	45%
Difference : (% points)	1.8	0.6	1.8	-1.0	-2.8	
<i>Panel B: Obese, BMI &gt; 30</i>						
Obese, Poor	19.8 (1.09)	19.0 (1.02)	26.8 (1.38)	34.6 (1.28)	34.2 (1.23)	62 %
Obese, Not Poor	13.3 (0.55)	13.5 (0.41)	21.3 (0.87)	29.5 (1.12)	33.9 (1.17)	155%
Difference: (% points)	6.5***	5.5***	5.5***	5.1*	0.3	

*Note:* For all analysis in this paper, poor is defined as less than or equal to 130% of the poverty line. Statistical significance indicated with \*, \*\*, or \*\*\* for p-values less than 0.1, 0.05, and 0.01, respectively. Standard errors corrected for complex sample design using the NHANES pseudo design variables. All years exclude pregnant and breastfeeding women.

Table 2: Adult Overweight by Income

$OW_{\alpha}$	Indices of Overweight	1971 - 1974	1976 - 1980	1988 - 1994	1999 - 2002	2003 - 2006	% Change 1971-2006
<i>Panel A: Poor (&lt; 130% poverty line)</i>							
$OW_0$	Prevalence	48.3 (1.54)	46.5 (1.40)	55.7 (1.40)	64.4 (1.41)	64.7 (1.53)	34%
$OW_1$	Depth	10.4 (0.46)	10.2 (0.48)	13.6 (0.68)	18.8 (0.68)	18.5 (0.71)	77%
$OW_2$	Severity	4.4 (0.35)	4.3 (0.39)	6.1 (0.51)	9.7 (0.62)	10.3 (1.06)	135%
<i>Panel B: Nonpoor (<math>\geq 130\%</math> poverty line)</i>							
$OW_0$	Prevalence	46.5 (0.77)	45.9 (0.84)	53.9 (1.09)	65.4 (1.03)	67.5 (1.09)	45%
$OW_1$	Depth	7.5 (0.22)	7.4 (0.18)	11.1 (0.43)	15.6 (0.55)	17.5 (0.53)	133%
$OW_2$	Severity	2.3 (0.14)	2.3 (0.10)	4.4 (0.33)	6.6 (0.41)	7.9 (0.37)	237%

*Note:* All overweight measures are multiplied by 100. Standard errors for  $OW_0$ ,  $OW_1$ , and  $OW_2$  are also multiplied by 100 and in parentheses. See notes for Table 1.

Table 3: Adult Overweight by Sex and Income

$OW_{\alpha}$	Indices of Overweight	1971 - 1974	1976 - 1980	1988 - 1994	1999 - 2002	2003 - 2006	% Change 1971-2006
<i>Panel A: Poor Men (&lt;130% poverty)</i>							
$OW_0$	Prevalence	43.8 (2.18)	42.8 (1.99)	51.5 (1.97)	63 (2.05)	62.7 (2.30)	43%
$OW_1$	Depth	7.0 (0.5)	6.8 (0.44)	10.2 (0.68)	15.6 (0.92)	13.9 (1.00)	98%
$OW_2$	Severity	2.2 (0.30)	2.1 (0.27)	3.6 (0.48)	7.5 (0.93)	7.6 (2.25)	252%
<i>Panel B: Nonpoor Men (<math>\geq 130\%</math> poverty)</i>							
$OW_0$	Prevalence	54.9 (1.20)	52.4 (0.92)	60.4 (1.29)	69.6 (1.05)	74.5 (1.21)	36%
$OW_1$	Depth	7.5 (0.28)	7 (0.21)	10.6 (0.44)	14.5 (0.56)	17.2 (0.55)	128%
$OW_2$	Severity	2.0 (0.19)	1.7 (0.08)	3.6 (0.34)	5.5 (0.47)	6.8 (0.32)	246%
<i>Panel C: Poor Women (&lt;130% poverty)</i>							
$OW_0$	Prevalence	51.7 (1.94)	49 (1.75)	59 (1.91)	65.5 (1.92)	66.4 (1.71)	29%
$OW_1$	Depth	12.9 (0.64)	12.5 (0.79)	16.4 (0.94)	21.3 (0.81)	22.3 (0.93)	73%
$OW_2$	Severity	6.0 (0.51)	5.8 (0.67)	8.0 (0.68)	11.4 (0.77)	12.6 (0.89)	108%
<i>Panel D: Nonpoor Women (<math>\geq 130\%</math> poverty)</i>							
$OW_0$	Prevalence	38.3 (0.89)	39.3 (1.12)	47.3 (1.29)	60.8 (1.60)	60.2 (1.54)	57%
$OW_1$	Depth	7.5 (0.33)	7.8 (0.32)	11.6 (0.59)	16.7 (0.70)	17.8 (0.67)	138%
$OW_2$	Severity	2.7 (0.21)	3.0 (0.21)	5.1 (0.47)	7.8 (0.52)	9.0 (0.57)	234%

Note: See note for Table 2.

Table 4: BMI and Income

## Multivariate OLS and Unconditional Quantile Regression Estimates

Income gradients, OLS and UQR estimators	1971	1976	1988	1999
	-74	-80	-94	2006
<i>Panel A: All Adults, BMI</i>				
OLS	<b>-.10***</b> (.02)	-.09* (.05)	<b>-.20**</b> (.08)	<b>-.22***</b> (.03)
Quantile: Underweight	<b>.06</b> (.05)	<b>.11***</b> (.02)	<b>.02***</b> (.00)	<b>.11**</b> (.04)
Quantile: Overweight	<b>-.10**</b> (.05)	.01 (.03)	<b>-.14**</b> (.06)	-.02 (.03)
Quantile: Obese	<b>-.44***</b> (.09)	<b>-.63***</b> (.10)	<b>-.40**</b> (.17)	<b>-.40***</b> (.04)
<i>Sample size</i>	<i>11,994</i>	<i>11,005</i>	<i>13,208</i>	<i>14,123</i>
<i>Panel B: Males, BMI</i>				
OLS	.10* (.05)	.08 (.07)	-.06 (.08)	.07* (.04)
Quantile: Underweight	.11 (.12)	.07 (.11)	.03** (.02)	.13 (.15)
Quantile: Overweight	<b>.13**</b> (.05)	<b>.22***</b> (.03)	-.05 (.08)	<b>.29***</b> (.06)
Quantile: Obese	-.17 (.23)	<b>-.33***</b> (.07)	-.24 (.15)	-.01 (.07)
<i>Sample size</i>	<i>4,632</i>	<i>5,276</i>	<i>6,351</i>	<i>7,213</i>
<i>Panel C: Females, BMI</i>				
OLS	<b>-.29***</b> (.03)	<b>-.28***</b> (.03)	<b>-.33***</b> (.07)	<b>-.48***</b> (.02)
Quantile: Underweight	-.04 (.05)	<b>.11***</b> (.01)	<b>-.03***</b> (.01)	.05 (.06)
Quantile: Overweight	<b>-.37***</b> (.10)	<b>-.28***</b> (.05)	<b>-.27***</b> (.06)	<b>-.36***</b> (.11)
Quantile: Obese	<b>-.68***</b> (.13)	<b>-.68***</b> (.08)	<b>-.52***</b> (.18)	<b>-.69***</b> (.02)
<i>Sample size</i>	<i>7,362</i>	<i>5,729</i>	<i>6,857</i>	<i>6,910</i>

*Note:* Weighted OLS and Unconditional Quantile Regressions of BMI on income, age, square of age, race and education levels. OLS standard errors are corrected for stratification and clustering. UQR standard errors are bootstrap estimates (reps=1,000), accounting for clustering. Bold indicates p-value < 0.05; \* p-value < 0.1, \*\* p-value < 0.05, \*\*\* p-value < 0.01.

Table 5: BMI and Income

## Bivariate OLS and Unconditional Quantile Regression Estimates

Income gradients, OLS and UQR estimators	1971	1976	1988	1999 2006
<i>Panel A: All Adults, BMI</i>				
OLS	<b>-.14***</b> (.02)	<b>-.15***</b> (.03)	<b>-.17***</b> (.06)	<b>-.17***</b> (.05)
Quantile: Underweight	<b>.09**</b> (.03)	<b>.12***</b> (.02)	.04 (.05)	<b>.12**</b> (.05)
Quantile: Overweight	<b>-.11**</b> (.05)	<b>-.05***</b> (.01)	-.09 (.06)	.00 (.03)
Quantile: Obese	<b>-.65***</b> (.10)	<b>-.81***</b> (.07)	<b>-.40**</b> (.16)	<b>-.33***</b> (.09)
<i>Sample size</i>	<i>12,397</i>	<i>11,295</i>	<i>13,274</i>	<i>14,134</i>
<i>Panel B: Males, BMI</i>				
OLS	<b>.13**</b> (.05)	<b>.13***</b> (.05)	.05 (.06)	<b>.16***</b> (.03)
Quantile: Underweight	.11 (.11)	.11 (.10)	.04 (.03)	.14 (.12)
Quantile: Overweight	<b>.26***</b> (.05)	<b>.30***</b> (.02)	<b>.15**</b> (.06)	<b>.45***</b> (.08)
Quantile: Obese	-.25 (.22)	<b>-.32***</b> (.05)	-.11 (.14)	.11 (.09)
<i>Sample size</i>	<i>4,793</i>	<i>5,406</i>	<i>6,392</i>	<i>7,219</i>
<i>Panel C: Females, BMI</i>				
OLS	<b>-.43***</b> (.03)	<b>-.45***</b> (.03)	<b>-.39***</b> (.07)	<b>-.47***</b> (.06)
Quantile: Underweight	.01 (.03)	<b>.11***</b> (.02)	-.01 (.04)	.07 (.06)
Quantile: Overweight	<b>-.57***</b> (.11)	<b>-.50***</b> (.03)	<b>-.38***</b> (.08)	<b>-.44***</b> (.07)
Quantile: Obese	<b>-1.00***</b> (.12)	<b>-1.11***</b> (.04)	<b>-.66***</b> (.17)	<b>-.70***</b> (.05)
<i>Sample size</i>	<i>7,604</i>	<i>5,889</i>	<i>6,882</i>	<i>6,915</i>

*Note:* Weighted OLS and Unconditional Quantile Regressions of BMI on income. OLS standard errors are corrected for stratification and clustering. UQR standard errors are bootstrap estimates (reps=1,000), accounting for clustering. Bold indicates p-value < 0.05; \* p-value < 0.1, \*\* p-value < 0.05, \*\*\* p-value < 0.01.



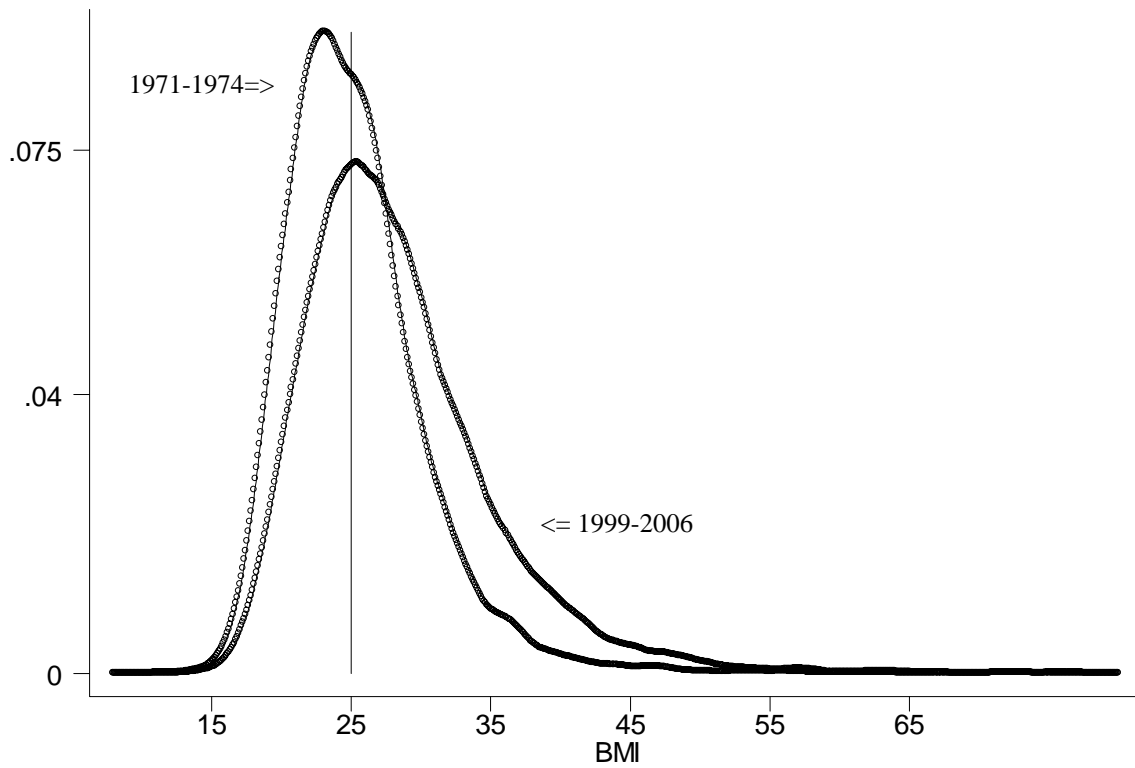


Figure 1: BMI density from 1971-1974 and 1999-2006

*Notes:* The Epanechnikov kernel is used to estimate the density functions with the smoothing parameter set to 0.75.

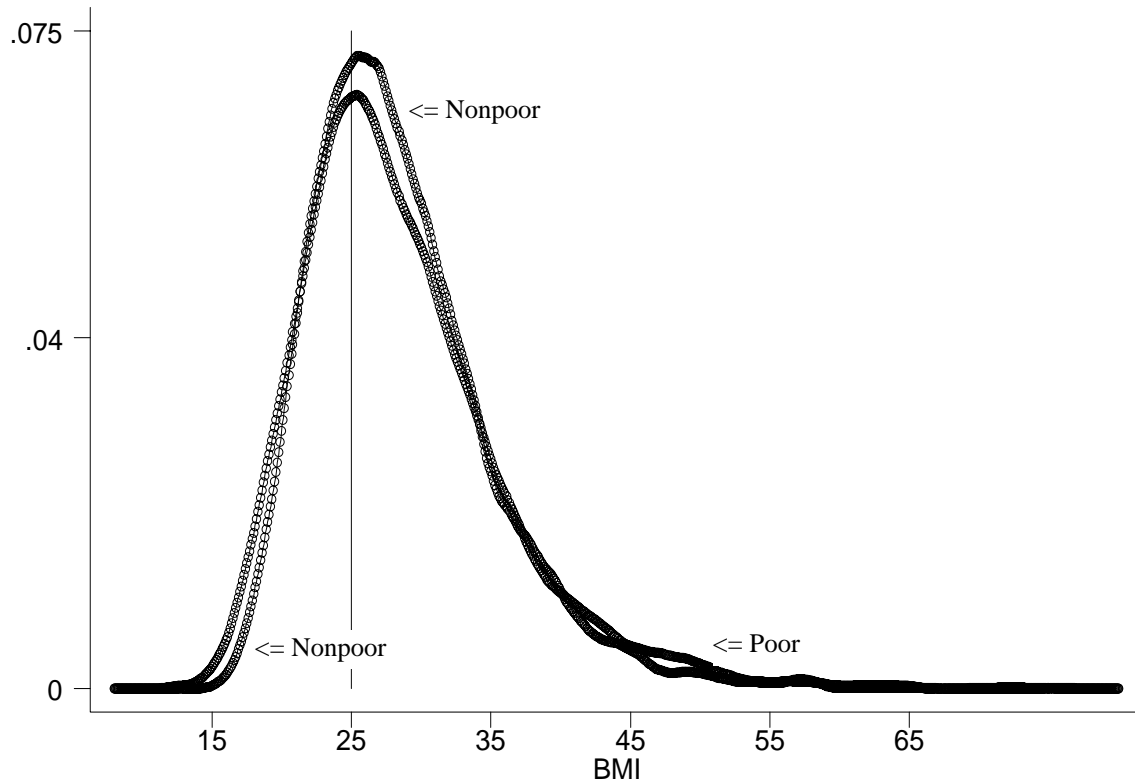
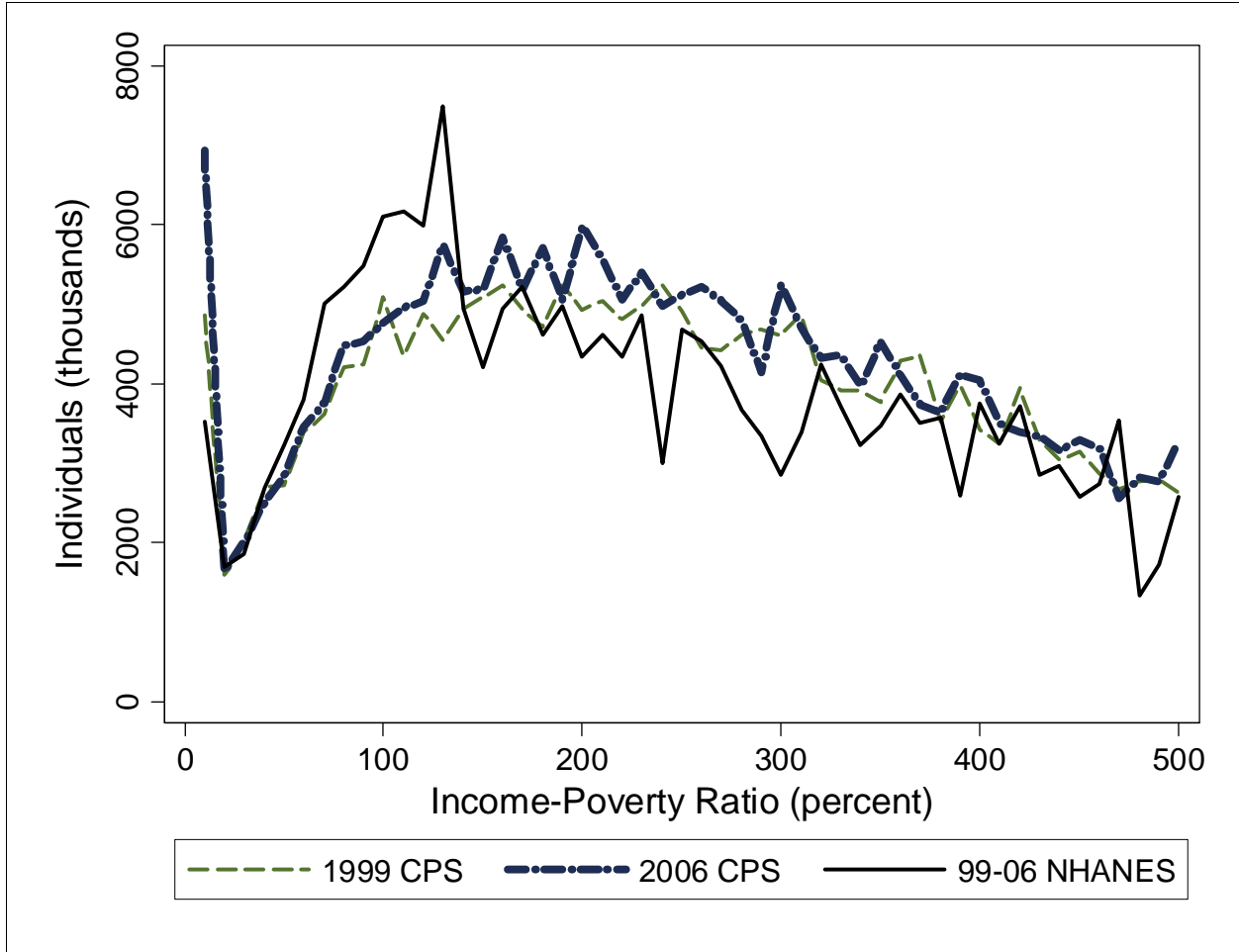


Figure 2: BMI density in 1999-2006 by poverty status

*Notes:* Poor persons are those with incomes less than 130 percent of the poverty line, nonpoor are those with incomes greater than 130 percent. The Epanechnikov kernel is used to estimate the density functions with the smoothing parameter set to 0.75 for the nonpoor and 1.5 for the poor.

**Appendix Figure 1: Income Estimates, Comparing NHANES to CPS**



*Sources:* NHANES 1999-2000, 2001-02, 2003-04, 2005-06 and Current Population Survey 1999 and 2006 (March Supplements)

Appendix Table 1: BMI and Income

Bivariate OLS and Unconditional Quantile Regression Estimates

Income gradients, OLS and UQR estimators	1971	1976	1988	1999	2003
<i>Panel A: All Adults, BMI</i>					
OLS	<b>-.14</b> (.02)	<b>-.15</b> (.03)	<b>-.17</b> (.06)	<b>-.23</b> (.03)	-.11 (.11)
Quantile: Underweight	<b>.09</b> (.03)	<b>.12</b> (.02)	.04 (.05)	<b>.14</b> (.05)	.09 (.11)
Quantile: Overweight	<b>-.11</b> (.05)	<b>-.05</b> (.01)	-.09 (.06)	<b>-.05</b> (.03)	.04 (.04)
Quantile: Obese	<b>-.65</b> (.10)	<b>-.81</b> (.07)	<b>-.40</b> (.16)	<b>-.37</b> (.10)	<b>-.28</b> (.07)
<i>Sample size</i>	<i>12,397</i>	<i>11,295</i>	<i>13,274</i>	<i>6,816</i>	<i>7,318</i>
<i>Panel B: Males, BMI</i>					
OLS	<b>.13</b> (.05)	<b>.13</b> (.05)	.05 (.06)	.08 (.05)	<b>.24</b> (.02)
Quantile: Underweight	.11 (.11)	.11 (.10)	.04 (.03)	.13 (.09)	.15 (.17)
Quantile: Overweight	<b>.26</b> (.05)	<b>.30</b> (.02)	<b>.15</b> (.06)	<b>.34</b> (.07)	<b>.56</b> (.07)
Quantile: Obese	-.25 (.22)	<b>-.32</b> (.05)	-.11 (.14)	.03 (.12)	<b>.18</b> (.01)
<i>Sample size</i>	<i>4,793</i>	<i>5,406</i>	<i>6,392</i>	<i>3,468</i>	<i>3,751</i>
<i>Panel C: Females, BMI</i>					
OLS	<b>-.43</b> (.03)	<b>-.45</b> (.03)	<b>-.39</b> (.07)	<b>-.50</b> (.05)	<b>-.44</b> (.19)
Quantile: Underweight	.01 (.03)	<b>.11</b> (.02)	-.01 (.04)	.10 (.06)	.03 (.03)
Quantile: Overweight	<b>-.57</b> (.11)	<b>-.50</b> (.03)	<b>-.38</b> (.08)	<b>-.43</b> (.09)	<b>-.47</b> (.02)
Quantile: Obese	<b>-1.00</b> (.12)	<b>-1.11</b> (.04)	<b>-.66</b> (.17)	<b>-.73</b> (.12)	<b>-.68</b> (.15)
<i>Sample size</i>	<i>7,604</i>	<i>5,889</i>	<i>6,882</i>	<i>3,348</i>	<i>3,567</i>

*Note:* Weighted OLS and Unconditional Quantile Regressions of BMI on income. OLS standard errors are corrected for stratification and clustering. UQR standard errors are bootstrap estimates (reps=1,000), accounting for clustering. Bold indicates p-value < 0.05.

Appendix Table 2: BMI and Income

## Multivariate OLS and Unconditional Quantile Regression Estimates

Income gradients, OLS and UQR estimators	1971	1976	1988	1999	2003
<i>Panel A: All Adults, BMI</i>					
OLS	<b>-.10</b> (.02)	-.09 (.05)	<b>-.20</b> (.08)	<b>-.32</b> (.03)	<b>-.12</b> (.05)
Quantile: Underweight	.06 (.05)	<b>.11</b> (.02)	<b>.02</b> (.00)	<b>.13</b> (.04)	.08 (.06)
Quantile: Overweight	<b>-.10</b> (.05)	.01 (.03)	<b>-.14</b> (.06)	<b>-.08</b> (.03)	<b>.03</b> (.01)
Quantile: Obese	<b>-.44</b> (.09)	<b>-.63</b> (.10)	<b>-.40</b> (.17)	<b>-.50</b> (.04)	<b>-.27</b> (.03)
<i>Sample size</i>	<i>11,994</i>	<i>11,005</i>	<i>13,208</i>	<i>6,806</i>	<i>7,317</i>
<i>Panel B: Males, BMI</i>					
OLS	.10 (.05)	.08 (.07)	-.06 (.08)	-.05 (.09)	<b>.19</b> (.01)
Quantile: Underweight	.11 (.12)	.07 (.11)	<b>.03</b> (.02)	.12 (.07)	.13 (.19)
Quantile: Overweight	<b>.13</b> (.05)	<b>.22</b> (.03)	-.05 (.08)	<b>.15</b> (.04)	<b>.42</b> (.07)
Quantile: Obese	-.17 (.23)	<b>-.33</b> (.07)	-.24 (.18)	-.17 (.09)	<b>.12</b> (.02)
<i>Sample size</i>	<i>4,632</i>	<i>5,276</i>	<i>6,351</i>	<i>3,463</i>	<i>3,750</i>
<i>Panel C: Females, BMI</i>					
OLS	<b>-.29</b> (.03)	<b>-.28</b> (.03)	<b>-.33</b> (.07)	<b>-.56</b> (.02)	<b>-.41</b> (.08)
Quantile: Underweight	-.04 (.05)	<b>.11</b> (.01)	<b>-.03</b> (.01)	.06 (.07)	.01 (.05)
Quantile: Overweight	<b>-.37</b> (.10)	<b>-.28</b> (.05)	<b>-.27</b> (.06)	<b>-.31</b> (.10)	<b>-.40</b> (.06)
Quantile: Obese	<b>-.68</b> (.13)	<b>-.68</b> (.08)	<b>-.52</b> (.13)	<b>-.81</b> (.09)	<b>-.59</b> (.04)
<i>Sample size</i>	<i>7,362</i>	<i>5,729</i>	<i>6,857</i>	<i>3,343</i>	<i>3,567</i>

*Note:* Weighted OLS and Unconditional Quantile Regressions of BMI on income, age, square of age, race and education levels. OLS standard errors are corrected for stratification and clustering. UQR standard errors are bootstrap estimates (reps=1,000), accounting for clustering. Bold indicates p-value < 0.05.