Working Papers

FEDERAL RESERVE BANK OF BOSTON

No. 11-2

Explaining Gender-Specific Racial Differences in Obesity Using Biased Self-Reports of Food Intake and Physical Activity

Mary A. Burke and Frank W. Heiland

Abstract:

Policymakers have an interest in identifying the differences in behavior patterns—namely, habitual caloric intake and physical activity levels—that contribute to demographic variation in body mass index (BMI) and obesity risk. While disparities in mean BMI and obesity rates between whites (non-Hispanic) and African-Americans (non-Hispanic) are well-documented, the behavioral differences that underlie these gaps have not been carefully identified. Moreover, the female-specificity of the black-white obesity gap has received relatively little attention. In the National Health and Nutrition Examination Surveys (NHANES) data, we initially observe a very weak relationship between self-reported measures of caloric intake and physical activity and either BMI or obesity risk, and these behaviors appear to explain only a small fraction of the black-white BMI gap (or obesity gap) among women. These unadjusted estimates echo previous findings from large survey datasets such as the NHANES. Using an innovative method to mitigate the widely recognized problem of measurement error in self-reported behaviors—proxying for measurement errors using the ratio of reported caloric intake to estimated true caloric needs—we obtain much stronger relationships between behaviors and BMI (or obesity risk). Behaviors can in fact account for a significant share of the BMI gap (and the obesity gap) between black women and white women and are consistent with the presence of much smaller gaps between black men and white men. The analysis also shows that the effects smoking has on BMI and obesity risk are small-to-negligible when measurement error is properly controlled.

JEL Classifications: I12, J15, J16

Mary A. Burke is a senior economist at the Federal Reserve Bank of Boston. Her e-mail address is mary.burke@bos.frb.org. Frank W. Heiland is an assistant professor of economics in the School of Public Affairs, Baruch College/CUNY. His e-mail address is frank.heiland@baruch.cuny.edu. The authors are grateful for the comments of seminar and session participants at the Brookings Institution, Florida State University, Clark University, the Southern Economic Association meetings, the Western Economic Association International meetings, and the North American Econometric Society summer meetings. In particular, we want to thank Peyton Young, Carol Graham, Robin Simon, and Chris Foote for helpful discussions and suggestions. For research assistance we thank Kevin Todd, Carl Nadler, and the staff of the research library at the Federal Reserve Bank of Boston. All errors are our own.

This paper, which may be revised, is available on the web site of the Federal Reserve Bank of Boston at http://www.bos.frb.org/economic/wp/index.htm. **This version: May 19, 2011**

1 Introduction

Since at least the late 1970s, obesity has been significantly more prevalent among non-Hispanic African-American women (henceforth "black women") than among non-Hispanic white American women (henceforth "white women").[1] Table 1 gives age-adjusted obesity rates for white women and black women, ages 20 to 74 years, over various time intervals between 1976 and 2006. These rates are based on the National Health and Nutrition Examination Surveys, hereafter NHANES.[2] In the 1976–1980 data, the obesity rate among black women exceeds that of white women by 15.6 percentage points. This gap is roughly unchanged in the 1988–1994 data (15.8 points) and becomes even greater in subsequent periods, reaching a maximum of 21.8 percentage points in the 2003–2006 data. In contrast, black men do not consistently exhibit higher obesity rates than white men. In periods in which black men's obesity rate exceeds that of white men by a significant margin, the gap in rates falls far short of the difference observed among women—witness, for example, the 3.3 percentage-point gap among men in the 2003–2006 data, versus the 21.8 percentage-point gap among women during the same period.

A recent editorial in the *The American Journal of Clinical Nutrition* draws attention to black women's higher obesity rate in relation to other sociodemographic groups, and argues that a greater understanding of disparities in obesity risk is critical for the design of appropriate policy interventions (Ogden 2009). While policymakers and researchers from a broad array of disciplines are well aware of differences in obesity prevalence and mean body mass index (BMI) by race and sex, systematic research into the causes of these patterns has been relatively scarce and the facts underlying the disparities remain poorly understood.

In this paper, using data from the NHANES 1999–2006 surveys, we perform a gender-specific multivariate analysis of obesity and BMI in order to determine the extent to which variation in relevant behaviors—including food intake, physical activity, and smoking—contribute to gender-

[1] The definition of obesity employed by the Centers for Disease Control (CDC) is a body mass index (BMI) value of 30 or greater. BMI is the ratio of weight, measured in kilograms, to squared height, measured in meters.

[2] The data are based on NHANES II, which covers 1976–1980, NHANES III (1988–1994), and NHANES 1999–2004. Age-adjusted data by race and sex are not available in surveys prior to NHANES II.

specific racial differences in outcomes. The analysis incorporates a new method for addressing measurement error in self-reported behaviors and is informed by a conceptual model of BMI based on biological and economic principles. To our knowledge, no previous paper has identified the contribution of behavior patterns categorized along racial and gender lines to the outcomes of interest using a large, nationally representative dataset such as NHANES. Although differences in behaviors are merely proximate to differences in BMI and obesity—and causal links from food intake and physical activity to BMI have been well established in previous studies—identifying the behavioral patterns that bear on the racial and gender specificity of obesity helps point the way toward appropriately targeted policy interventions.

NHANES data are advantageous in that weight and height values (and therefore BMI) are measured by trained examiners rather than self-reported, and hence not subject to systematic bias. The surveys also contain extensive data on food intake and physical activity patterns, although these data are largely self-reported and subject to systematic biases in addition to random noise.[3] To a greater extent than previous studies that make use of similar data, we employ rigorous methods to mitigate measurement error in food intake and physical activity. These innovations turn out to be quite important: after eliminating highly implausible observations and proxying for remaining measurement error with a continuous index of data validity, we achieve large increases in the explanatory power of caloric intake and physical activity and, hence, in the ability of these factors to account for gender-specific racial differences in mean BMI and obesity.

Among women, we find that lower levels of physical activity in both work-related and leisure-time pursuits contribute significantly to black women's higher mean BMI and greater obesity prevalence. While higher caloric intake predicts higher BMI and increased obesity risk among women (as expected), the difference in mean caloric intake between black women and white women is small and imprecisely measured, suggesting that dietary intake may not contribute much to racial disparities in BMI and obesity. Our findings related to measurement error indicate, however, that

[3] Accelerometer-based measures of physical activity are available for some subjects in NHANES 2003–2006; these also are subject to measurement error based on user compliance, in addition to potential sample-selection bias. See, for example, Beyler et al. 2008.

caloric intake patterns likely play a greater role than that implied by racial differences in self-reported intakes, since we present evidence that black women underreport calories to a greater extent than white women.

Black men report lower caloric intake than white men, behavior that, all else equal, would predict lower BMI and obesity risk compared to white men. However, black men's greater tendency to be sedentary has a partly offsetting effect. Furthermore, black men appear to underreport calories to a greater extent than white men, and this disparity helps to reconcile the existence of a small but positive black-versus-white BMI gap among men in the uncontrolled data.

Smoking has been negatively associated with BMI in previous studies (Perkins 1993 and Chou, Grossman, and Saffer 2004, among many others). In the NHANES 1999–2006 data, black women are significantly less likely to smoke than white women, while black men are more apt to smoke than are white men. However, we find that smoking patterns contribute at best only marginally to race-by-gender patterns in BMI (or obesity) when measurement error in caloric intake and physical activity are mitigated. This result occurs because smoking acts on BMI (if at all) only indirectly, via its influences on caloric intake and/or caloric expenditure. While the measured effects of smoking on BMI and obesity in our models are more robust among men than among women, smoking's contribution to racial patterns in BMI among men is physiologically, if not statistically, insignificant.

Our findings cannot rule out the possibility that variation in unobserved biological endowments—such as basal metabolism—contribute to the large racial disparity in female outcomes. However, our findings suggest that the contribution of such factors declines when measurement error in behaviors is mitigated. In addition, racial differences in such endowments would have to be more pronounced among women, or even specific to women, in order to explain the residual differences in BMI and obesity risk by race and sex.

Numerous studies across various disciplines have documented the presence of racial disparities in mean BMI and obesity prevalence in the United States over various (mostly fairly recent) time periods.[4] Some of these have pointed out the gender specificity of the black-white patterns in particular, such as Burkhauser and Cawley (2008). Other studies (Kumanyika 1987) have offered numerous hypotheses to explain black women's excess obesity prevalence, yet do not come to any firm conclusions and do not address the problem of gender specificity. A number of studies have investigated the contribution of socioeconomic status to racial and/or race-by-gender patterns in BMI and/or obesity risk, and have concluded, nearly uniformly, that the demographics of obesity are highly robust to controls for household income, educational attainment, and other indicators of status.[5] To our knowledge, no previous study has undertaken a systematic analysis of how behaviors (including food intake, physical activity, and smoking) contribute to mean BMI and obesity rates categorized by race and gender using a large nationally representative dataset.

Rashad (2006) also uses self-reported data on caloric intake and physical activity from NHANES surveys (waves I, II, and III) to estimate a structural model of BMI determination. Her main objective is to identify the contributions to BMI of exogenous variation in food intake and cigarette smoking, as predicted by instruments such as restaurant prices and cigarette taxes. She acknowledges potential measurement error in self-reported caloric intake and attempts to mitigate it using a simple heuristic strategy. In gender-specific, ordinary-least-squares (OLS) models that use (adjusted) caloric intake measures directly, the estimated effects of caloric intake on BMI are an order of magnitude smaller than our own estimates, suggesting possible attenuation bias.[6] The use of instruments might in principle be expected to mitigate attenuation bias, but the effects of caloric intake on BMI are in most cases smaller—and have lower or no significance—in the instrumental variables (IV) models compared to OLS models.[7] These results highlight the need for a more

[4]An early report documenting excess weight by race and gender was published by the Bureau of State Services in 1954. Kuczmarski et al. (1994) document trends in obesity prevalence and mean BMI by race and sex for NHANES surveys spanning the years 1960–1991. Flegal et al. (1998) update these trends for the NHANES III (1988–1994) survey, and Ogden et al. (2006) describe the trends for NHANES 1999–2004. Komlos and Brabec (2010) examine obesity trends by birth cohort and race for the period 1882–1986.

[5]See, for example, Burke and Heiland (2008), Mujahid et al. (2005), Chang and Lauderdale (2005), Chou, Grossman, and Saffer (2004), Denney et al. (2004), and Zhang and Wang (2004).

[6]Rashad's caloric intake measure is adjusted for measurement error and also adjusted for an estimate of the physical activity level.

[7]Rashad attributes the decline to the removal of endogeneity bias, although endogeneity could result in either a downward or an upward bias on the effect of calories, depending on assumptions. See our discussion below.

sophisticated strategy for mitigating measurement error in self-reported behaviors related to BMI.

Our goal is to identify race and gender variation in health behaviors that help to account, in a proximate sense, for race and gender variation in BMI and obesity. While an analysis of proximate causes begs the question of why behaviors differ, for policy purposes it may be easier to target behaviors rather than underlying causes of such behaviors. For example, suppose that less-educated individuals are more likely to drink sugar-sweetened soft drinks than are well-educated individuals (for causal reasons) and that consuming such drinks contributes to a high BMI. While it remains to be seen, a tax on soft drinks may be a more cost-effective way to reduce soft drink consumption (and BMI) than a policy seeking to redress differences in education, however noble the latter goal.

Our analysis contributes to the debate concerning the contribution of metabolic endowments to observed demographic patterns in obesity— since variations not explained by behaviors must be attributable to such endowments, barring the imperfections of the empirical estimation. In particular, we show that failing to address self-reporting bias in behaviors could lead to the mistaken conclusion that individual behaviors explain only a modest portion of the difference in mean BMI between black women and white women, with the corresponding implication that basal metabolism must differ systematically between these groups.

The remainder of the paper is organized as follows: Section II presents the conceptual model of how BMI is produced. Section III describes the data and sample selection. Section IV explains our method for mitigating measurement error in self-reported behaviors, Section V describes the empirical analysis, and Section VI discusses policy implications and questions for future research.

2 Conceptual Framework

An individual's choice of weight represents a complex problem.[8] In the proximate or structural sense, one's body weight at a point in time, denoted W_i^T, is a stock that reflects lifetime energy balance, or caloric intake net of caloric expenditure, depicted as follows:

$$W_i^T = W_i^0 + \kappa \sum_{t=1}^{T} (EI_i^t - TEE_i^t). \tag{1}$$

In the above equation, W_i^0 is weight at birth (measured in kilograms), EI_i^t is total energy intake or food consumption in a given period (measured in kilocalories or kcal), and TEE_i^t is total energy expenditure in a given period (also measured in kcal). The term κ is a constant that translates energy balance in a given period (measured in kcal) into change in body weight (measured in kilograms). For example, if EI_i^t exceeds TEE_i^t, energy balance is positive and body weight increases over the period. Ignoring childhood and adolescence, when weight and height are generally increasing over time, we focus on adulthood and adopt the stylized assumption that weight is largely stable over time. Holding height constant as well, we can apply the energy balance equation to BMI as well, although the value of κ will differ.[9] In order to maintain weight or BMI at a stable value over time, a necessary condition is that $EI_i^t = TEE_i^t$ within each period.

While the basic energy-balance relationship appears straightforward, caloric intake and caloric expenditure are determined by a complex array of putatively voluntary choices and largely involuntary biological processes. In order to analyze BMI empirically, we have to abstract considerably from such complexity. The goal is to develop a simple optimization framework that will guide the empirical estimation of BMI as a function of (endogenous) behavioral choices such as food intake and physical activity. Such estimation seeks to identify behavioral factors that may account for (in a proximate sense) gender-specific racial variation in BMI and obesity.

[8] Because we are focusing on adults, we assume height is fixed so that only changes in weight can induce changes in BMI, and therefore one's weight choice is equivalent to making a BMI choice.

[9] While it is more realistic to think of a lifecycle path of BMI than a fixed value, we abstract from age in the discussion immediately following. We control for age differences in the empirical analysis.

The following equation represents BMI as the joint product of individual choices and biological factors, subject to an implicit stability condition:

$$BMI_i^s = G(EI_i^s, PAL_i^s, g_i, \epsilon_i). \tag{2}$$

In the above equation, BMI_i^s refers to the stable BMI value that results when the individual adopts a daily caloric intake level equal to EI_i^s and a daily physical activity level equal to PAL_i^s. The physical activity level (PAL) is defined formally as the ratio of total daily energy expenditure, TEE_i^s (in kcal), to the basal metabolic rate, BMR_i^s (also measured in kcal). The basal metabolic rate, in turn, refers to the calories expended per day just in maintaining basic involuntary bodily functions, such as breathing, while in a resting (not sleeping) and fasting state.[10] As such, the PAL represents the average energy-intensity of physical activities performed in a typical day, where energy-intensity is expressed in relation to energy (kcal) expended at rest.[11] Intuitively, PAL increases with time spent in sports and other voluntary activities, and with the intensity of the chosen activities. Beginning from any initial BMI value, an individual who fixes her daily choices at EI_i^s and PAL_i^s will eventually stabilize at BMI_i^s. This stable BMI value also depends on the individual's gender, denoted g_i, and on an idiosyncratic metabolic endowment, ϵ_i, which is described below. We posit that the individual chooses a physical activity level (PAL) rather than total energy expenditure (TEE) — based on the fact that it is harder to observe caloric expenditure than to observe time spent and intensity of physical activities. Total energy expenditure is chosen indirectly, however, as a function of PAL and other factors, and can be expressed as follows:

[10]The term "resting metabolic rate" or RMR is sometimes used interchangeably with BMR. However, measurement standards are stricter for basal metabolism than for resting metabolism. See, for example, http://www.caloriesperhour.com/tutorial_BMR.php.

[11]The PAL is equivalent to the average MET score of activities in a given day, where a MET, or metabolic equivalent, is defined as the calories expended per minute of an activity relative to calories expended per minute at rest. The estimated range of PAL values for sustainable lifestyles runs from 1.2 to 2.5, where 1.2 is indicative of a bed-bound or chair-bound individual and 2.5 represents a very physically active individual (Scrimshaw, Waterlow, and Schurch 1994).

$$TEE_i^s = PAEE_i^s(PAL_i^s, BMI_i^s, g_i) + BMR_i^s(BMI_i^s, g_i, \epsilon_i). \quad (3)$$

In the above equation, the stable value of total energy expenditure, TEE_i^s (in kcal), is separated into two components: physical-activity-related energy expenditure, denoted as $PAEE_i^s$, which depends on the chosen (stable) physical activity level (PAL_i^s), the (stable) BMI value, and gender, g_i; and the basal metabolic rate (BMR_i^s), defined above. BMR depends on the individual's current BMI value, her gender, and on the idiosyncratic metabolic endowment, ϵ_i. The dependence of both components of energy expenditure on BMI reflects the fact that a larger body requires more energy to perform a given amount of work. This dependence also means that a stable BMI must satisfy a fixed-point condition, since the BMI depends on total energy expenditure, which in turn depends on the BMI. The idiosyncratic endowment, ϵ_i, acts as a random shock to the relationship between BMR and BMI. Given this heterogeneity, a stable BMI may vary across individuals for identical behavior choices.[12]

Individuals have their own preferences regarding food intake, physical activity level, and BMI itself. In addition, they face constraints such as income, food prices, and the cost of engaging in physical activity. Taking the BMI production function, $G[.]$, as given, individuals can be seen as jointly choosing their BMI, food intake, and physical activity level to maximize utility subject to constraints. For a given individual and a given BMI value, we assume there is a unique combination of behavioral choices that maximize utility subject to achieving the given BMI. Given indirect (maximized) utility as a function of BMI, individuals choose their BMI to maximize utility globally, and optimal behaviors consistent with the chosen BMI are implied.[13]

Optimal BMI is expressed as a function of behaviors, preferences, endowments, and constraints,

[12]The PAEE may also vary idiosyncratically, although we abstract from such variation for simplicity. This simplification results in only a minor loss of generality.

[13]Optimal BMI is taken to be unique subject to individual parameters. A parameterized structural model of BMI appears in Burke and Heiland (2007), which gives necessary and sufficient conditions for the existence of optimal and stable BMI.

as follows:

$$BMI_i^* = K[EI_i^*(\mu_i, p_i, \epsilon_i), PAL_i^*(\mu_i, p_i, \epsilon_i)]. \tag{4}$$

In the above equation, gender has been dropped for simplicity and the relationship is assumed to be gender-specific. Note that the function $K[.]$ differs from the function $G[.]$ in the previous equation because the latter equation imposes optimality conditions in addition to stability. The term μ_i denotes a vector of individual preferences over food, physical activity, and BMI. Such preferences may reflect physiological factors (an inherent taste for various foods) as well as social and cultural factors such as body size norms and the importance of sports in social life, all of which are taken as exogenous to the individual. Constraints are given by p_i, a vector that includes income, food prices, and the cost of engaging in physical activity.[14]

The use of tobacco can also be incorporated into the model of BMI. While smoking should have no independent effect on energy balance once total caloric intake and total caloric expenditure are taken into account, empirical analysis often finds that smoking has a significant (negative) association with BMI (see, for example, Chou, Grossman, and Saffer 2004). Previous research has found that nicotine may raise the resting metabolic rate (Perkins et al. 1989) and inhibit food consumption via a number of pathways (Miyata et al. 1999 and Grunberg 1982). Since caloric intake and caloric expenditure are typically measured with error, smoking likely acts as a proxy for unmeasured variation in these factors. The choice to smoke (or not to quit once a smoking habit is established) is likely to be simultaneous with BMI if individuals perceive the effects of smoking on body weight. The optimization framework can be readily extended to include smoking as an additional choice available to the individual, under assumptions analogous to those governing the choices of food intake and physical activity.

[14]The direct cost of physical activity in terms of displeasure is captured by preferences.

2.1 Empirical estimation of BMI and the problem of measurement error in self-reported behaviors

There are several obstacles to estimating the equilibrium relationship between BMI and behavioral choices using self-reported data on behaviors such as those contained in the NHANES. First, for even a simple parameterization, optimal BMI is nonlinear in both food intake and the physical activity level (see Burke and Heiland 2007). Second, behavior choices are endogenous in the production of BMI. Third, and the main object of our concern here, behavioral choices are likely to be measured with error for reasons discussed below.[15] We address each of these limitations and interpret the results accordingly. First, while the (nonlinear) structural model of optimal BMI cannot be estimated directly, it can be shown formally that a linear model offers a reasonable approximation to the nonlinear model. In order to discuss the remaining estimation issues, it is useful to express this approximate linear model as follows:

$$BMI_i^s = \alpha + \beta EI_i^s + \gamma PAL_i^s + \mu * \epsilon_i. \tag{5}$$

The above equation represents a linear approximation of the structural model of equilibrium BMI. It assumes that all variables, including BMI, energy intake (EI), and the physical activity level (PAL), are observed at their respective (and mutually consistent) steady-state equilibrium values. The term ϵ_i represents the idiosyncratic metabolic endowment, which has marginal effect μ on BMI, where $\mu < 0$ implies that those individuals with a large metabolic shock (i.e. faster basal metabolic rate) will have lower BMI in equilibrium, all else equal. Ignoring the issue of measurement error in behaviors for the moment, we first address the estimation bias that may arise because behaviors are potentially endogenous with respect to the unobserved endowment term, ϵ_i, which also affects BMI directly yet is unobserved in our data. If behaviors are correlated

[15] Reverse causality—from BMI to behavior—is a separate problem. We assume that BMI and behaviors are determined simultaneously, such that BMI cannot precede behavior. In a fully rational model, any feedbacks from BMI to constraints or preferences (such as occur in models of habit-formation) would be fully anticipated, hence restoring the assumption of simultaneity.

with the metabolic endowment—for example, if an individual with a lower metabolic endowment chooses to eat less than someone with a higher endowment (as occurs in the model of Burke and Heiland 2007)—the estimated effect of calories on BMI will be biased downward relative to the (positive) treatment effect.[16] By similar reasoning, if an individual chooses to engage in more (less) physical activity to offset a lower (or higher) endowment, the estimated effect of physical activity will be biased towards zero relative to (expected negative) treatment effect.[17] Since we cannot rule out the dependence of behaviors on the metabolic endowment, we take these possible biases into account in the remaining discussion. We also note that we are not primarily concerned with demonstrating the existence of treatment effects on BMI of food intake and physical activity, as in broad terms these effects are already well established.

Turning now to the issue of measurement error, recall that the structural relationship described in equation 5 assumes that individuals are in steady state, in which neither BMI nor relevant behaviors fluctuate over time. In reality, of course, such behaviors, and also BMI, may fluctuate significantly over time within an individual around average levels and may exhibit long-term trends in a particular direction. Ignoring trends, fluctuations around steady-state imply that snapshots of behaviors and BMI at a point in time—even if perfectly accurate—constitute noisy measures of average or habitual (i.e., steady state) behaviors and outcomes.[18] Given this potential for classical measurement error, the estimated relationships between behaviors and BMI are likely to be attenuated in the data, ignoring other sources of bias. Given the design of the NHANES surveys, the data on caloric intake are potentially more noisy than the data on physical activity: caloric intake data represent either a single day's calories (in the 1999–2006 sample) or the average caloric intake across two separate days (in the 2003–2006 sample), whereas physical activity questions ask

[16]While this example seems reasonable, positive bias or zero bias are possible under alternative assumptions regarding the correlation between metabolism and caloric intake.

[17]Behaviors are also endogenous with respect to individual preferences and constraints; however, these factors do not directly produce BMI in the physiological sense and hence do not separately enter equation 5 once behaviors are included.

[18]Behaviors could be observed at steady-state values and yet be inconsistent with an observed (non-steady-state) BMI. For example, this will be the case for an individual who recently lost weight due to an illness and just resumed her habitual eating pattern, since it will take time to revert back to her steady-state BMI.

about activities in a "typical" day, week, or month, rather than activities on a specific day or set of days.[19] Concerning measurement error in BMI itself, we remind the reader that our BMI values are not based on self-reported weight and height but on weight and height values measured by trained examiners. For any given individual, the BMI value we observe may nonetheless deviate from its long-run average or steady-state value; however, any given one-day fluctuation in energy intake (or energy expenditure) will result in a much smaller (percentage) change in BMI over the day. For example, an individual who fasts for a day, thereby reducing caloric intake by 100 percent, will lose less than 1 percent of body weight.[20] As a stylized approximation of these facts, we assume that observed BMI represents its steady-state value and yet we do not assume that observed behaviors necessarily represent their respective steady-state values.

Fluctuations around steady state are not the only potential source of measurement error in behaviors. Evidence from previous studies finds that self-reported dietary practices tend on average to understate food intake (Macdiarmid and Blundell 1998). There are several reasons for such self-reporting bias: first, individuals may have an imperfect recall of food intake; second, individuals may deliberately fail to report certain food items and/or may understate true portion sizes; third, prior to the interview people may eat less than normal.[21] In cases of deliberate underreporting, individuals may feel that their food intake is excessive and may be ashamed to report true consumption to the interviewer—an example of a "social desirability" effect, or bias toward giving a socially normative response.[22] In cases of deliberate undereating prior to the interview, subjects may (consciously or not) feel compelled to consume a socially normative level of food intake based on the knowledge that they will have to reveal their intake to an interviewer.

[19] Some yes/no questions about physical activity participation refer to activities conducted in the past 30 days, whereas questions that elicit frequency and duration of activity instruct subjects to describe activity in a "typical" day, week, or month.

[20] Weight loss percentage will vary by individual; the figure is an approximate upper-bound, assuming no change in physical activity level from normal habits.

[21] In the NHANES surveys, individuals are asked to report all items consumed in previous 24-hour period. They know in advance they will be asked to recall intake but are not instructed to write things down as they go, because doing so may induce restraint relative to habitual intake.

[22] Social desirability effects are a long-standing concern in organizational research. See, for example, Ganster et al. 1983. Effects have been observed in survey data on political attitudes (Streb et al. 2008) and church attendance (Smith 1998), among other contexts.

Given its motivating causes, underreporting is likely to be more severe among individuals with high true caloric intake, thus resulting in non-classical measurement error (Macdiarmid and Blundell 1998). Specifically, we expect a negative covariance between true caloric intake and its measurement error (defined as self-reported calories minus true calories). When measurement error has this latter property, Black, Berger, and Scott (2000) show that, under relatively weak conditions, the estimated coefficient on reported calories will be biased towards zero.[23] Therefore, both potential sources of measurement error in caloric intake—mean-zero fluctuations and systematic self-reporting bias—are expected to bias the coefficient on calories in the same direction.

There is evidence that self-reports of physical activity on average overstate true activity levels, due to imperfect recall as well as due to social desirability effects (Beyler et al. 2008). Similar to the predicted biases in caloric intake reports, measurement error in physical activity might be negatively correlated with true activity levels. However, based on our measures of physical activity, we expect overstatement of activity to be less severe than understatement of caloric intake. For example, we use discrete measures of both leisure-time physical activity and "usual" daily activities in the regression analysis. Still, we expect that the effects of physical activity on BMI will be biased towards zero based on the potential for noise and systematic reporting biases.

To mitigate the potential biases caused by measurement error in self-reported behaviors, we adopt two complementary strategies. First, we identify self-reports of food intake that appear invalid in relation to the reported physical activity level and exclude invalid observations from the regression. Note, as explicated below, that we cannot determine whether the reported food intake is too low or too high in an absolute sense, only whether it appears inconsistent with the reported physical activity level, which may also be biased up or down. Second, because the exclusion criterion has relatively low sensitivity when using single-day food intake data, we construct a continuous proxy for the measurement error in food intake relative to physical activity and add this as a control variable in some models. The error proxy helps to correct for residual

[23]The necessary condition is that the sum of the variance of the measurement error and the (negative) covariance of true intake and measurement error must be positive. Our results suggest that the coefficient on self-reported calories is indeed biased towards zero relative to the coefficient on true caloric intake.

underreporting bias as well as classical attenuation bias.

The validity criterion springs from the steady-state condition stated above, $EI_i^s = TEE_i^s$, which states that an individual's daily total energy intake must exactly match her daily total energy expenditure in order to maintain a stable BMI. Recall also that total energy expenditure depends on BMI, which takes its steady-state value in the condition. Dividing both sides of the steady-state condition by the basal metabolic rate (BMR), we obtain the related condition:

$$EI_i^s/BMR_i^s = TEE_i^s/BMR_i^s, \tag{6}$$

which simply states that, in steady state, the ratio of daily energy intake to the BMR must equal the ratio of total energy expenditure to the BMR. The former ratio is called the energy-intake ratio, while the latter ratio is the physical activity level or PAL, defined above. Using this notation, we can rewrite equation 6 as follows: $EI_i^s/BMR_i^s = PAL_i^s$. We will work with these ratios, rather than with caloric intake and caloric expenditure levels directly, because self-reports of physical activity can be more readily translated into PAL values—which are comparable across individuals—than into actual levels of caloric expenditure.[24] If, in the data, the estimated energy-intake ratio (based on self-reported intake and estimated BMR) differs markedly from the estimated PAL value (based on self-reported engagement in physical activities), the self-reported behaviors, with respect to either food intake and/or physical activity, are likely to be inconsistent with steady-state behavior, plus or minus reasonable mean-zero fluctuations.

Consistent with this intuition, Black (2000) describes a method for constructing a 95-percent confidence interval for the energy-intake ratio which is centered around an assumed PAL value. In our application of the method, we construct PAL values (and associated confidence intervals) that are specific to each individual, based on responses to the NHANES physical-activity questionnaire.[25] Each confidence interval allows for (homoscedastic) within-subject variance in caloric

[24] The energy-intake ratio is calculated by dividing reported caloric intake by a predicted value of BMR. The BMR-prediction equation, due to Mifflin et al. 1990, predicts BMR as a linear function of weight, height, age, and sex.

[25] We map the NHANES physical-activity data into PAL values using guidelines drawn from the NHANES and

intake and caloric expenditure, respectively. These variance estimates are based on previous studies (cited in Black 2000) in which individual energy intake and/or expenditure were measured over extended periods using sophisticated methods. The fewer days on which caloric intake is observed, the wider the confidence interval. Because we observe at most two days' worth of caloric intake reports, our confidence intervals are fairly generous. Nonetheless, based on these intervals we exclude roughly 24 percent of relevant observations among women and 21 percent among men. The exclusion criterion identifies a greater number of implausibly low reports of food intake than implausibly high reports, indicating (and also mitigating) a bias toward underreporting of caloric intake.

In addition to the sample restriction, we contribute an additional method for mitigating measurement bias. In particular, we construct a continuous proxy for the joint measurement error in self-reported behaviors, termed "EI-check," short for "energy-intake check." To compute EI-check, we divide the energy-intake ratio by the PAL. Since in steady state these two latter quantities are equal to each other, we obtain the following result:

$$EI\ check^s \equiv (EI_i^s/BMR_i^s)/PAL_i^s = 1. \tag{7}$$

Note that this condition holds for all individuals in steady state, regardless of the individual's BMI. For convenience of exposition in the empirical analysis, we multiply EI-check by 100 with no loss of generality.[26] If all individuals were observed in steady state, we should obtain EI-check values of 100 uniformly, as in the equation above. A value less than 100 indicates that the energy intake ratio falls short of the PAL, which will occur if energy intake is understated, holding PAL at its steady-state value, or if physical activity is overstated, holding intake at its steady-state value. For analogous reasons, EI-check will exceed 100 when self-reported energy intake is overstated relative

from the World Health Organization (FAO/WHO/UNU 2001).

[26]Ignoring the normalization, the measure is equivalent to the ratio of daily energy intake (in calories) to daily energy expenditure (in calories). This is seen by dividing the left-hand side of equation 6 by the right-hand side, where BMR cancels out of the resulting ratio.

to self-reported physical activity.[27] As such, EI-check acts as an index of the joint measurement error in self-reported behaviors. Under classical measurement error alone, deviations of EI-check from 100 would be random with respect to true behaviors and BMI, and the expected value of EI-check would still be 100. With the non-classical measurement error of the kind discussed above, we expect EI-check to be less than 100 on average (bias toward understatement of caloric intake) and it may exhibit a negative covariance with true (steady-state) caloric intake.[28]

Including EI-check in the regression should alleviate the expected bias on the coefficients on caloric intake and physical activity under certain conditions. We can illustrate how the control works by writing a modified version of equation 5 as follows:

$$BMI_i^s = \hat{\alpha} + \hat{\beta} EI_i^{SR} + \hat{\gamma} PAL_i^{SR} + \mu * \epsilon_i + \omega_i. \tag{8}$$

In the above modification, EI_i^{SR} and PAL_i^{SR} refer to self-reported (SR) behaviors, and the coefficients are denoted $\hat{\beta}$ (and $\hat{\alpha}$, and so on), rather than simply β, to reflect potential biases in the empirical estimation of the model. The new term, ω_i, represents the joint contribution of unmeasured variation in behaviors (both energy intake and physical activity) to BMI.[29] Referring back to equation 5, any two individuals with the same *true* (steady-state) values of food intake and physical activity (age and gender constant) have the same steady-state BMI in expectation; their respective realized steady-state BMI values will differ from each other only if the individuals have different metabolic endowments (different ϵ_i). Among individuals with the same set of *self-reported* behaviors, however, variation in BMI may reflect either variation in metabolic endowments, ϵ_i, or variation in measurement error in behaviors, ω_i. Because we observe neither ϵ_i nor ω_i in the

[27] As explained above, survey design predicts that physical activity reports are less susceptible to small-sample fluctuations than caloric intake data. PAL values are also somewhat immune to deliberate overstatement, because the scale of estimated PAL values has an upper bound.

[28] If measurement error in self-reported caloric intake has a negative covariance with true caloric intake, we may or may not obtain a negative covariance between EI-check and true intake. For example, if all individuals understate caloric intake by 15 percent, we will have a negative covariance between true food intake and its measurement error, whereas EI-check will equal 85 uniformly (assuming no error in PAL values). A similar argument applies to measurement error in physical activity.

[29] Since our mitigation strategy cannot separately assess the measurement error in these components, we represent the joint contribution of measurement error as a single factor.

data, variation in BMI conditional on the observed factors could reflect either of these unobserved effects, and correlations between either unobserved factor and any of the regressors could lead to biased coefficient estimates.

If we assume, for the moment, that ϵ_i is orthogonal to the observed variables, then the only potential source of bias is measurement error. In particular, ω_i may be correlated with either self-reported energy intake, self-reported physical activity, or both. (Under classical measurement error, such correlation is guaranteed.) When we include EI-check in the regression, EI-check acts as a proxy for ω_i, the purpose of which is to pick up the variation in BMI—conditional on self-reported behaviors—that can be attributed to the joint measurement error in behaviors. If the proxy works as intended, the estimated relationship between self-reported caloric intake and BMI will be closer to its expected value in the absence of measurement error, and similarly for the relationship between self-reported physical activity patterns and BMI.

Two caveats are in order. Including EI-check does nothing to alleviate the bias to estimated coefficients that may arise (discussed above) if the metabolic endowment (ϵ_i) is correlated with the observed behaviors. Furthermore, the correction method itself depends on certain assumptions. To calculate EI-check, we use predicted values of the basal metabolic rate rather than measured values, which are unavailable in NHANES. These predictions assume, for lack of better information, that all individuals have the mean value (zero) of the metabolic endowment. Therefore, our estimates of EI-check are themselves subject to error. For example, an individual with a low EI-check (based on our estimate) would appear to be understating her food intake relative to her (estimated) caloric needs, but she may simply have a lower-than-expected BMR (relative to our estimate) and hence have lower caloric needs/expenditures than estimated. (In this example, our estimated EI-check values would be correlated with the unobserved metabolic endowments, resulting in inconsistent estimation.) As a result, we may attribute too much variation in BMI to variation in measurement error and too little to variation in metabolic endowments. However, Black (2000) argues that low metabolic endowments are likely not sufficient to explain improbably low values of self-reported food intake, basing this claim on a study by Pryer et al. (1997) involving sophisticated measurements of caloric needs and corresponding self-reports of food intake. In the latter study, self-reported intakes fell below measured caloric needs on average, and by a larger margin among subjects with greater caloric needs. For the method to work as intended, we must assume in general that our EI-check values are uncorrelated with the metabolic endowment.

3 Data and Sample Selection

The empirical analysis is conducted using data from the NHANES, a nationally representative series of cross-sectional studies conducted by the Centers for Disease Control (CDC). The NHANES data include observations of weight, height, and other physical features measured by direct examination, as well as information about demographic and socioeconomic characteristics, life circumstances, behavioral choices, and health conditions, collected via in-person interviews. The empirical analysis here uses the NHANES surveys from the years 1999–2006.[30] NHANES data for 2007 and later are also available, but changes in the design of the physical activity questionnaire beginning in 2007 make it difficult to compare activity levels in the later data with the earlier data.[31]

We restrict the NHANES samples to individuals between the ages of 20 to 65 years who had their height and weight measured by direct examination.[32] Imposing a maximum age of 65 years helps to minimize differences in the age distribution across groups and across survey periods. We select a minimum age of 20 years because we are interested in fully developed adults and, beginning at age 20, the adult criterion for obesity applies. We do not exclude any racial categories from the analysis, although our study focuses on the discrepancies between just two groups: non-Hispanic whites and non-Hispanic blacks. Pregnant women—based on self-reported pregnancy status—

[30] We pool the 1999–2006 data as recommended in the NHANES analytical guidelines (National Center for Health Statistics 1996). Complex survey design is accounted for using Stata's "svy" commands.

[31] Since 1999, the survey has been conducted annually, with statistics reported in two-year increments.

[32] In interview sessions conducted prior to the physical examinations, individuals gave self-reports of weight and height. Individuals who were not subsequently examined are excluded. This omission minimizes measurement error and does not affect representativeness, since survey weights are provided that pertain to use of the examination-only sample.

are excluded from the sample on the grounds that BMI during pregnancy is likely to be above its typical value. The sample that results from these criteria will be called the "full" regression sample. We impose additional sample-selection criteria related to the validity of self-reports of food intake relative to physical activity, as described below, to create an alternative, "validity-restricted" sample. The respective sample sizes for women are 6,225 (full) and 4,730 (validity-restricted), and for men sample sizes are 6,208 (full) and 4,879 (validity-restricted).

We describe the construction of variables used in the empirical analysis in the Appendix.

4 Empirical Analysis

The goal of the empirical analysis is to identify behaviors that may account for gender-specific racial variation in BMI and obesity. We first describe raw differences in food intake, physical activity, and smoking behavior by race and sex. Then we estimate empirical models of BMI and obesity status separately for men and women as functions of these behavioral measures. Descriptive statistics and regression are performed for each of the full sample and the validity-restricted sample. Results are given in Tables 3 through 11.

4.1 Descriptive analysis of behavior

Table 3 shows the mean values of the variables of interest by sex and race, calculated for each full regression sample and validity-restricted regression sample, respectively. As expected, statistically significant differences in mean BMI and obesity rates between blacks and whites are specific to women. The black-white female obesity gap appears slightly greater in the restricted sample than in the full sample, indicating that the restrictions are more likely to exclude obese white women (relative to non-obese white women) than to exclude obese black women.

Both black women and black men are significantly more likely to be classified as sedentary (engaging in no leisure-time physical activity) than their white counterparts, although the gap in the sedentary fraction is greater for black women than black men. White women are significantly more likely to fall into the "vigorous activity" category (highest level of leisure-time physical activity) than black women. White men are also more likely to achieve "vigorous activity" than black men, although the gap in the men's shares is smaller than it is for women and becomes insignificant in the validity-restricted sample. In normal daily activities, black women are significantly more likely than white women to sit most of the time (31 percent versus 25 percent, in either sample). In contrast, black men and white men are about equally likely (21 percent and 22 percent, respectively) to report sitting for most of the day. Daily caloric intake appears marginally higher among black women than white women, but the difference is not statistically significant in either sample. Black men report lower caloric intake than white men, but the difference is only marginally significant (p-value .09) in the restricted sample. White women are significantly more likely to smoke (26 percent are current smokers) than black women (22 percent), while the opposite holds for men—31 percent of white men and 35 percent of black men are smokers. (Smoking figures are for the full sample; the rates in the restricted sample are not significantly different.)

In the full sample, the mean EI-check value is significantly less than 100 (the benchmark value for valid reporting) for each of the four demographic groups of interest, indicating an average bias toward underreporting of caloric intake relative to physical activity. For each group, the mean of EI-check increases significantly in the restricted sample and moves closer to 100. In the restricted-sample, black women have a lower average, at 92.7, than white women, at 94.4, although the difference is not significant. Also for the restricted sample, black men report a significantly lower mean EI-check (95.5) than white men (98.6). These values indicate that, even after eliminating extreme cases, the reported caloric intake still falls below the estimated caloric needs on average, and more so for women than men.

4.2 Regression analysis

Table 4 shows the results of linear (OLS) regressions of female BMI against various groups of regressors for each of two different samples of women. Both samples are unrestricted or "full" samples, as defined above. The first four columns of the table report results pertaining to the NHANES 1999–2006 data, including all women aged 20-65 years who reported at least one day of food intake and with non-missing values for all other regressors. Columns 5 through 8, included as robustness analysis, report results pertaining to NHANES 2003–2006, including all women aged 20-65 years who reported two days of food intake and who had non-missing values for all other variables. Intake values used in regressions in Columns 7 and 8 represent the average of Day 1 intake and Day 2 intake.[33] Results for the 2003–2006 sample are qualitatively and quantitatively similar to those for the 1999–2006 data and are not discussed separately.

The first row in each column of Table 4 gives the estimated difference in mean BMI between black women and white women conditional on the other covariates. The baseline difference (controlling for age and foreign-born status) is 4.0 units (model 1)—roughly 23.5 pounds for a woman of average height (5 feet 4 inches). Model 3 includes the leisure-time physical activity categories, daily activity categories, smoking status, and daily food intake, in addition to the baseline controls. While many of the behavior variables have significant effects on BMI in the directions we would expect, the black-white difference in (conditional) mean BMI remains large and significant, at 3.5 units.

Models 2 and 4 are similar to models 1 and 3, respectively, but in each case we add the EI-check variable as a control for measurement error. When EI-check is included, the effects of (most of) the behavior variables become larger and/or more statistically significant. For example, between models 3 and 4 the effect of the variable "stands"—which refers to standing rather than sitting during normal daily activities—increases in absolute value by a factor of roughly 3, from -0.9 to -2.9. Also, the effect of caloric intake increases from zero to a statistically significant value of 0.015. (The effects of smoking, discussed below, constitute an exception to this pattern.) Consequently, the conditional black-white gap in female mean BMI becomes significantly smaller in model 4 (2.3 units), which conditions on behaviors as well as on EI-check, than in model 3 (3.5 units), which

[33]NHANES 1999–2006 data include indicators for the day of the week on which a given day's food intake was reported. Results are robust to including these indicators as control variables and the effects of the indicators are generally not significant (results not shown).

conditions on behaviors only.

R-squared values are substantially higher when EI-check is included.[34] Between models 1 and 2 the increase in R-squared reflects the (negative) correlation between BMI and EI-check. Between models 3 and 4, the increase reflects the fact that the explanatory power of certain behaviors is enhanced when EI-check is included. In addition, the explanatory power of EI-check itself increases when behaviors are included: note the greater absolute coefficient on EI-check in model 4 compared to model 2. The negative coefficient on EI-check (for example, in model 2) means that, all else equal, an individual with a lower value of EI-check—that is, for whom reported caloric intake appears lower in relation to her estimated caloric needs—has a higher BMI than an individual with a higher value of EI-check. This result agrees with previous findings (for example, Pryer et al. 1997) that higher-BMI individuals underreport caloric intake more than lower-BMI individuals.

Table 5 shows results of linear regressions of female BMI performed on restricted samples, from which we exclude observations with invalid reports of food intake based on the criteria described above. The restrictions result in a loss of 1,495 observations in each of models 1 through 4 (808 observations in models 5 through 8). The R-squared values are in most cases greater under the restricted samples (in all but models 2 and 6), an expected consequence of using less noisy data. However, the gain in explanatory power from sample restriction (for example, seen in comparing the restricted-sample version of model 3 to the full-sample model 3) is considerably less than the gain in explanatory power from adding EI-check (seen comparing models 3 and 4 for either the restricted or full sample.) Compared to the estimates from corresponding models on the full sample, the point estimates of the black-white difference in (conditional) mean BMI are smaller for the restricted sample (for each of models 1 through 8), reaching a low of 1.6 in model 4. However, the pairwise differences between the coefficients in the corresponding models are generally not significant. The effect of caloric intake becomes significant in model 3 due to the sample restriction (in the absence of EI-check), but the effect remains attenuated compared

[34] All R-squared values are adjusted R-squareds.

to model 4, in which EI-check is included. As predicted, sample restrictions alone do not fully mitigate the measurement error in reporting of food intake.

Tables 6 and 7 show the results of Poisson models of obesity risk for women for each of the four samples described in the context of the linear regressions.[35] The first row in each column indicates the estimated ratio of relative obesity risk for black women compared with white women, conditional on the other covariates, and subsequent rows indicate the remaining coefficients of relative risk. The results follow the same basic patterns seen in the linear models.

Looking at Table 6, between model 1 and model 4, the marginal effect of black race on female obesity risk (relative to whites) falls significantly, from 1.65 to 1.24. Again, the effects of caloric intake and physical activity on BMI are stronger and/or more significant when EI-check is included, and the augmented model accounts for a greater portion of the black-white obesity gap. Restricting the sample (Table 7) increases the estimated effect of caloric intake on obesity risk (comparing model 3 in the restricted and full samples), but including EI-check results in a much greater amplification of food's effect. Sample restriction alone does not have significant consequences for the impact of physical activity on obesity risk, with the exception that the effect of "standing" becomes significantly less than one between the full and restricted versions of model 3.

Tables 8 through 11 provide analogous results for men. In linear models on the full sample, black men have a somewhat higher mean BMI than white men in the baseline models (model 1 and model 5), yet the differences become insignificant in models 4 and 8. For the restricted sample, no significant differences in mean BMI between blacks and whites are observed. Again, including EI-check (models 4 and 8) strengthens the effects of caloric intake and physical activity on BMI and results in large increases in explanatory power. In general, the marginal effects of caloric intake and physical activity categories (such as "vigorous" leisure-time activity) on BMI (or obesity risk) are weaker for men than for women and maximum R-squared values (observed

[35]Given the high prevalence of obesity in the subject population, odds ratios from a logistic regression do not approximate relative risks of obesity and are therefore hard to interpret. According to McNutt et al. (2003), a Poisson regression produces reliable estimates of relative risks and conservative confidence intervals. Logistic regressions with identical right-hand side variables yield qualitatively similar results.

in models 4 and 8) are also lower for men. The coefficient on EI-check is negative and significant in all models, but the point estimates are smaller for men than for women.

The contrasts that we observe between results pertaining to women and results pertaining to men are consistent with prior evidence that BMR has a higher variance among men than women (Mifflin et al. 1990). Recall that the variance in basal metabolism (around its predicted value) enters the model residuals and introduces noise into the calculations of EI-check. Therefore, a higher variance in BMR predicts a lower R-squared and an attenuated coefficient on EI-check, such as we observe for men compared to women. A noisier EI-check among men would also help to explain the finding that caloric intake and physical activity have weaker effects on BMI and obesity risk because a noisier EI-check will do less well as an antidote to attenuation bias.

In models that exclude EI-check, being a current smoker is associated with significantly lower BMI for both men and women, and with lower obesity risk in most models, compared to the BMI of nonsmokers. Unlike the effects of caloric intake and physical activity, the effects of smoking on BMI (or obesity risk) generally become smaller and/or lose significance when EI-check is included, although smoking's effects are more robust among men than among women. As explained above, smoking has no effect on BMI in the theoretical model of BMI that includes perfect measures of caloric intake and caloric expenditure. In an empirical model, however, smoking may proxy for unmeasured variation in intake and/or expenditure.[36] By including EI-check as a proxy for measurement error, smoking's proxy effect is apparently reduced. Since we suspect that EI-check is noisier for men, it also makes sense that EI-check would draw less explanatory power from smoking in the case of men compared to the case of women.

Taking the estimated effects of behaviors on BMI in the multivariate analysis, combined with the demographic patterns in the behaviors themselves, we can get a quantitative sense of how these

[36]Whether smoking proxies primarily for caloric intake or caloric expenditure is subject to debate. While smoking has been found to cause acute increases in resting metabolism (for example, Perkins et al. 1989), Perkins (1992) finds that the chronic effects of smoking on resting metabolism are negligible. In addition, Stamford et al. (1986) find that weight gain after smoking cessation is largely explained by increased caloric intake rather than changes in resting metabolism. Smoking's effects on food intake may reflect a number of mechanisms: nicotine may suppress appetite and/or dull taste buds and smoking can serve as a physical and psychological substitute for eating (Miyata et al. 1999 and Grunberg 1982).

behaviors help to account for the black-white BMI gap among women. In results from NHANES 1999–2006 data for the restricted sample (Table 5), the female BMI gap falls by approximately 2.1 units between the baseline model (model 1) and the most inclusive model (model 4). Of these 2.1 BMI units, roughly 0.51 units are accounted for by the 34-calorie-per-day gap in mean caloric intake between black women and white women (seen in Table 3). Lower levels of physical activity among black women (across various indicators) jointly account for about 1.03 units, while the contribution from smoking is negligible. The lower average value of EI-check among black women accounts for 0.57 units, which means that a significant portion of the BMI gap may owe to racial differences in caloric intake and physical activity that are obscured by measurement error. However, we cannot neatly separate out the contribution of measurement error in caloric intake from measurement error in physical activity, nor can we rule out the possibility that EI-check proxies (at least partly) for unmeasured variation in basal metabolism.

Among men, we observe only a small BMI gap between blacks and whites in the baseline model, and the gap is significant only for the unrestricted sample (Table 8). In the comprehensive models, the gaps do not differ significantly from zero. Taking the estimated BMI gaps for the restricted sample at face value, the gap declines by roughly 0.36 units between model 1 and model 4 (Table 9). Lower levels of physical activity among black men can jointly account for most (roughly 0.3 units) of this gap. Complicating the picture, lower caloric intake by black men predicts a negative BMI gap of about 0.54 units and their higher smoking rate predicts a further negative gap of 0.03 units. However, the latter effects are more than offset by the fact that black men have a lower mean EI-check than white men, a difference which predicts a positive BMI gap of 0.64 units. The data suggest, therefore, that any excess BMI among black men likely reflects a greater tendency towards sedentary behavior. This may be offset at least partly by lower caloric intake, but black men's lower reported intake is called into question by the EI-check analysis.

5 Discussion and Conclusion

Using NHANES data from 1999–2006, we describe the contribution that behavioral factors make to explaining the gender-specific differences in mean BMI and obesity risk between blacks and whites in the United States. Using an innovative strategy to mitigate measurement error in self-reported behaviors, we find that a combination of lower physical activity levels and higher caloric intakes can account for roughly half of the difference in mean BMI between black women and white women and more than half of black women's excess obesity risk. Among men, the relatively small black-white gaps in BMI and obesity in uncontrolled data are most likely due to black men's greater tendency to be sedentary, a difference which may be offset partly by lower caloric intake among black men. However, for both men and women, the significant contribution of joint measurement error in self-reported behaviors implies that we cannot determine the relative contributions of physical activity and caloric intake with precision. When measurement error in caloric intake and physical activity are mitigated, smoking's contribution to racial differences in BMI, among both women and men, becomes small-to-negligible.

Our analysis shows that measurement error, if uncontrolled, may severely limit the ability to explain the variation in BMI on the basis of self-reported caloric intake and physical activity. When we take steps to minimize measurement error, the effects of calories consumed and physical activities on outcomes become much stronger (and smoking's effects become weaker), enabling these behaviors to account for a larger share of the black-white BMI gap (or obesity gap) for women and enabling higher R-squared values in all linear regressions. If we were to take at face value the results of the uncontrolled models, we might be led to believe that the lion's share of the BMI gap (or the obesity gap) is driven by a systematic racial difference in basal metabolism, a relatively immutable factor. Our findings indicate that the role of any such differences is likely to be limited. In studies that have measured basal metabolism directly among diverse subject pools, evidence of systematic racial differences has been inconclusive (Martin et al. 2004; Sharp et al. 2002; Weyer et al. 1999; Carpenter et al. 1998; Foster, Wadden, and Vogt 1997; and Yanovski et al. 1997). One caveat to this argument is that our control methods themselves are subject to error and our models may understate the contribution of unobserved factors (basal metabolism, in particular) to the outcomes of interest.

The fact that behaviors related to diet, physical activity, and smoking can account for a significant portion of the obesity gap between black women and white women is important from a policy perspective. While wholesale behavioral change is hard to achieve, our findings suggest that a combination of modest, sustained adjustments on both sides of the energy balance equation may go a long way towards closing the BMI and obesity gaps between black women and white women. For example, our results imply that a relatively small reduction in caloric intake—such as 50 calories per day—could reduce relative obesity risk among black women by 10 percent. While black women engage in significantly less leisure-time physical activity than white women, interventions aimed at increasing black women's participation in such activities only address one portion of overall physical activity. Our results indicate that normal daily activities, which likely pertain to actions on the job or related to home production, may have a significant impact on BMI and obesity risk. If leisure-time is scarce, targeting leisure-time activities may be less effective than, for example, encouraging employers to facilitate physical activity during the workday among workers required to sit most of the time.

Our findings raise the obvious question of why behaviors differ between black women and white women. Using our conceptual framework, black women can be seen as choosing a higher BMI simultaneously with higher caloric intake and lower physical activity levels. The implication is that there must be underlying, gender-specific racial differences in preferences, constraints, and/or biological endowments that predict behavioral differences. Concerning preferences, a number of studies find evidence that black women hold a higher ideal BMI than white women (for example, see Anderson et al. 1992, Furnham and Alibhai 1983, Cusumano and Thompson 1997, and Burke and Heiland 2008). Other studies find that black women in particular face lower social and economic penalties associated with obesity (Averett and Korenman 1996, 1999 and Cawley 2004). While socioeconomic status (or SES, based on income and education) is widely cited as an important underlying determinant of BMI and obesity risk, both black women and black men have significantly lower socioeconomic status than their white counterparts, rendering the explanation insufficient to capture gender specificity. In any event, previous research has shown that the stylized facts are largely robust to controls for socioeconomic status (Burke and Heiland 2008). Concerning constraints, previous research has shown that African-Americans tend to live in neighborhoods that provide greater exposure to fast food restaurants, reduced access to fresh produce, and restricted opportunities for safe exercise (for example, Currie et al. 2009 and Lovasi et al. 2009). However, it is not known whether the differences in such constraints apply to (or influence) African-American women to a greater extent than African-American men. Further investigation of underlying determinants remains a critical area for future research.

While so far we have emphasized the female-specificity of the black-white obesity gap, this fact breaks down among the younger cohorts in our sample. As seen in Table 2, black men in the youngest age group (20–29 years) have a 50 percent higher obesity rate than white men the same age (full sample) and black men ages 30–39 years have a roughly 20 percent higher obesity rate than their white counterparts. The relative risks are still significantly smaller than those observed among women in the same age groups—note that the black-white female obesity gap is also greatest within the youngest age cohort—but the emergence of excess obesity risk among young black men raises important questions for public policy and future research.

REFERENCES

Anderson, Judith L., Charles B. Crawford, Joanne Nadeau, and Tracy Lindberg. 1992. "Was the Duchess of Windsor Right? A Cross-Cultural Review of the Socioecology of Ideals of Female Body Shape." *Ethology and Sociobiology* 13(3): 197–227.

Averett, Susan, and Sanders Korenman. 1996. "The Economic Reality of the Beauty Myth," *Journal of Human Resources* 31: 304–330.

Averett, Susan, and Sanders Korenman. 1999. "Black-White Differences in Social and Economic Consequences of Obesity," *International Journal of Obesity* 223(2): 166–173.

Beyler, Nicholas, Sarah Nusser, Wayne Fuller, and Gregory Welk. 2008. "Relating Self-report and Accelerometer Physical Activity with Application to NHANES 2003–2004." *Proceedings of the Survey Research Methods Section*. American Statistical Association. Available at http://www.amstat.org/sections/srms/proceedings/allyearsf.html.

Black, Alison E. 2000. "Critical Evaluation of Energy Intake Using the Goldberg Cut-Off for Energy Intake:Basal Metabolic Rate. A Practical Guide to its Calculation, Use and Limitations." *International Journal of Obesity* 24(9): 1119–1130.

Black, Dan A., Mark C. Berger, and Frank A. Scott. 2000. "Bounding Parameter Estimates with Nonclassical Measurement Error." *Journal of the American Statistical Association* 95(451): 739–748.

Bureau of State Services, Division of Special Health Services. 1954. "Estimated Prevalence of Overweight in the United States." *Public Health Reports* 69(11): 1084–1086.

Burke, Mary A., and Frank Heiland. 2007. "Social Dynamics of Obesity." *Economic Inquiry* 45(3): 571–591.

Burke, Mary A., and Frank Heiland. 2008. "Race, Obesity, and the Puzzle of Gender Specificity." Federal Reserve Bank of Boston, Working Paper No. 08-7. Available at http://bos.frb.org/economic/wp/wp2008/wp0808.pdf

Burkhauser, Richard V., and John Cawley. 2008. "Beyond BMI: The Value of More Accurate Measures of Fatness and Obesity in Social Science Research." *Journal of Health Economics* 27(2): 519–529.

Carpenter, William H., Tekum Fonong, Michael J. Toth, Philip A. Ades, Jorge Calles-Escandon, Jeremy D. Walston, and Eric T. Poehlman. 1998. "Total Daily Energy Expenditure in Free-Living Older African-Americans and Caucasians." *American Journal of Physiology* 274(37): E96–E101.

Cawley, John. 2004. "The Impact of Obesity on Wages," *Journal of Human Resources* 39: 451–74.

Chang, Virginia W., and Diane S. Lauderdale. 2005. "Income Disparities in Body Mass Index and Obesity in the United States, 1971–2002," *Archives of Internal Medicine* 165: 2122–2128.

Chou, Shin-Yi, Michael Grossman, and Henry Saffer. 2004. "An Economic Analysis of Adult Obesity: Results from the Behavioral Risk Factor Surveillance System." *Journal of Health Economics* 23(3): 565–587.

Currie, Janet, Stefano Della Vigna, Enrico Moretti, and Vikram Pathania. 2009. "The Effect of Fast Food Restaurants on Obesity and Weight Gain." NBER Working Paper No. 14721. Cambridge, MA: National Bureau of Economic Research.

Cusumano, Dale L., and J. Kevin Thompson. 1997. "Body Image and Body Shape Ideals in Magazines: Exposure, Awareness, and Internalization." *Sex Roles* 37(9/10): 701-721.

Denney, Justin T., Patrick M. Krueger, Richard G. Rogers, and Jason D. Boardman. 2004. "Race/Ethnic and Sex Differentials in Body Mass Among US Adults." *Ethnicity & Disease* 14: 389–398.

Food and Agricultural Organization of the United Nations. 2004. *Human Energy Requirements: Report of a Joint FAO/WHO/UNU Expert Consultation.* FAO Food and Nutrition Technical Report Series 1. Rome: Food and Agricultural Organization of the United Nations. Available at ftp://ftp.fao.org/docrep/fao/007/y5686e/y568e00.pdf.

Flegal, Katherine M., Margaret D. Carroll, Robert J. Kuczmarski, and Clifford L. Johnson. 1998. "Overweight and Obesity in the United States: Prevalence and Trends, 1960–1994." *International Journal of Obesity* 22(1): 39–47.

Foster, Gary D., Thomas A. Wadden, and Renee A. Vogt. 1997. "Resting Energy Expenditure in Obese African American and Caucasian Women," *Obesity Research* 5: 1–8.

Furnham, Adrian, and Naznin Alibhai. 1983. "Cross-Cultural Differences in the Perception of Female Body Shapes." *Psychological Medicine* 13(4): 829–837.

Ganster, Daniel C., Harry W. Hennessey, and Fred Luthans. 1983. "Social Desirability Response Effects: Three Alternative Models." Academy of Management Journal 26(2): 321–331.

Grunberg, Neil E. 1982. "The Effects of Nicotine and Cigarette Smoking on Food Consumption and Taste Preferences." *Addictive Behaviors* 7(4): 317–331.

Komlos, John, and Marek Brabec. 2010. "The Trend of Mean BMI Values of U.S. Adults by Centiles, Birth Cohorts 1882–1986." Working Paper No. 16252. Cambridge, MA: National Bureau of Economic Research.

Kuczmarski, Robert J., Katherine M. Flegal, Stephen M. Campbell, and Clifford L. Johnson. 1994. "Increasing Prevalence of Overweight Among U.S. Adults: The National Health and Nutrition Examination Surveys, 1960 to 1991." *Journal of the American Medical Association* 272(3): 205–211.

Kumanyika, Shiriki. 1987. "Obesity in Black Women." *Epidemiologic Reviews* 9(1): 31–50.

Lovasi, Gina S., Malo A. Hutson, Monica Guerra, and Kathryn M. Neckerman. 2009. "Built Environments and Obesity in Disadvantaged Populations." *Epidemiologic Reviews* 31: 7–20.

Macdiarmid, Jennie, and John Blundell. 1998. "Assessing Dietary Intake: Who, What, and Why of Under-reporting." *Nutrition Research Reviews* 11(2): 231–253.

Martin, Kelley, Penny Wallace, Philip F. Rust, and W. Timothy Garvey. 2004. "Estimation of Resting Energy Expenditure Considering Effects of Race and Diabetes Status." *Diabetes Care* 27(6): 1405–1411.

McNutt, Louise-Anne, Chuntao Wu, Xiaonan Xue, and Jean Paul Hafner. 2003. "Estimating the Relative Risk in Cohort Studies and Clinical Trials of Common Outcomes." *American Journal of Epidemiology* 157(10): 940–943.

Mifflin, Mark D., Sachiko T. St. Jeor, Lisa A. Hill, Barbara J. Scott, Sandra A. Daugherty, and Young O. Koh. 1990. "A New Predictive Equation for Resting Energy Expenditure in Healthy Individuals." *American Journal of Clinical Nutrition* 51(2): 241–247.

Miyata, Go, Michael M. Meguid, Serguei O. Fetissov, Giovanni F. Torelli, and Hyune-Ju Kim. 1999. "Nicotine's Effect on Hypothalamic Neurotransmitters and Appetite Regulation." *Surgery* 126(2): 255–263.

Mujahid, Mahasin S., Ana V. Diez Roux, Luisa N. Borrell, and F. Javier Nieto. 2005. "Cross-Sectional and Longitudinal Associations of BMI with Socioeconomic Characteristics." *Obesity Research* 13(8): 1412–1421.

National Center for Health Statistics Centers for Disease Control and Prevention. 1996. *Analytic Reporting Guidelines. The Third National Health and Nutrition Examination Survey. NHANES III 1988–1994.* Hyattsville, MD: Center for Disease Control. Available at http://www.cdc.gov/nchs/data/nhanes/nhanes3/nh3gui.pdf.

Ogden, Cynthia L. 2009. "Disparities in Obesity Prevalence in the United States: Black Women at Risk." *American Journal of Clinical Nutrition* 89(4): 1001–102.

Ogden, Cynthia L., Margaret D. Carroll, Lester R. Curtin, Margaret A. McDowell, Carolyn J. Tabak, and Katherine M. Flegal. 2006. "Prevalence of Overweight and Obesity in the United States, 1999–2004." *Journal of the American Medical Association* 295(13): 1549–1555.

Perkins, Kenneth A. 1992. "Metabolic Effects of Cigarette Smoking." *Journal of Applied Physiology* 72(2): 401–409.

Perkins, Kenneth A. 1993. "Weight Gain Following Smoking Cessation." *Journal of Consulting and Clinical Psychology* 61(5): 768–777.

Perkins, Kenneth A., Leonard H. Epstein, Richard L. Stiller, Bonita L. Marks, and Rolf G. Jacob. 1989. "Acute Effects of Nicotine on Resting Metabolic Rate in Cigarette Smokers." *American Journal of Clinical Nutrition* 50(3): 545–550.

Pryer, Jane A., Martine Vrijheid, Robert Nichols, Matthew Kiggins, and Paul Elliott. 1997. "Who Are the 'Low Energy Reporters' in the Dietary and Nutritional Survey of British Adults?" *International Journal of Epidemiology* 26(1): 146–154.

Rashad, Inas. 2006. "Structural Estimation of Caloric Intake, Exercise, Smoking, and Obesity." *Quarterly Review of Economics and Finance* 46(2): 268–283.

Scrimshaw, Nevin S., John C. Waterlow, and Beat Schurch, eds., 1994. Energy and Protein Requirements, Proceedings of an IDECG workshop.
URL: http://unu.edu/unupress/food2/UID01E/UID01E08.HTM.

Sharp, Teresa A., Melanie L. Ball, Gary K. Grunwald, Kathryn H. Schmitz, Stephen Sidney, Cora E. Lewis, Kim Tolan, and James O. Hill. 2002. "Differences in Resting Metabolic Rate between White and African-American Young Adults." *Obesity Research* 10(8): 726–732.

Smith, Tom W. 1998. "A Review of Church Attendance Measures." *American Sociological Review* 63: 131–136.

Stamford, Bryant A., Sharleen Matter, Ronald D. Fell, and Paula Papanek. 1986. "Effects of Smoking Cessation on Weight Gain, Metabolic Rate, Caloric Consumption, and Blood Lipids." *American Journal of Clinical Nutrition* 43(4): 486–494.

Streb, Matthew J., Barbara Burrell, Brian Frederick, and Michael A. Genovese. 2008. "Social Desirability Effects and Support for a Female American President." *Public Opinion Quarterly* 72(1): 76–89.

Weyer, Christian, Sören Snitker, Clifton Bogardus, and Eric Ravussin. 1999. "Energy Metabolism in African Americans: Potential Risk Factors for Obesity." *American Journal of Clinical Nutrition* 70(1): 13–20.

Yanovski, Susan Z., John C. Reynolds, Andrew J. Boyle, and Jack A. Yanovski (1997), "Resting Metabolic Rate in African-American and Caucasian Girls," *Obesity Research* 5(4): 321–325.

Zhang, Qi, and Youfa Wang. 2004. "Socioeconomic Inequality of Obesity in the United States: Do Gender, Age, and Ethnicity Matter?" *Social Science and Medicine* 58(6): 1171–1180.

Table 1: Obesity Rates (in Percentages) of U.S. Adults Ages 20–74 years, by Gender, Race, and Poverty Status[1]

	1960–62	**1971–74**	**1976–80**	**1988–94**	**1999–02**	**2003–06**
			20–74 Years, Age adjusted[3]			
Total population:						
Both sexes	13.3	14.6	15.1	23.3	31.1	34.1
Male	10.7	12.2	12.8	20.6	28.1	33.1
Female	15.7	16.8	17.1	26.0	34.0	35.2
Not Hispanic or Latino:						
White, male	—	—	12.4	20.7	28.7	33.0
White, female	—	—	15.4	23.3	31.3	32.5
Black/African American, male	—	—	16.5	21.3	27.9	36.3
Black/African American, female	—	—	31.0	39.1	49.6	54.3
Percent of poverty level:[2]						
Below 100%	—	20.7	21.9	29.2	36.0	35.9
100%–less than 200%	—	18.4	18.7	26.6	35.4	36.7
200% or more	—	12.4	12.9	21.4	29.2	33.1

Sources: NHES I, NHANES II, NHANES III, NHANES 1999–2006
Notes: —Data not available. [1]Based on CDC Trend tables and Chartbook Tables in Excel format, 2006 Edition, Table 73 (http://www.cdc.gov/nchs/hus.htm). [2]Poverty level is based on family income and family size. Persons with unknown poverty level are excluded. [3]Age adjusted to the 2000 standard population using five age groups: 20–34 years, 35–44 years, 45–54 years, 55–64 years, and 65 years and over (65–74 years for estimates for 20–74 years).

Table 2: Obesity Rates (in Percentages) of U.S. Adults Ages 20-65 years, by Gender, Race, and Regression Sample, NHANES 1999–2006

	Women				Men			
	Whites		Blacks		Whites		Blacks	
	Full	Restricted	Full	Restricted	Full	Restricted	Full	Restricted
	N=2,817	N=2,206	N=1,403	N=977	N=2,920	N=2,386	N=1,311	N=932
Age								
Ages 20–29 Years	21.3	16.9	47.7	44.1	21.4	19.8	32.6	30.3
	(2.2)	(2.4)	(3.9)	(5.1)	(2.3)	(2.3)	(3.3)	(3.6)
Ages 30–39 Years	27.7	24.4	50.3	47.6	27.5	26.1	33.5	34.6
	(2.6)	(2.8)	(3.5)	(4.6)	(2.0)	(2.4)	(3.2)	(4.3)
Ages 40–49 Years	36.5	35.7	52.1	50.0	33.9	33.1	31.8	27.8
	(2.2)	(2.6)	(3.0)	(3.7)	(2.3)	(2.7)	(2.7)	(3.0)
Ages 50–65 Years	37.5	35.5	53.8	51.5	36.0	33.3	33.9	31.5
	(1.9)	(2.0)	(2.9)	(3.4)	(1.8)	(2.1)	(2.6)	(3.4)
Foreign-born Status								
Not U.S.-born	17.6	16.0	27.6	25.4	26.0	26.7	13.5	12.3
	(3.4)	(3.4)	(4.1)	(4.9)	(4.3)	(4.6)	(2.8)	(3.5)
U.S.-born	32.7	30.4	53.0	50.6	30.8	29.1	35.2	33.3
	(1.3)	(1.3)	(1.8)	(2.4)	(1.3)	(1.3)	(1.8)	(2.2)
Education								
High School Dropout	38.7	33.7	50.6	46.2	33.9	32.2	29.6	28.6
	(2.6)	(3.0)	(3.1)	(4.4)	(3.9)	(4.6)	(2.4)	(3.0)
High School Graduate	37.1	32.9	53.0	50.6	31.8	29.6	34.4	31.1
	(1.8)	(1.9)	(3.4)	(4.0)	(1.7)	(2.0)	(2.7)	(3.0)
Some College	28.9	27.9	50.5	48.8	29.5	28.2	34.3	32.6
	(1.5)	(1.7)	(2.3)	(2.9)	(1.6)	(1.6)	(2.9)	(3.3)
Income								
Low Income	37.8	32.7	53.0	48.6	28.6	27.9	29.3	27.4
	(1.7)	(2.2)	(2.5)	(3.1)	(1.9)	(2.4)	(2.3)	(2.9)
Middle Income	35.1	32.6	55.0	53.6	33.6	31.8	36.2	32.3
	(2.8)	(2.9)	(3.4)	(3.9)	(2.7)	(2.9)	(2.8)	(3.5)
High Income	28.0	27.0	43.2	43.4	30.0	28.0	35.5	34.4
	(1.5)	(1.7)	(3.3)	(3.5)	(1.6)	(1.7)	(2.8)	(3.7)
Physical Activity								
Sedentary	41.1	38.2	52.9	50.0	35.5	32.8	29.6	28.8
	(1.6)	(1.7)	(2.4)	(2.7)	(2.0)	(2.2)	(2.3)	(3.2)
Light Activity	38.4	36.2	57.3	56.4	31.9	31.9	31.5	31.5
	(2.1)	(2.3)	(4.5)	(5.4)	(2.2)	(2.5)	(3.4)	(4.0)
Moderate Activity	29.1	24.5	47.5	49.8	30.4	28.8	41.4	37.8
	(3.0)	(2.9)	(5.0)	(5.0)	(3.3)	(3.3)	(4.5)	(4.7)
Vigorous Activity	20.3	18.6	43.3	35.1	26.5	24.6	34.8	31.1
	(1.7)	(1.9)	(2.8)	(3.3)	(1.8)	(2.0)	(2.8)	(3.3)
Sits	39.9	39.6	51.7	48.6	36.5	35.6	33.1	33.5
	(2.3)	(2.6)	(3.2)	(3.4)	(2.3)	(2.6)	(3.8)	(4.1)
Stands	32.2	29.2	52.1	50.7	31.0	28.7	34.5	32.7
	(1.6)	(1.7)	(2.2)	(3.0)	(2.1)	(2.3)	(2.3)	(2.8)
Light Lifting	21.2	17.7	47.5	41.9	25.4	24.4	33.5	26.7
	(2.2)	(2.1)	(3.4)	(5.3)	(2.2)	(2.4)	(4.1)	(4.4)
Heavy Lifting	34.1	31.1	42.2	35.4	27.8	26.1	24.9	24.8
	(6.8)	(7.6)	(8.9)	(11.8)	(2.0)	(2.4)	(3.9)	(4.6)
Smoking								
Nonsmoker	32.5	32.2	52.0	49.1	31.5	30.4	37.2	35.2
	(1.7)	(1.9)	(2.2)	(3.1)	(1.9)	(2.0)	(2.5)	(3.2)
Former Smoker	31.5	27.2	49.3	47.3	29.8	27.9	28.7	27.0
	(1.3)	(1.4)	(2.7)	(3.3)	(1.5)	(1.6)	(1.9)	(2.7)
Current Smoker	28.7	23.0	45.4	42.4	23.6	21.8	24.7	24.2
	(1.7)	(1.6)	(3.3)	(4.3)	(1.8)	(1.8)	(2.3)	(3.0)

Source: Authors' calculations based on NHANES 1999–2006.
Notes: Standard errors given in parentheses.

Table 3: Sample Means, Dependent and Independent Variables, by Gender, Race, and Regression Sample, NHANES 1999-2006

	Women				**Men**			
	Whites		Blacks		Whites		Blacks	
	Full N=2,817	Restricted N=2,206	Full N=1,403	Restricted N=977	Full N=2,920	Restricted N=2,386	Full N=1,311	Restricted N=932
BMI	27.84	27.377	31.59	30.78	28.18	27.97	28.47	28.00
	(0.21)	(0.22)	(0.31)	(0.23)	(0.16)	(0.16)	(0.26)	(0.29)
Obese	0.32	0.27	0.51	0.49	0.31	0.29	0.33	0.31
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Age	42.58	42.67	40.41	40.51	41.87	42.10	39.57	39.40
	(0.32)	(0.33)	(0.38)	(0.37)	(0.31)	(0.36)	(0.45)	(0.50)
Ages 20–29 Years	0.19	0.18	0.22	0.22	0.21	0.20	0.26	0.27
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Ages 30–39 Years	0.23	0.23	0.26	0.26	0.23	0.23	0.24	0.25
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Ages 40–49 Years	0.26	0.27	0.26	0.26	0.25	0.26	0.25	0.24
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Ages 50–65 Years	0.33	0.33	0.26	0.26	0.32	0.32	0.25	0.24
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Not U.S.-born	0.05	0.05	0.08	0.08	0.05	0.05	0.11	0.11
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Sedentary	0.30	0.30	0.49	0.51	0.27	0.27	0.39	0.38
	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Light Activity	0.24	0.25	0.19	0.20	0.19	0.20	0.15	0.17
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Moderate Activity	0.14	0.15	0.10	0.10	0.15	0.15	0.10	0.11
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Vigorous Activity	0.33	0.30	0.22	0.19	0.39	0.37	0.35	0.33
	(0.02)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Sits	0.25	0.25	0.31	0.31	0.22	0.22	0.21	0.21
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
Stands	0.53	0.53	0.52	0.53	0.43	0.43	0.52	0.52
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Light Lifting	0.20	0.19	0.14	0.13	0.21	0.21	0.16	0.16
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Heavy Lifting	0.03	0.03	0.02	0.03	0.14	0.14	0.11	0.11
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Total Kcal	1878.80	2030.53	1894.14	2064.58	2767.56	2866.38	2587.42	2792.43
	(15.59)	(15.64)	(29.89)	(26.59)	(23.71)	(22.88)	(34.33)	(38.73)
Nonsmoker	0.51	0.51	0.66	0.65	0.43	0.44	0.49	0.49
	(0.01)	(0.01)	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.02)
Former Smoker	0.23	0.23	0.12	0.12	0.26	0.26	0.16	0.17
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
Current Smoker	0.26	0.26	0.22	0.23	0.31	0.30	0.35	0.34
	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
EI-check	86.64	94.44	84.84	92.70	95.02	98.55	88.22	95.51
	(0.76)	(0.64)	(1.51)	(1.16)	(0.82)	(0.68)	(1.18)	(1.15)

Source: Authors' calculations based on NHANES 1999–2006.
Notes: Standard errors given in parentheses.

Table 4: Determinants of BMI, Linear Regressions, Females Aged 20–65 Years, Full Sample

	NHANES 1999–2006				**NHANES 2003-2006**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black	4.004***	3.911***	3.531***	2.283***	4.224***	4.071***	4.042***	2.471***
	(0.335)	(0.307)	(0.342)	(0.251)	(0.491)	(0.421)	(0.524)	(0.410)
Ages 30–39	1.208***	1.201***	0.879**	1.386***	1.409**	1.379**	1.166*	1.565***
	(0.365)	(0.360)	(0.361)	(0.293)	(0.622)	(0.637)	(0.591)	(0.477)
Ages 40–49	2.422***	2.395***	1.927***	3.010***	2.721***	2.790***	2.361***	3.520***
	(0.388)	(0.379)	(0.382)	(0.290)	(0.458)	(0.461)	(0.458)	(0.465)
Ages 50–65	3.028***	2.827***	2.353***	4.943***	3.113***	3.020***	2.553***	5.605***
	(0.338)	(0.355)	(0.349)	(0.302)	(0.500)	(0.534)	(0.520)	(0.475)
Not Born US	-2.799***	-2.616***	-3.190***	-0.713**	-3.018***	-3.014***	-3.229***	-0.722*
	(0.345)	(0.369)	(0.308)	(0.288)	(0.455)	(0.516)	(0.415)	(0.406)
Light Activity			-0.231	-0.304			-0.025	0.165
			(0.296)	(0.186)			(0.529)	(0.297)
Moderate Activity			-2.049***	-1.491***			-1.477***	-0.488
			(0.393)	(0.239)			(0.474)	(0.292)
Vigorous Activity			-2.400***	-5.516***			-2.010***	-5.004***
			(0.306)	(0.221)			(0.527)	(0.369)
Current Smoker			-1.305***	-0.220			-0.698*	-0.036
			(0.219)	(0.191)			(0.359)	(0.341)
Former Smoker			0.601	0.407			1.212*	0.614
			(0.406)	(0.317)			(0.711)	(0.501)
Stands			-0.943***	-2.916***			-0.831	-2.843***
			(0.247)	(0.254)			(0.526)	(0.448)
Light Lifting			-2.035***	-5.370***			-1.824***	-5.530***
			(0.393)	(0.332)			(0.617)	(0.508)
Heavy Lifting			-1.332**	-7.459***			-1.607	-7.898***
			(0.625)	(0.492)			(0.975)	(0.878)
Total Kcal			0.0002	0.0149***			0.0006	0.0166***
			(0.0002)	(0.0005)			(0.0004)	(0.0008)
EI-check		-0.041***		-0.327***		-0.054***		-0.370***
		(0.003)		(0.010)		(0.007)		(0.017)
Constant	26.080***	29.707***	28.457***	30.291***	26.004***	30.595***	26.955***	29.895***
	(0.311)	(0.432)	(0.491)	(0.396)	(0.476)	(0.892)	(1.006)	(0.806)
R^2 (adj.)	0.067	0.112	0.105	0.522	0.073	0.123	0.103	0.546
N	6,225	6,225	6,225	6,225	2,868	2,868	2,868	2,868

Source: Authors' calculations.
Notes: Standard errors are reported in parentheses. All regressions also control for Mexican, other Hispanic, and other Race/Ethnicity. *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Table 5: Determinants of BMI, Linear Regressions, Females Ages 20–65 Years, Restricted Sample

	NHANES 1999–2006				**NHANES 2003–2006**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black	3.684***	3.605***	2.970***	1.593***	4.043***	4.044***	3.355***	1.875***
	(0.344)	(0.315)	(0.375)	(0.242)	(0.450)	(0.426)	(0.492)	(0.355)
Ages 30-39	1.641***	1.602***	1.551***	2.080***	2.127***	2.072***	2.131***	2.106***
	(0.393)	(0.382)	(0.393)	(0.266)	(0.611)	(0.624)	(0.520)	(0.376)
Ages 40-49	2.959***	2.890***	2.809***	3.777***	3.260***	3.293***	3.285***	4.031***
	(0.408)	(0.398)	(0.387)	(0.255)	(0.499)	(0.476)	(0.501)	(0.414)
Ages 50-65	3.456***	3.301***	3.564***	5.855***	3.458***	3.400***	3.816***	6.192***
	(0.423)	(0.422)	(0.415)	(0.292)	(0.593)	(0.601)	(0.542)	(0.426)
Not Born US	-2.648***	-2.670***	-2.582***	-0.525*	-2.655***	-2.645***	-2.455***	-0.322
	(0.382)	(0.399)	(0.333)	(0.285)	(0.438)	(0.473)	(0.374)	(0.406)
Light Activity			-0.254	-0.259			0.105	0.272
			(0.323)	(0.208)			(0.592)	(0.278)
Moderate Activity			-2.293***	-1.517***			-1.777***	-0.797***
			(0.432)	(0.236)			(0.555)	(0.286)
Vigorous Activity			-2.673***	-5.358***			-2.880***	-5.013***
			(0.335)	(0.239)			(0.534)	(0.407)
Current Smoker			-1.575***	-0.577***			-0.700	-0.136
			(0.264)	(0.209)			(0.471)	(0.359)
Former Smoker			0.207	-0.005			0.362	0.102
			(0.419)	(0.306)			(0.470)	(0.393)
Stands			-1.223***	-3.119***			-1.100**	-2.870***
			(0.296)	(0.281)			(0.504)	(0.358)
Light Lifting			-2.688***	-5.791***			-2.094***	-5.558***
			(0.427)	(0.358)			(0.573)	(0.442)
Heavy Lifting			-2.400***	-8.495***			-2.860***	-8.052***
			(0.645)	(0.525)			(1.042)	(0.731)
Total Kcal			0.0027***	0.0151***			0.0036***	0.0158***
			(0.0002)	(0.0005)			(0.0006)	(0.0008)
EI-check		-0.041***		-0.319***		-0.050***		-0.334***
		(0.005)		(0.012)		(0.009)		(0.017)
Constant	25.217***	29.130***	22.529***	28.637***	24.867***	29.491***	19.873***	27.597***
	(0.328)	(0.619)	(0.616)	(0.506)	(0.514)	(1.007)	(1.162)	(0.958)
R^2 (adj.)	0.073	0.096	0.168	0.576	0.089	0.114	0.192	0.592
N	4,730	4,730	4,730	4,730	2,060	2,060	2,060	2,060

Source: Authors' calculations.
Notes: Standard errors are reported in parentheses. All regressions also control for Mexican, other Hispanic, and other Race/Ethnicity. *Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Table 6: Determinants of Obesity, Poisson Regressions, Females Ages 20–65 Years, Full Sample

	NHANES 1999-2006				NHANES 2003-2006			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black	1.654***	1.609***	1.522***	1.242***	1.702***	1.658***	1.639***	1.321***
	(0.074)	(0.065)	(0.071)	(0.056)	(0.097)	(0.086)	(0.098)	(0.084)
Ages 30-39	1.238***	1.247***	1.170**	1.219***	1.382**	1.387**	1.307*	1.362**
	(0.100)	(0.098)	(0.092)	(0.082)	(0.195)	(0.200)	(0.176)	(0.163)
Ages 40-49	1.499***	1.507***	1.386***	1.555***	1.620***	1.660***	1.514***	1.760***
	(0.117)	(0.113)	(0.107)	(0.088)	(0.138)	(0.142)	(0.118)	(0.141)
Ages 50-65	1.566***	1.530***	1.418***	1.849***	1.629***	1.620***	1.484***	2.028***
	(0.109)	(0.110)	(0.101)	(0.128)	(0.174)	(0.181)	(0.152)	(0.188)
Not Born US	0.574***	0.588***	0.538***	0.789***	0.565***	0.565***	0.544***	0.773**
	(0.048)	(0.050)	(0.044)	(0.063)	(0.058)	(0.061)	(0.054)	(0.076)
Light Activity			0.991	1.016			1.027	1.085
			(0.051)	(0.049)			(0.075)	(0.069)
Moderate Activity			0.758***	0.879*			0.867	1.020
			(0.058)	(0.060)			(0.083)	(0.098)
Vigorous Activity			0.610***	0.368***			0.635***	0.396***
			(0.039)	(0.027)			(0.057)	(0.045)
Current Smoker			0.850***	1.001			0.915	0.973
			(0.040)	(0.050)			(0.066)	(0.084)
Former Smoker			1.027	1.027			1.157*	1.078
			(0.056)	(0.049)			(0.094)	(0.068)
Stands			0.943	0.747***			1.014	0.794**
			(0.045)	(0.038)			(0.100)	(0.073)
Light Lifting			0.705***	0.492***			0.747*	0.479***
			(0.064)	(0.041)			(0.107)	(0.066)
Heavy Lifting			0.869	0.420***			0.957	0.489***
			(0.139)	(0.060)			(0.212)	(0.097)
Total Kcal			1.0000	1.0020***			1.0001**	1.0022***
			(0.0000)	(0.0001)			(0.0000)	(0.0001)
EI-check		0.993***		0.952***		0.992***		0.948***
		(0.001)		(0.002)		(0.001)		(0.003)
N	6,225	6,225	6,225	6,225	2,868	2,868	2,868	2,868

Source: Authors' Calculations.
Notes: Coefficients represent relative risks. Standard errors are reported in parentheses. All regressions also control for Mexican, other Hispanic, and other Race/Ethnicity.
*Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Table 7: Determinants of Obesity, Poisson Regressions, Females Ages 20–65 Years, Restricted Sample

	NHANES 1999-2006				NHANES 2003–2006			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black	1.703***	1.683***	1.504***	1.216***	1.785***	1.782***	1.621***	1.281***
	(0.098)	(0.091)	(0.092)	(0.081)	(0.130)	(0.130)	(0.113)	(0.106)
Ages 30-39	1.347***	1.340***	1.294**	1.319***	1.625***	1.604***	1.592***	1.518***
	(0.142)	(0.139)	(0.129)	(0.112)	(0.259)	(0.261)	(0.233)	(0.199)
Ages 40-49	1.711***	1.699***	1.630***	1.741***	1.894***	1.9100***	1.8582***	2.0341***
	(0.161)	(0.156)	(0.148)	(0.122)	(0.222)	(0.219)	(0.221)	(0.230)
Ages 50-65	1.760***	1.721***	1.748***	2.111***	1.790***	1.757***	1.860***	2.335***
	(0.167)	(0.164)	(0.159)	(0.170)	(0.253)	(0.245)	(0.254)	(0.267)
Not Born US	0.591***	0.587***	0.600***	0.844*	0.602***	0.603***	0.610***	0.899
	(0.055)	(0.056)	(0.054)	(0.073)	(0.069)	(0.073)	(0.065)	(0.102)
Light Activity			1.004	1.037			1.069	1.126
			(0.059)	(0.058)			(0.090)	(0.092)
Moderate Activity			0.710***	0.849**			0.865	1.080
			(0.065)	(0.068)			(0.103)	(0.121)
Vigorous Activity			0.572***	0.367***			0.546***	0.349***
			(0.048)	(0.038)			(0.056)	(0.048)
Current Smoker			0.760***	0.899*			0.864	0.896
			(0.045)	(0.054)			(0.090)	(0.101)
Former Smoker			0.961	0.980			1.036	0.998
			(0.064)	(0.054)			(0.095)	(0.089)
Stands			0.865**	0.684***			0.970	0.780**
			(0.053)	(0.043)			(0.111)	(0.081)
Light Lifting			0.591***	0.430***			0.703**	0.436***
			(0.065)	(0.039)			(0.100)	(0.064)
Heavy Lifting			0.723*	0.359***			0.865	0.484***
			(0.138)	(0.065)			(0.207)	(0.102)
Total Kcal			1.0004***	1.0020***			1.0005***	1.0022***
			(0.0000)	(0.0001)			(0.0001)	(0.0001)
EI-check		0.993***		0.955***		0.990***		0.946***
		(0.001)		(0.002)		(0.003)		(0.003)
N	4,730	4,730	4,730	4,730	2,060	2,060	2,060	2,060

Source: Authors' calculations.
Notes: Coefficients represent relative risks. Standard errors are reported in parentheses. All regressions also control for Mexican, other Hispanic, and other Race/Ethnicity.
*Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Table 8: Determinants of BMI, Linear Regressions, Males Ages 20–65 Years, Full Sample

	NHANES 1999–2006				**NHANES 2003–2006**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black	0.536***	0.349	0.515	0.136	1.051**	0.506	0.877**	0.287
	(0.298)	(0.293)	(0.310)	(0.254)	(0.408)	(0.371)	(0.399)	(0.271)
Ages 30-39	1.208***	1.317***	0.939***	1.512***	1.631***	1.906***	1.466***	1.873***
	(0.246)	(0.235)	(0.247)	(0.229)	(0.319)	(0.293)	(0.315)	(0.254)
Ages 40-49	2.083***	2.127***	1.780***	2.752***	2.446***	2.593***	2.260***	2.924***
	(0.276)	(0.272)	(0.276)	(0.271)	(0.417)	(0.374)	(0.408)	(0.280)
Ages 50-65	2.013***	1.851***	1.486***	3.239***	2.076***	1.933***	1.692***	3.235***
	(0.284)	(0.266)	(0.281)	(0.229)	(0.447)	(0.374)	(0.432)	(0.304)
Not Born US	-2.062**	-2.179***	-2.256***	-1.301***	-1.982***	-2.041***	-2.185***	-1.117***
	(0.347)	(0.327)	(0.347)	(0.259)	(0.488)	(0.469)	(0.456)	(0.321)
Light Activity			-0.180	-0.250			0.109	-0.011
			(0.240)	(0.206)			(0.402)	(0.294)
Moderate Activity			-0.221	-0.565**			-0.401	-0.546*
			(0.312)	(0.261)			(0.492)	(0.301)
Vigorous Activity			-0.803***	-3.754***			-0.548	-3.385***
			(0.269)	(0.255)			(0.366)	(0.263)
Current Smoker			-1.657***	-0.655**			-1.660***	-0.640***
			(0.250)	(0.256)			(0.266)	(0.229)
Former Smoker			0.076	0.087			-0.420	-0.020
			(0.248)	(0.212)			(0.2660)	(0.217)
Stands			-0.842***	-1.746***			-0.407	-1.553***
			(0.286)	(0.247)			(0.390)	(0.353)
Light Lifting			-1.310***	-3.122***			-0.768**	-2.753***
			(0.327)	(0.321)			(0.360)	(0.397)
Heavy Lifting			-0.817***	-4.882***			-0.326	-5.143***
			(0.302)	(0.356)			(0.498)	(0.536)
Total Kcal			-0.0001	0.0070***			-0.0004**	0.0075***
			(0.0001)	(0.0003)			(0.0002)	(0.0004)
EI-check		-0.027***		-0.211***		-0.047***		-0.237***
		(0.003)		(0.009)		(0.005)		(0.009)
Constant	26.846	29.403***	28.920***	30.571***	26.711***	31.063***	29.292***	31.585***
	(0.228)	(0.327)	(0.462)	(0.386)	(0.344)	(0.475)	(0.690)	(0.580)
R^2 (adj.)	0.032	0.067	0.058	0.323	0.041	0.117	0.066	0.372
N	6,208	6,208	6,208	6,208	2,805	2,805	2,805	2,805

Source: Authors' calculations.
Notes: Standard errors are reported in parentheses. All regressions also control for Mexican, other Hispanic, and other Race/Ethnicity.
*Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Table 9: Determinants of BMI, Linear Regressions, Males Ages 20–65 Years, Restricted Sample

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black	0.304	0.210	0.345	-0.053	0.583	0.376	0.625	0.327
	(0.336)	(0.334)	(0.326)	(0.270)	(0.413)	(0.414)	(0.383)	(0.286)
Ages 30-39	1.184***	1.291***	0.953***	1.683***	1.969***	2.033***	1.812***	2.223***
	(0.260)	(0.259)	(0.263)	(0.263)	(0.347)	(0.330)	(0.347)	(0.240)
Ages 40-49	2.221***	2.249***	2.053***	3.089***	2.876***	2.874***	2.787***	3.368***
	(0.287)	(0.283)	(0.280)	(0.272)	(0.438)	(0.412)	(0.433)	(0.287)
Ages 50-65	1.880***	1.782***	1.780***	3.647***	2.407***	2.372***	2.364***	3.933***
	(0.281)	(0.269)	(0.263)	(0.229)	(0.394)	(0.347)	(0.407)	(0.283)
Not Born US	-2.055***	-2.171***	-1.976***	-1.077***	-1.627***	-1.625***	-1.631***	-0.647**
	(0.337)	(0.329)	(0.338)	(0.270)	(0.447)	(0.427)	(0.426)	(0.310)
Light Activity			0.023	-0.148			0.010	-0.087
			(0.276)	(0.236)			(0.395)	(0.309)
Moderate Activity			-0.111	-0.507*			-0.327	-0.426
			(0.333)	(0.290)			(0.407)	(0.271)
Vigorous Activity			-1.015***	-3.905***			-0.877**	-3.639***
			(0.270)	(0.244)			(0.340)	(0.304)
Current Smoker			-1.666***	-0.726***			-1.433***	-0.578*
			(0.253)	(0.258)			(0.325)	(0.286)
Former Smoker			0.082	0.074			-0.090	0.140
			(0.287)	(0.218)			(0.305)	(0.240)
Stands			-0.933***	-1.792***			-0.648	-1.588***
			(0.318)	(0.244)			(0.425)	(0.355)
Light Lifting			-1.624***	-3.305***			-1.124**	-2.892***
			(0.359)	(0.325)			(0.441)	(0.412)
Heavy Lifting			-1.424***	-5.283***			-0.925*	-5.477***
			(0.362)	(0.329)			(0.490)	(0.511)
Total Kcal			0.0009***	0.0073***			0.0008***	0.0078***
			(0.0001)	(0.0003)			(0.0002)	(0.0004)
EI-check		-0.030***		-0.216***		-0.044***		-0.239***
		(0.003)		(0.008)		(0.005)		(0.011)
Constant	26.633***	29.585***	26.047***	30.167***	26.064***	30.407***	25.332***	30.236***
	(0.218)	(0.372)	(0.529)	(0.428)	(0.311)	(0.579)	(0.818)	(0.601)
R^2 (adj.)	0.034	0.056	0.083	0.355	0.052	0.091	0.086	0.40
N	4,879	4,879	4,879	4,879	2,165	2,165	2,165	2,165

Source: Author's calculations.
Notes: Standard errors are reported in parentheses. All regressions also control for Mexican, other Hispanic, and other Race/Ethnicity.
*Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Table 10: Determinants of Obesity, Poisson Regressions, Males Ages 20–65 Years, Full Sample

	NHANES 1999–2006				**NHANES 2003–2006**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black	1.132**	1.081	1.121**	0.976	1.255**	1.114	1.215**	1.004
	(0.064)	(0.061)	(0.064)	(0.049)	(0.105)	(0.085)	(0.103)	(0.070)
Ages 30-39	1.245***	1.282***	1.176**	1.364***	1.308**	1.400***	1.249*	1.435***
	(0.096)	(0.099)	(0.091)	(0.110)	(0.151)	(0.157)	(0.144)	(0.169)
Ages 40-49	1.436***	1.461***	1.344***	1.733***	1.498***	1.570***	1.403***	1.681***
	(0.117)	(0.119)	(0.108)	(0.139)	(0.179)	(0.180)	(0.160)	(0.178)
Ages 50-65	1.479***	1.435***	1.324***	1.947***	1.441***	1.417***	1.290**	1.809***
	(0.129)	(0.121)	(0.112)	(0.154)	(0.166)	(0.146)	(0.143)	(0.168)
Not Born US	0.585***	0.574***	0.559***	0.712***	0.599***	0.594***	0.571***	0.737**
	(0.065)	(0.063)	(0.064)	(0.074)	(0.088)	(0.084)	(0.084)	(0.097)
Light Activity			0.925	0.944			0.952	0.935
			(0.064)	(0.060)			(0.109)	(0.092)
Moderate Activity			0.878	0.830**			0.831	0.775**
			(0.080)	(0.072)			(0.114)	(0.088)
Vigorous Activity			0.811***	0.407***			0.808**	0.390***
			(0.059)	(0.032)			(0.084)	(0.050)
Current Smoker			0.723***	0.871**			0.720***	0.853*
			(0.053)	(0.057)			(0.073)	(0.076)
Former Smoker			1.019	1.003			0.979	1.027
			(0.061)	(0.058)			(0.068)	(0.076)
Stands			0.893	0.791***			0.925	0.780**
			(0.069)	(0.054)			(0.090)	(0.074)
Light Lifting			0.783***	0.605***			0.9043	0.646***
			(0.069)	(0.049)			(0.101)	(0.079)
Heavy Lifting			0.846**	0.428***			0.863	0.358***
			(0.058)	(0.035)			(0.077)	(0.039)
Total Kcal			1.0000	1.0014***			0.9999	1.0015***
			(0.0000)	(0.0001)			(0.0001)	(0.0001)
EI-check		0.994***		0.956***		0.989***		0.948***
		(0.001)		(0.002)		(0.002)		(0.004)
N	6,208	6,208	6,208	6,208	2,805	2,805	2,805	2,805

Source: Authors' calculations.
Notes: Coefficients represent relative risks. Standard errors are reported in parentheses. All regressions also control for Mexican, other Hispanic, and other Race/Ethnicity.
*Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

Table 11: Determinants of Obesity, Poisson Regressions, Males Ages 20–65 Years, Restricted Sample

	NHANES 1999–2006				**NHANES 2003–2006**			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Black	1.137*	1.113	1.141*	0.985	1.226**	1.165*	1.245**	1.081
	(0.081)	(0.078)	(0.079)	(0.064)	(0.109)	(0.103)	(0.104)	(0.089)
Ages 30-39	1.303***	1.344***	1.238**	1.484***	1.461**	1.488***	1.400**	1.602***
	(0.126)	(0.132)	(0.117)	(0.152)	(0.208)	(0.211)	(0.194)	(0.212)
Ages 40-49	1.534***	1.552***	1.479***	1.963***	1.704***	1.721***	1.634***	1.948***
	(0.157)	(0.157)	(0.145)	(0.196)	(0.257)	(0.251)	(0.233)	(0.229)
Ages 50-65	1.508***	1.471***	1.480***	2.271***	1.656***	1.643***	1.592***	2.284***
	(0.150)	(0.145)	(0.134)	(0.213)	(0.237)	(0.216)	(0.228)	(0.279)
Not Born US	0.594***	0.579***	0.598***	0.752***	0.654***	0.656***	0.644***	0.839
	(0.070)	(0.067)	(0.071)	(0.079)	(0.099)	(0.094)	(0.099)	(0.120)
Light Activity			0.985	0.968			0.947	0.924
			(0.078)	(0.066)			(0.122)	(0.108)
Moderate Activity			0.888	0.808**			0.841	0.757*
			(0.094)	(0.083)			(0.129)	(0.106)
Vigorous Activity			0.773***	0.369***			0.737**	0.330***
			(0.064)	(0.036)			(0.092)	(0.055)
Current Smoker			0.716***	0.873*			0.769*	0.904
			(0.062)	(0.069)			(0.105)	(0.112)
Former Smoker			0.987	0.952			1.055	1.072
			(0.072)	(0.065)			(0.098)	(0.102)
Stands			0.851*	0.746***			0.858	0.749***
			(0.080)	(0.060)			(0.097)	(0.076)
Light Lifting			0.734***	0.561***			0.810*	0.574***
			(0.075)	(0.053)			(0.098)	(0.072)
Heavy Lifting			0.762***	0.377***			0.791**	0.321***
			(0.075)	(0.039)			(0.091)	(0.045)
Total Kcal			1.0002***	1.0014***			1.0002*	1.0017***
			(0.0000)	(0.0001)			(0.0001)	(0.0001)
EI-check		0.992***		0.953***		0.988***		0.944***
		(0.001)		(0.002)		(0.003)		(0.005)
N	4,879	4,879	4,879	4,879	2,165	2,165	2,165	2,165

Source: Authors' calculations.
Notes: Coefficients represent relative risks. Standard errors are reported in parentheses. All regressions also control for Mexican, other Hispanic, and other Race/Ethnicity.
*Statistically significant at the .10 level; **at the .05 level; ***at the .01 level.

A Data Appendix

This appendix describes the construction of variables used in the empirical analysis. All variables are constructed using data from the NHANES, survey years 1999–2006. Methods used conform to those in the NHANES analytical guidelines, available at
http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/nhanes_analytic_guidelines_dec_2005.pdf.

A.1 Body Mass Index (BMI)

We calculate BMI as examined weight in kilograms divided by the square of examined height in meters. We determine obesity status using the standard criterion of BMI greater than or equal to 30.

A.2 Physical activity

We construct two discrete variables to capture physical activity levels: a variable that describes typical daily activities including (but not limited to) work-related tasks, and a variable that measures leisure-time physical activity (LTPA). The former variable consists of responses to a survey question asking subjects to characterize their "usual daily activities" in the course of paid work, housework, attending classes if a student, or the typical activities of retired persons.[37] Available responses include "mostly sitting," "mostly standing or walking with little heavy lifting," "lifting light loads and/or stair-climbing," and "lifting heavy loads or other heavy work." (Subjects must choose a single best response.) The leisure-time activity variable draws on a series of questions pertaining to "moderate" and "vigorous" leisure-time physical activities. "Moderate" activities are described as those that cause light sweating or a slight-to-moderate increase in breathing or heart rate, and "vigorous" activities are those that cause heavy sweating or large increases in breathing or heart rate. For each of these categories, respondents are first asked whether they engaged in any such activity for at least 10 minutes during the past 30 days; if so, they report the frequency and typical duration of each qualifying activity. For example, a subject could report running (a vigorous activity) three times per week for 30 minutes each in a typical week as well as swimming vigorously once a week for 45 minutes, and analogously for moderate physical activities.

We take the set of LTPA-related responses for a given individual and aggregate these into a measure of total intensity-weighted LTPA per month using the concept of MET-minutes. A MET, or metabolic equivalent, is defined as the ratio of energy expended in an activity to energy expended at rest (but not asleep).[38] MET-minutes are calculated as the product of minutes spent in an activity and the MET value of the activity, where 30 minutes of a 2-MET activity, such as washing dishes, represents the same total MET-minutes (60) and the same total caloric

[37]Beginning in 2007, this question was replaced with a series of questions eliciting time spent in "vigorous work" and "moderate work" activities, including housework, yard work, and work done away from home.

[38]MET values are based on a formula for resting energy expenditure per kilogram of body weight per unit of time. Because individual energy expenditure at rest may deviate from the formula, METs must be understood as approximations of true energy intensity.

expenditure (for a given person) as 15 minutes of a 4-MET activity, such as walking briskly.[39] Using the recommended MET values for activities provided in the NHANES data, together with the frequency and duration data, we compute total MET-minutes (per month) across all moderate and vigorous leisure-time physical activities. We then assign each individual into one of four levels of leisure-time physical activity: "sedentary" individuals are those with zero MET-minutes of LTPA, meaning they responded "no" to the initial yes/no questions about minimal participation in physical activity; the "low activity" category includes those with up to 1999 MET-minutes per month of LTPA; the "moderate activity" category includes those with at least 2,000 MET-minutes per month but less than 4,000; and the "vigorous activity" category includes those with 4,000 or greater MET-minutes per month of LTPA. These categories correspond roughly to the four levels of physical activity described in the U.S. Department of Health and Human Services' physical activity guidelines for adults.[40]

A.3 Smoking

The NHANES surveys contain self-reported information on both past and present smoking activity. Based on this information, we construct a smoking variable with three categories: "nonsmoker," for those individuals who report having smoked fewer than 100 cigarettes in their lifetime; "current smoker," for those who smoked at least 100 cigarettes in their lifetime and also reported smoking regularly at the time of the interview; and"former smoker," for those who smoked at least 100 cigarettes in their lifetime but were not still smoking regularly at the time of the interview.

A.4 Dietary intake

In the NHANES data from 1999–2006, a subset of survey participants reported extensive information on food consumption using the 24-hour recall method, in which individuals are asked to describe every food and beverage item that they consumed in the previous 24-hour period. Based on these reports, the survey uses standardized nutrition information to compute total kilocalories consumed per day.[41] For the years 1999–2002, in interviews conducted at the mobile examination center, respondents supplied information on food intake for at most a single day. Beginning in 2003, the survey attempts to collect food intake data for a second (non-consecutive) day in a follow-up phone interview. "Day 1" intake refers to food intake data collected in the examination center, regardless of whether the individual also provided "Day 2" intake data over the phone. In our baseline regression models, encompassing data from 1999–2006, the variable "total kcal" represents Day 1 caloric intake only. In additional models estimated for robustness using the

[39]However, the same MET-minutes may entail different levels of caloric expenditure for different individuals based on BMI, gender, and other factors.

[40]The threshold of 2,000 MET-minutes per month (the low end of our "moderate" category) corresponds roughly to the minimal physical activity requirement for adults set forth in these guidelines. The threshold of 4,000 corresponds to twice that amount, a level which the agency's guidelines indicate would provide "additional and more extensive health benefits." See http://www.health.gov/paguidelines/pdf/paguide.pdf for details.

[41]In colloquial speech, the term "calorie" refers to a kilocalorie (kcal for short). We will use the terms interchangeably. Quantities consumed of specific food components, such as carbohydrates and fats, are also provided in the data. Models involving specific nutrients are suppressed in the interest of brevity, as the results do not add substantively to the questions of interest.

2003–2006 subsample, "total kcal" represents the individual's average caloric intake across Day 1 and Day 2.[42]

A.5 Race

NHANES 1999–2006 contains five racial/ethnic categories: non-Hispanic white, non-Hispanic black, Mexican, other Hispanic, and "other," which embeds Asian-Americans and all other identities.[43] In the regression analysis, we include dummy variables for each racial/ethnic category, letting whites be the omitted group, such that our reported contrasts pertain explicitly to blacks versus whites. Effects for other racial/ethnic groups are suppressed in the results tables.

A.6 Foreign-born status

The NHANES data indicate whether or not each participant was born in the United States. One's country of origin (United States or other) is an exogenous factor that may have a significant impact on an individual's BMI. Since we want to assess racial differences in BMI that are not driven by an individual's country of origin, a dummy variable for "not born in the United States" is included as a control in all empirical models.

A.7 Age

NHANES 1999–2006 reports age in years for all interviewees. Within the sample of adults aged 20-65 years, we construct four age groups: 20–29 years, 30–39 years, 40–49 years, and 50-65 years. We include the age category as a control in all regression models to control for racial differences (by sex) in age composition. While the results are not shown, for robustness we test alternative models that are linear and quadratic in age.

[42]NHANES provides two alternate sets of weights that apply, respectively, when using (1) the sample of individuals supplying at least Day 1 data or (2) the sample of individuals supplying both Day 1 and Day 2 data. The weights account for sample selection into data provision as well as the representativeness of the day/s of the week for which food intake was provided. We use appropriate weights accordingly.

[43]We use the race/ethnicity variable named "ridreth1" in NHANES terminology.