Does obesity lead to poor school performance? Estimates from propensity score matching

Hongyun Han

Department of Sociology, University of Wisconsin-Madison

March 26, 2012

ABSTRACT

High body weight is negatively associated with test scores among elementary and middle school students. Are these negative outcomes due to preexisting differences, or are they a casual effect of childhood obesity? To better understand the causal mechanisms underlying this pattern, I use a propensity score matching approach to control for biases from observable preexisting differences, and conduct sensitivity analysis to assess the impact of biases from unobserved variables. Using data from the Early Childhood Longitudinal Study, the matching models reveal that obese eighth graders, on average, score 0.17 standard deviations lower in reading and 0.16 standard deviations lower in math, a reduction roughly equivalent to one sixth of the racial achievement gap. Obesity penalties are larger for girls than for boys in both subjects. Differences between obese and normal-weight children decline slightly after adjusting for missing values. Findings from sensitivity analyses indicate that unmeasured variables would need to increase the odds of becoming obese by at least 20 percent to change the conclusion.

Key words: obesity, academic achievement, propensity score matching

Does obesity lead to poor school performance? Estimates from propensity score matching

Childhood obesity has become a public health crisis in the United States. The rates of obesity among children and adolescents have tripled over the past four decades(Wang and Beydoun 2007). Roughly one in five children and adolescence ages 2 through 19 was obese (Ogden et al. 2010). Treatments for obesity-related conditions in the United States cost roughly $150 billion per year (Cawley 2010). Past research has revealed substantial negative impacts of obesity on public health and the health care system (Finkelstein, Ruhm and Kosa 2005).

Few studies, however, have examined the causal effect of childhood obesity on academic achievement. Although past studies have consistently shown that obesity is associated with lower levels of cognitive function (Li et al. 2008; Miller et al. 2006; Shore et al. 2008), scholars disagree about the impact of obesity on standardized test scores, and associated gender differences in the impacts (Averett and Stifel 2010; Datar, Sturm and Magnabosco 2004; Kaestner and Grossman 2009). Further, the methods employed in previous studies are insufficient to establish a causal effect of obesity on academic achievement—any observed negative effect may be due to preexisting differences rather than a casual relationship. Children who gain excessive weight may come from more disadvantaged families or possess other unobserved characteristics that lead to worse outcomes. For instance, the ability to concentrate may be an unobserved characteristic that affects both weight and school performance; drawing conclusions about the causal relationship between obesity and poor test scores is difficult because of the potential for unobserved characteristics.

To identify the casual effect of childhood obesity on academic achievement, I employ propensity score matching to reduce preexisting differences associated with observed variables. I use a sensitivity analysis to evaluate the strength of the matching estimates against the bias associated with unobserved variables. To alleviate the possibility of reverse causality, I also use predictor variables measured in fifth grade to predict outcomes in eighth grade. Using data from the Early Childhood Longitudinal Study, the matching models reveal that obese eighth graders score, on average, 0.17 standard deviations lower in reading and 0.16 standard deviations lower in math, a reduction roughly equivalent to one sixth of the racial achievement gap. These estimates are robust unless an unmeasured variable would have to increase the odds of becoming obese by at least 20 percent to change the conclusion. Obesity penalties are larger for girls than for boys in both subjects.

This paper makes three important contributions. First, this is the first study to apply propensity score matching to investigate the causal effect of childhood obesity on academic achievement. Second, the analyses include important controls for determinants of obesity (i.e., nutrition and exercise) that have not been used in past studies. Third, this is the first paper to use formal sensitivity analysis in the study of obesity and school performance to evaluate the robustness of matching estimates in the face of selection bias.

POTENTIAL CAUSAL PATHWAYS

Theoretical perspectives focus on four characteristics through which childhood obesity may be associated with poor school performance: behavioral problems, reduced educational expectation, poor work habits and school absence. First, obese individuals have lower levels of self-esteem, and exhibit more behavioral problems that disturb their cognitive development (Falkner et al. 2001; Miller and Downey 1999; Strauss 2000). Second, obese individuals and their parents have lower educational expectations and lower levels of subsequent parental investment (Ball, Crawford and Kenardy 2004; Crandall 1991; Crandall 1995). Third, obese students may also lack attention, concentration, task persistence, and flexibility that are key to effective learning (Rimm and Rimm 2004). Research has firmly established that behavioral problems, reduced levels of educational expectations and poor work habits have a detrimental impact on academic achievement (Bub, McCartney and Willett 2007; Campbell et al. 2006; Fan and Chen 2001; Farkas 2003; McLeod and Kaiser 2004). Thus, if the prevalence of these three factors is higher among obese individuals, it may lead obese children to score poorly on standardized tests.

Fourth, frequent school absences among obese students, due to obesity-related health problems, may hinder the learning process. Obesity is often associated with chronic physical health problems including sleep apnea, asthma, and cardiovascular disease (Daniels 2006). These chronic health problems can lead to fatigue, difficulty concentrating in class, and frequent school absences due to treatment or discomfort (Currie 2009). While obesity is associated with frequent school absences (Geier et al. 2007; Schwimmer, Burwinkle and Varni 2003), chronic health problems can curb school attendance by an average of two more days per year (Bonilla et al. 2005). Thus, if the prevalence of chronic health problems is higher among obese children, and school attendance is crucial to academic succeess (Perez-Chada et al. 2007), obese students, with frequenct school absences, may do poorly in school.

Obesity may affect girls' academic achievement more than boys' achievement. First, heavy girls are more aware of their weight because girls mature earlier than boys (Rimm and Rimm 2004). Second, the degree of stigmatization (such as teasing, and verbal and physical bullying) is higher among obese girls than obese boys (Fikkan and Rothblum 2011; Tang-Peronard and Heitmann 2008). Third, reactions to weight bias are more problematic among girls than boys. A number of studies have consistently found a larger negative impact of obesity on self-esteem for women than men (Miller et al. 2006). Finally, lower educational expectations are more prevalent among girls than among boys. In sum, a stronger degree of stigma and consequently more severe problems suggest a larger obesity penalty in academic achievement for girls than boys.

Despite these initial observations, a few arguments suggest that obesity may be a marker rather than a causal factor. First, any observed negative effect may be due to preexisting differences rather than a casual relationship between childhood obesity and poor school performance. Children who gain excessive weight may come from more disadvantaged families, have poor nutrition, or possess other unobserved characteristics that lead to worse outcomes. For instance, Datar and colleagues (2004) found that obesity differences in first-grade reading scores become insignificant after including socioeconomic and behavioral variables. Similarly, poor nutrition among obese children may diminish their ability to think and concentrate. Children who consume high-sugar drinks may often feel tired because eating sweets leads to a drop in blood sugar. Those who skip meals may not have enough energy for learning (Rimm and Rimm 2004). Further, obese children may have unobserved characteristics, such as a low IQ, that are counterproductive to learning. Additionally, lower test scores may cause excessive weight gain. For some children, poor school performance can be a stressor, causing them to seek comfort from highly caloric foods. Given these considerations, obesity may not have a causal effect on school performance.

PREVIOUS EMPIRICAL RESEARCH

Empirical studies of adolescents have consistently found obesity penalties in school performance and educational attainment, and have revealed larger obesity effects among female students than among male students (Crosnoe 2007; Sabia 2007). In contrast, empirical investigations of younger students have had mixed results, depending on the measures of cognitive development, methods, and age.

Several earlier studies reported a negative effect of obesity on standardized test scores and grade point averages, using ordinary least squares (OLS) regression with cross-sectional data (Datar and Sturm 2006; Judge and Jahns 2007; Shore et al. 2008). However, an OLS approach using contemporaneous measures of obesity and test scores has two limitations when making causal inferences: not only it is impossible to establish the temporal order of obesity and educational outcomes, these analyses also suffer from omitted-variable bias. For example, OLS estimates in the work of Datar and Sturm (2006) and Judge and Johns (2007)were likely biased because the models lacked measures of nutrition, neighborhood, and school characteristics, which may affect both obesity and educational outcomes (Jyoti, Frongillo and Jones 2005). Biases also arise when unmeasured variables, such as intelligence and genetic factors, simultaneously determine both obesity and test scores (Cawley 2004). OLS regressions that control for prior weight (Crosnoe 2007; Datar, Sturm and Magnabosco 2004) alleviate reverse causality to some extent, but remain vulnerable to omitted variable bias.

To control for unobserved genetic factors that may affect both obesity and test scores, some scholars have adopted an instrumental variable approach (Averett and Stifel 2010; Ding et al. 2009; Kaestner and Grossman 2009; Sabia

2007). Kaestner and Grossman (2008) reported no negative effect of obesity on test scores; however, using the same sample of children from the National Longitudinal Study of Youth 1979, Averett and Stifel (2010) identified subtle racial and gender differences in obesity penalties. For instance, Averett and Stifel (2010) found that obese white boys scored approximately one standard deviation lower in reading than their normal-weight counterparts. Three limitations might undermine this causal conclusion. First, the use of either maternal weight or children's prior weight as variables may violate the exclusion restriction, because these variables can affect school performance through characteristics other than obesity, such as birth weight. One recent study found that controlling for actual obesity-related genetic factors removes the negative impact of obesity on the probability of employment among respondents in their mid-twenties (Norton and Han 2008). Second, neither study controlled for nutrition intake or exercise levels, which are crucial determinants of childhood obesity. Third, these studies did not use formal sensitivity analyses to assess the robustness of the instrument variable estimates in the face of selection bias.

Finally, Morris (2007) combined propensity score matching and the instrumental variable approach to examine the obesity gap in adult employment. Use of propensity score matching reduces the biases associated with observed preexisting differences between obese and non-obese individuals; however Morris examined employment rather than educational outcomes, and did not use formal sensitivity analysis to test the robustness of the matching estimates. In short, prior empirical studies are insufficient to establish the causal effects of childhood obesity on academic achievement.

In this study I extend previous investigations by applying propensity score matching and sensitivity analysis to the study of educational outcomes. To establish causal order and alleviate the possibility of reverse causality, I use predictors from fifth grade (or earlier) and outcomes measured in eighth grade. To reduce omitted variable bias, I control for nutritional intake, exercise level, and neighborhood features. The use of propensity score matching also mitigates bias from observed preexisting differences between obese students and their normal-weight counterparts. I conduct Rosenbaum bound sensitivity analysis to evaluate the strength of matching estimates in the face of selection bias.

RESEARCH HYPOTHESES

In this study I investigate the causal effect of childhood obesity on standardized test scores, and explore associated gender differences. Based on the weight stigma and physical health perspectives, I expect that obesity will have a negative impact on reading scores, and math scores. I further expect that due to gender differentials in the degree of stigma and responses to stigma, obesity penalties in academic achievement will be larger among girls than among boys.

METHODOLOGY

Data

I use the kindergarten, fifth-grade, and eighth-grade public-use data from the Early Childhood Longitudinal Study (ECLS-K) for the analysis. The Department of Education sampled 19,000 children enrolled in kindergarten in the fall of 1998, and followed them through eighth grade. The main purpose of the data collection was to track students' academic trajectories. The survey included seven waves of data collection: fall of kindergarten (1998), spring of kindergarten (1999), fall of first grade (1999), spring of first grade (2000), spring of third grade (2002), spring of fifth grade (2004) and spring of eighth grade (2006). The ECLS-K had a multiple-stage sample design, drawing respondents from students within schools located in the primary sampling units.[1]

The ECLS-K kindergarten, fifth-grade, and eighth-grade data is suitable for two reasons. First, the data include complete measures of nutrition, physical activity, family socioeconomic background, neighborhood safety, and school characteristics. Specifically, student-reported nutrition measures are only available for fifth grade. Relatively comprehensive measures not only reduce the likelihood of omitted-variable bias, but also ensure the validity of matching estimates by improving matching quality. Second, estimates based on data from the eighth graders (14-15 years old) can be compared to studies using other datasets in the United States, such as The National Longitudinal Study of Adolescent Health (Add Health), the National Longitudinal Survey of Youth (NLSY) and the Panel Study of Income Dynamics (PSID). Despite these

[1] Detailed information can be found at http://nces.ed.gov/ecls/kindergarten.asp.

advantages, the ECLS-K lacks measures of intelligence, maternal weight, genetic factors, and time-use information, all of which may affect both obesity and test scores. Although a matching approach cannot directly control these unmeasured factors, the sensitivity analysis can reveal the effect of these unmeasured variables on the robustness of the matching estimates.

Sample

The analytic sample consists of 4,460 white children with complete data on item response theory (IRT)-scale test scores and obesity status in fifth and eighth grade.[2] Nearly 39 percent of these children are classified as obese, with a BMI at or above the 95th percentile in both fifth and eighth grade. The case-complete sample includes 2,631 white children. To gauge the enduring effect of obesity, I excluded the 106 obese children whose BMI dropped below the 95th percentile, and the 200 children whose BMI moved above the 95th percentile, between fifth and eighth grade. I used the final sample of 2,631 children with complete data on all covariates in the primary analysis, and supplemented the primary results with an analysis of the imputed sample.

Examining the patterns among the missing covariates revealed that 12 of 20 covariates had missing values for at least some students, and the rate of missing values for four variables (reading scores in kindergarten, math scores in kindergarten, father's occupational status, and free lunch recipient status)

[2] I limit the analytic sample to white children because there are not enough African American and Hispanic children to allow effective matching.

account for the majority (approximately 75 percent) of missing data. Assuming the observations are missing randomly, I use imputation by chained equations implemented via the ICE procedure in STATA (Raghunathan et al. 2001; van Buuren and Oudshoorn 2000). The imputation process yielded five imputed samples of 4,460 cases. Compared to children in the case-complete sample, those with missing values were more likely to receive free lunch, live in unsafe neighborhoods, and have lower math scores in kindergarten.

Method

The analysis consisted of three steps: OLS regression, propensity score matching, and sensitivity analysis. I began with OLS regression, predicting the effect of obesity on test scores, net of all confounding variables. The OLS estimates provide benchmarks for the estimates of the obesity penalty in both reading and math.

To reduce the bias associated with measured covariates, I use propensity score matching to estimate the average treatment effect of obesity on test scores. Propensity score matching assumes that unmeasured variables do not correlate with children’s propensity to be obese or with test scores. The validity of matching estimates relies on the balance in the distributions of covariates between obese students and their normal-weight counterparts. I adopt three strategies to ensure high-quality matching: a clear definition of the treatment and control groups, controls for comprehensive covariates, and the adoption of various matching regimes (Morgan and Harding 2006). The first strategy was to implement clear definitions of the treatment and control groups. I define the treatment category as "being obese" if a child's BMI was equal to or above the 95^{th} percentile of the BMI z-score distribution in both fifth and eighth grade.[3] The control group includes children whose BMI z-scores were between the 5^{th} and the 75^{th} percentiles of the distribution.[4] The educational outcomes are IRT-scale test scores in reading and math in eighth grade. Hence, the causal question is: Does obesity depress test scores among eighth graders?

The second strategy I used to ensure high-quality matching was to control for confounding variables. Three theoretical considerations guided the selection of covariates: (1) all confounding were derived from theories and endorsed by empirical evidence; (2) all confounding variables preceded and determined the probability of being obese, those that are consequences of being obese were excluded (Morgan and Winship 2007); and (3) all confounding variables determined test scores (Angrist and Hahn 2004). In accordance with these three rules, I followed the ecological theory of childhood obesity (Procter

[3] The number of children in the treatment and control groups remained relatively stable (>90 percent) between fifth grade and eighth grade. Only 106 of 1,400 obese children lost weight and moved into the normal-weight group, while 200 of 3,000 normal-weight children gained weight and moved into the obese group. To test the sensitivity of the treatment definition, I also examined whether being extremely obese ($\geq 97^{th}$ percentile of BMI z-scores) was more harmful to test scores than being obese ($\geq 95^{th}$ percentile of BMI z-scores) .

[4] This narrow definition of the control group is more meaningful than the more commonly used category (5^{th} -85^{th} percentile) for two reasons. First, I defined the control group based on a plot of the quadratic relationship between BMI z-scores and test scores. Second, large variations in the quadratic relationship between the 75th and 85th percentile of the BMI distribution suggest differences within this group. Despite these advantages, I also used the alternative definition of normal weight (5^{th}-85^{th} percentile) to test the sensitivity of matching estimates based on the narrowly defined control group (5^{th}-75^{th} percentile).

2007) and identified a set of individual, social, cultural, and environmental determinants of obesity, including birth weight, food and drink intake, exercise, family socioeconomic background, neighborhood safety, and school characteristics. These confounders occur at or before third grade. Because the treatment occurs in fifth grade, and outcomes occur in eighth grade, this selection not only meets the criteria of matching, but also alleviates the possibility of reverse causality.

The third strategy I adopted to ensure high-quality matching was to use various matching regimes. Because matching estimates may vary with changes in the selection criteria for control and treatment cases, the consistency of matching estimates across multiple matching regimes indicates robustness (Morgan and Harding 2006). After identifying the obesity penalty via matching model that achieved the best balance in the distribution of covariates between the treatment and control groups (Sekhon 2009),[5] I compared the estimated average treatment effect on the treated group across eight matching regimes. These matching regimes include nearest-neighbor matching with four variants in replacement and ratio, stratified matching, full matching, and optimal matching (Hansen 2004; Ho et al. 2007; Rosenbaum 1989). In addition to this

[5] The final matching model includes covariates, square terms, and interaction terms. I tried a variety of combinations of high-order and interaction terms before choosing this final matching model. I used the Kolmogorov-Smirnov Bootstrap p-values and the Q-Q plot to evaluate the balance in the distribution of the confounders between the treatment and control groups. I selected the matching regime that achieved the best balance in the distribution of the covariates with a fair number of matched cases, realizing that there is a tradeoff between the precision of the matches and the number of matched cases.

comprehensive approach, I also evaluated the consistency of matching estimates with estimates for severe obesity ≥97th percentile), estimates using an alternative definition of the control group (5th -85th percentile), and estimates determined after the adjustment of missing values with multiple imputation.

To assess the robustness of the matching estimates in the face of bias related to unobserved variables, I adopted three methods, including regression adjustment of the matched data (Abadie et al. 2001), adding a pre-treatment outcome (Smith and Todd 2005), and conducting Rosenbaum bound sensitivity analysis (DiPrete and Gangl 2004). To adjust matching estimates via regression, I useed the Zelig program with the least squares model for continuous variables (Imai, King and Lau 2007), and included a reading test score in kindergarten as a pre-treatment outcome to make the difference-in-difference adjustment. Next, I used the Rosenbaum bounds method of sensitivity analysis to reveal how strong the selection bias would need to be to alter the matching estimates (Rosenbaum 2002; Rosenbaum 2005). The analysis proceeded in three steps. First, I specified the strength of the correlation between an unmeasured confounding variable and a student's propensity of being obese as gamma (Γ), expressed as log odds. Then I calculated the upper and lower bounds of matching estimates for each assumed level of Γ. Third, I identified a cutoff point of Γ that rendered the matching estimates statistically insignificant. I conducted the sensitivity tests for reading and math scores based on the Wilcoxon signed-rank test and the Hodges-Lehmann point estimate (Keele 2009). The Rosenbaum bounds approach represents the “worst-case” scenario with respect to the robustness of the matching estimates (DiPrete and Gangl 2004), because the method assumes that there is a strong relationship between an unobservable variable and test scores.

Measurement

The outcomes in this study include IRT-scale test scores in reading and math in eighth grade. The reading test measures students’ skills in nine dimensions: letter recognition, beginning sounds, ending sounds, sight words, comprehension of words in context, literal inference, extrapolation, evaluation, and critical evaluation of literal works. The math test evaluates students’ skills in number and shape, relative size, ordinal and sequence, addition and subtraction, division and multiplication, place value, rate and measurement, fractions, area, and volume. Reading scores have a mean of 171 points and a standard deviation of 27.6 points; math scores have a mean of 142 points and a standard deviation of 22 points. Both scores are slightly skewed toward the bottom. I convert original scores to standard deviation units to facilitate comparisons across studies. The treatment is being obese ($\geq 95^{th}$ percentile of the BMI z-score distribution) in both fifth and eighth grade, and the control group consists of normal-weight children (5^{th}-75^{th} percentile of the BMI z-score distribution).

To improve matching quality and alleviate the problem of reverse causality, I controlled covariates measured in or before third grade. That is, confounders occur before the treatment in fifth grade and the outcome in eighth grade. All covariates are measured via retrospective parental report, except soda consumption, which was reported by students. Birth weight is a continuous variable based on parental report in kindergarten. Enrollment in a free or reduced-price lunch program in school, and weekly soda consumption are two measures of nutritional intake. For weekly soda consumption, I converted the original value range (0, 1-2 times per week, 3-4 times per week, 1 time per day, 2 times per day, 3 times per day, 4 or more times per day) into a continuous variable, by taking the mid-point of each value range (0, 1.5, 3.5, 7, 14, 21 and 30 times per week). Sedentary and active behaviors measure the level of physical activity. Sedentary behavior is indicated by the total hours of viewing television, videotapes, or DVDs per week. Active behavior is measured by the number of days in a typical week children get twenty or more minutes of exercise vigorous enough to cause rapid breathing, perspiration, and a rapid heartbeat. Family socioeconomic status includes maternal education, paternal occupation, and family income. Maternal education is measured by years of schooling and father's occupation is measured by percentage of college graduates among jobholders in a specific occupation.[6] Family income is the total income of all persons in a child's household, including salaries, interest, retirement, and other sources. Neighborhood safety is a dummy variable indicating that a neighborhood is deemed safe for children to play outside. School type is a dummy variable indicating a student attends public school. A pre-treatment

---

[6] This measure follows Hauser's (2008) strategy.

variable, reading IRT-scale scores in kindergarten, measures students' reading ability at school-entry.

RESULTS

Observed obesity penalty and gender differences

In this section I review the descriptive analysis of the relationship between BMI z-scores and test scores in standard deviation units. Results in Figure 1 show a quadratic relationship between BMI z-scores and eighth-grade test scores for both reading and math. Obese children (≥ 95th percentile) generally score below the mean, while normal-weight children (5th-75th percentiles) fluctuate around the mean.[7] Results in Table 1 show that the average difference between the two groups is 0.35 standard deviations in reading and 0.29 standard deviations in math, without adjusting any covariates (Row 1 and 4 of Table 1). These differences are statistically significant per a t-test (p-values <0.00). The results for girls reveal an even more striking pattern. Obese girls have an average loss of 0.43 standard deviations in reading scores, approximately 50 percent larger than the loss among boys (Row 2 and 3 of Table 1). Together, Figure 1 and Table 1 reveal considerable differences in reading and math test scores between obese children and their normal-weight counterparts.

[7] Fluctuations in the relationship between BMI and test scores for children between the 75th and 85th percentile BMI z-scores suggest that the often-used categorical measure of normal weight (5th-85th percentile) does not sufficiently reflect the obesity effect on academic achievement.

It is essential to evaluate preexisting differences in confounders between the control and treatment groups before matching. Table 1 shows Kolmgorov-Smirnov (KS) Bootstrap p-values by comparing the distributions of covariates used in genetic matching[8] (Sekhon 2009). The variables are ordered by types of determinants of obesity. Clearly, there are significant differences between obese and normal-weight children before matching with regard to gender, birth weight, nutrition, physical activity, parental social status and reading ability in kindergarten. Compared to their normal-weight counterparts, obese children, on average, drink more soda, watch more television, and do less intensive exercise; their mothers have fewer years of schooling and their fathers hold less prestigious jobs. Further, these preexisting differences persist in the imputed sample after adjusting for missing values (results not shown). Without adequately controlling for these preexisting differences, the OLS regression may yield biased estimates of the causal effect of obesity on academic achievement.

Obesity penalty in test scores: matching estimates

Does childhood obesity cause poor school performance? The weight stigma and physical health perspectives imply that obese children tend to earn lower test scores than their normal-weight peers. To test these theoretical arguments, I

[8] Genetic matching is a multivariate matching method that uses an evolutionary search to maximize the balance of observed covariates across matched treated and control units. It uses a search algorithm to iteratively check and improve covariate balance, thus it eliminates the need to manually and iteratively check the propensity score. and it is a generalization of propensity score and Mahalanobis Distance matching.

conducted OLS regression models and propensity score matching. Table 2 presents the estimated effects of obesity derived from OLS regression results and genetic matching[9] for the primary sample (N=2,631).

Results from the OLS regression generally support the first hypothesis, that obesity is negatively associated with test scores among eighth graders. Column 1 of Table 2 shows that, compared to their normal-weight peers (BMI z-score is between 5th and 75th of the distribution), obese children have an average loss of 0.105 standard deviations in reading and 0.062 standard deviations in math in eighth grade. These estimates must be interpreted cautiously, however, as the case-complete sample includes children whose chances of gaining weight vary substantially. The estimates could also be biased if there are unmeasured variables that determine both the propensity of being obese and poor school performance.

To minimize bias related to observable variables, I use propensity score matching to adjust an individual's propensity of being obese, and to reduce the preexisting differences between obese students and their normal-weight counterparts. As shown in the Table 1 and Appendix Figure 1, among a number of matching regimes applied, genetic matching yields the best balance in the

[9] Before reporting the matching estimates, it is necessary to evaluate the matching quality. Genetic matching significantly improved the balance in the distribution of covariates between obese and normal-weight children. Table 1 compares the standardized bias of the covariates before and after genetic matching for the primary sample. The KS Bootstrap p-values for all covariates and associated interaction/ high-order terms are at or above the 0.05 significance level, therefore, the distributions of all covariates are balanced between obese and normal-weight children.

distributions of covariates between obese students and their thinner peers. That is, obese students are similar to their thinner peers in terms of age, gender, birth weight, nutritional intake, physical activity, family socioeconomic background, neighborhood, school and initial reading ability. Results in the Column 3 of Table 2 show that, once obese students were similar to their thinner peers, they on average scored 0.17 standard deviations lower in reading and 0.16 standard deviations lower in math. Results from the genetic matching model reveal larger obesity penalties than the OLS regression, because matching excludes hundreds of normal-weight students whose chances of being obese are different from their obese peers. The obesity penalties found in this study are consistent with prior reports using data from the Add Health and the children sample of the NSLY 1979(Averett and Stifel 2010; Crosnoe and Muller 2004; Sabia 2007).

How large are the obesity gaps? In eighth grade, African American students on average scored 26.7 points or 0.97 standard deviations lower in reading than their white counterparts. The black-white gap in math was 21.3 points or 0.97 standard deviations. Thus, an obesity gap of 0.17 standard deviations in reading is equivalent to 17 percent of the racial achievement gap in eighth grade, and the obesity penalty in math is approximately one sixth of the racial achievement gap. Yet, these obesity penalties are smaller than those reported by Averett and Stifel (2010), probably due to differences in age groups, sample coverage and covariates between the two studies.

To further test the consistency of the matching estimates in the context of potential bias from observed variables, I compared the estimated obesity penalties across eight matching schemes ranging from strict to lenient distance choice.[10] Figure 2 depicts the estimated obesity penalties in reading and their 95 percent confidence intervals across the eight matching regimes. Overlapping confidence intervals suggest marked consistency among the estimated obesity penalties in reading. These estimates range from a loss of 0.09 to 0.17 standard deviations. The stricter the selection criteria, the fewer control cases selected, and the greater the deviation between the estimates. Figure 3 shows similar variations in the case of math scores, which range from a decrease of 0.09 to 0.15 standard deviations. Taken together, the small fluctuations in the obesity effects across the eight matching schemes indicate that the estimated obesity penalties in reading and math are fairly consistent.

Sensitivity analysis of the obesity penalty in test scores

Despite the consistent obesity penalty in the matching models, the validity of the matching estimates may be questionable if they are affected by unobserved covariates. To explore the robustness of the matching estimates in the context of

[10] The estimated treatment effects were calculated based on eight matching regimes: nearest neighbor (NN) without replacement (control-to-match ratio=1), NN with replacement (control-to-match ratio=2, caliper=0.25), NN with replacement (control-to-match ratio=2, caliper=0.5), stratified matching, full matching, optimal matching, genetic matching (control-to-match ratio=1), and genetic matching (control-to-match ratio=2). Full matching is a special form of stratified matching, and genetic matching is a special form of optimal matching (Ho et al. 2007).

bias associated with unobserved variables, I calculate the Rosenbaum bounds for the estimated average treatment effects of obesity in reading and math, and report the results in Table 3. Column 1 of Table 3 reflects the assumed odds ratio of being obese (Γ) associated with an unmeasured variable (such as intelligence). Columns 2-4 of Table 3 show the lower and upper bounds of the Hodges-Lehmann estimates and the maximum p-values for the Wilcoxon signed-rank test. An upper bound of zero or a p-value above 0.05 indicates a critical level of Γ that renders the matching estimates invalid.

Results from Table 3 reveal that the estimated treatment effect of obesity on reading scores among eighth graders is relatively robust to biases related to unmeasured variables. To illustrate the results, consider intelligence, an unmeasured covariate that may simultaneously determine a student's propensity of being obese and school performance. If the student's intelligence is not associated with his chance of being obese (Γ=1), the estimated average treatment effect of obesity on reading (-0.17 standard deviations) from the random experiment remains valid. However, the negative impact of obesity would become insignificant if intelligence elevated the odds of becoming obese by 20 percent (Γ=1.20), judging by the 0.07 p-values from the Wilcoxon signed-rank test. In the case of the Hodges-Lehmann test, intelligence would have to increase the odds of becoming obese by 10 percent (Γ=1.10) for the upper bound of the Hodges-Lehmann estimates to move near zero. Further, in the case of math, the Hodges-Lehmann test requires that intelligence boosts the student's odds of becoming obese at least by 15 percent. Notably, the Rosenbaum bounds are a “worst-case” scenario in which an unmeasured confounding variable correlates strongly with test scores (DiPrete and Gangl 2004). Thus, it is reasonable to conclude that the causal effects of obesity on reading and math test scores for eighth graders are relatively robust against selection bias.

What does a Γ value of 1.20 mean in practice? To illustrate its magnitude, I express the selection bias designated by specific levels of Γ in terms of the equivalent effects of observed covariates on treatment assignment from the propensity score model,[11] following the strategy described in DiPrete and Gangl (2004). Columns 5-7 of Table 3 present the selection bias equivalent values. The critical level of Γ= 1.20 is attained at a difference of 0.78 pounds of birth weight, 3.04 hours of television viewing, or 2.03 days that include 20 minutes of intensive exercise each week. Thus, the estimated causal impact of obesity on reading and math scores would be questionable if an unmeasured confounder (such as intelligence) affected the treatment assignment to at least this extent.

In conclusion, the estimated average treatment effects of obesity on reading and math among eighth graders determined via propensity score matching models are relatively robust to possible selection bias. Findings from the sensitivity analysis lend support to the causal link between childhood obesity and poor test scores.

[11] Table A1 in the appendix reports the specific estimates in log-odds ratio.

Gender differentials

Compared to obese boys, stronger social stigma and lower educational expectations among obese girls suggest a larger obesity penalty for girls' test scores. To test this hypothesis, I added gender interaction terms to the model to examine whether obesity differences vary by gender. Because the gender interaction terms were highly significant in the OLS regression (p-values<0.05), I divided the primary sample into boys and girls, and present gender-specific matching estimates in Table 4.

As expected, the results reveal clear gender differences in the obesity penalties for test scores. As is shown in Column 2 of Table 4, obese girls, on average, score 0.221 standard deviations lower in reading and 0.214 standard deviations lower in math than their normal-weight peers. For boys, the size of the obesity penalty for math is 20 percent smaller than for girls; the obesity difference on reading scores among boys is insignificant and trivial. The estimated obesity penalties for test scores produced by the imputed values follow a similar pattern (results not shown). Overall, these findings support the second hypothesis that being obese has larger negative impacts for girls than for boys.

Calculations of the Rosenbaum bounds also reveal striking gender differences in the robustness of matching estimates. As shown in Columns 3-4 of Table 4, the matching estimates for boys are fairly weak, oscillating around zero. In contrast, the estimates for girls are relatively robust in the face of selection bias. According to the Wilcoxon signed-rank test, a selection bias of Γ around

1.35 would be necessary to render the negative effects of obesity on reading and math among girls spurious. This magnitude of selection bias equals the impact of the 1.30 pounds of birth weight, 5.0 hours of television viewing per week, or 3.4 days of intensive exercise per week. Similarly, in the case of math, an unobserved measure (such as intelligence) would have to increase the probability of being obese by 30 percent to make the matching estimates invalid. In addition, analysis of the imputed samples shows that, for both boys and girls, the negative impacts of obesity on reading scores are comparable between the case-complete sample and the imputed samples (results not shown). In summary, the striking gender differences in the findings of the matching and sensitivity analyses further confirm the second hypothesis that obese girls suffer more from heavy weight than boys with regard to school performance.

Robustness Check

To further check the robustness of the matching estimates, I assessed the effects of both alternative definitions of normal weight and severe obesity and the adjustment of missing values on the estimated obesity penalties. A high level of consistency of matching estimates across these comparisons will help confirm the validity of these estimates. First, the definition of normal weight (5^{th}-75^{th} percentile) in the primary analytic sample excludes children whose BMI z-scores fall between the 75^{th} and 85^{th} percentile. Therefore, this exclusion may overestimate the obesity penalties in reading and math. Second, the definition of obesity (≥95th percentile) in the primary sample may mask the severity of extreme obesity. If obesity has a negative impact on academic achievement, extremely heavy children (≥97th percentile) are expected to suffer more from excessive weight. Thus, matching estimates using the 95th percentile as a cut-off point may underestimate the impact of extreme obesity. Third, the primary analytic sample excludes students with missing values on covariates. If children who have missing values for television viewing are less likely to be obese, the estimates derived from the primary sample may overestimate the obesity penalty by excluding these children. To test these possibilities, I used the imputed samples and ran the same matching models for the primary analytic sample with alternative specifications of normal weight (5th-85th percentile) and severe obesity (≥97th percentile).[12]

As expected, results in Table 2 show that the estimated obesity penalties in reading and math decline slightly but remain quite large when the alternative specification of normal weight (5th-85th percentile) is used. For instance, obese children, on average, score 0.136 standard deviations lower in reading than their normal-weight peers (5th-85th percentile). This result translates into a roughly 20 percent reduction in the estimated obesity gap, compared to results using the more stringent category of normal weight. In the case of math, changing the definition of normal weight leads to a 5 percent reduction in obesity differences. Overall, these results suggest that the estimated negative impact of

[12]I used the ICE command in STATA to implement multivariate imputation.

obesity on test scores is relatively stable, regardless of the definition of normal weight.

Further, larger impacts of heavy weight among extremely obese children ($\geq 97^{th}$ percentile) are evident in the last column of Table 2. Compared to their normal-weight peers, extremely obese children have an average loss of 0.23 standard deviations in reading and math scores. These estimates indicate a nearly 34 percent increase in the estimated obesity penalty in reading and a 50 percent increase in math. These findings suggest a dose-response relationship between obesity and poor test scores—as the degree of obesity goes up, the extent of the obesity penalty increases accordingly. Thus, these findings lend additional support to the causal effect of obesity on test scores.

Finally, the obesity penalties remain substantial after adjusting for missing values in the confounding variables. Table A2 in the appendix reports the average estimates and associated standard errors for the five imputed samples.[13] Indeed, relative to their normal-weight peers, obese children scored 0.127 standard deviations lower in reading and 0.117 standard deviations lower in math. These obesity penalties are roughly 20 percent lower than those in the complete-case sample. In short, the adjustment of missing values slightly reduces the estimated obesity penalties in reading and math.

[13] I used Rubin's correction method to calculate the average effect of obesity and associated standard errors across the five imputed samples (Allison 2001, p. 30)

DISCUSSION AND CONCLUSION

Childhood obesity is not only a public health crisis, but may also have far-reaching influences on the status attainment process. However, empirical investigations of obesity and school performance have often suffered from selection bias and omitted-variable bias. By applying propensity score matching to data from the Early Childhood Longitudinal Study, I demonstrate that obese eighth graders score, on average, 0.17 standard deviations lower on reading tests and 0.16 standard deviations lower on math tests, differences that equal roughly one sixth of the black-white achievement gap. The estimated harmful effects of obesity on academic performance are relatively robust in the face of hidden bias, and findings from sensitivity analyses reveal that an unmeasured confounder must increase the odds of becoming obese by at least 20 percent to alter the conclusions. Obesity penalties in reading and math test scores are stronger for girls than for boys. The estimated obesity penalties in test scores are relatively consistent across the eight matching regimes; however, the penalties decline slightly after adjusting for the missing values in the confounders and using an alternative definition of normal weight (5^{th}-85^{th} percentile). In sum, these findings support the weight stigma and physical health perspectives whose proponents assert that obesity has a negative impact on academic achievement. The findings provide compelling evidence for the necessity of policy interventions that seek to reduce childhood obesity, such as the “Let’s Move” campaign.

The current findings are consistent with prior reports using data from Add Health (Crosnoe and Muller 2004; Sabia 2007), the CNLSY79 (Averett and Stifel 2010), and earlier waves of the ECLS-K(Cesur and Kelly 2010), as well as a study using genetic markers as instruments (Ding et al. 2009). However, the size of the obesity gaps in reading and math are smaller than those reported by Averett and Stifel (2010), whose estimates of the obesity penalties in math were roughly 0.71 standard deviations. In addition to differences in age groups and sample coverage, one possible reason for the difference is that in this study I controlled for nutritional intake, physical exercise, and neighborhood safety, while these data were not available in the sample used by Averett and Stifel (2010).

In addition, the larger negative impact of obesity among girls is consistent with prior findings using Add Health data (Sabia 2007), and the research using genetic markers as instruments (Ding et al. 2009). However, results differ from those of a study using the first-grade data from ECLS-K, which suggested a larger penalty for boys than for girls (Datar, Sturm and Magnabosco 2004). Two factors may account for larger obesity penalties among girls in the current study.[14] First, this study focuses on the test scores of eighth graders (ages 14-15) while the sample used by Datar, Sturm, and Magnabosco (2004) consisted of first graders (ages 6-7). Prior studies have documented that

[14] I defined the control group as children whose BMI z-scores fell between the 5^{th} and the 75^{th} percentile, a much narrower definition than the control group (5^{th}-85^{th} percentile of BMI z-scores) in Datar, Sturm, and Magnabosco (2004) .

negative stereotypes about obese individuals increase with age (Lerner, Karabenick and Meisels 1975). Therefore, larger obesity effects for girls than boys at older ages are not surprising. Second, differences in estimation methods and sample may partially explain the discrepancies. Datar, Sturm, and Magnabosco (2004) used OLS methods with a variable for prior weight status, whereas the matching method used in this study restricted the sample to a group of normal-weight and obese children who had comparable distributions of confounders. Hence, the current estimates of the average treatment effects of obesity on test scores reveal larger negative impacts of obesity among girls.

Through what mechanisms does obesity lead to poor test scores? Proponents of the social stigma and health problem perspectives have proposed three mechanisms: behavior problems (Crosnoe 2007), low educational expectations (Crandall 1995), and school absenteeism due to obesity-related health problems (Geier et al. 2007). The rich information in ECLS-K allows a test of these mechanisms (results not shown), with the exception of weight discrimination. Although the ECLS-K has no direct measures of weight discrimination, the survey did ask parents to report whether their children were often "picked on" or bullied in eighth grade.[15] Because some bullying behaviors are a form of severe weight discrimination, the data from this question can serve as a proxy measure of the degree of weight discrimination. The preliminary findings support the proposition that obesity operates through all three

[15] It is unclear whether the bullying behavior is specifically weight-based, and whether the bullying is verbal or physical.

mechanisms. Compared to their normal-weight peers, obese children are 22 percent more likely to report being bullied, 22 percent more likely to report being worried, 23 percent more likely to complain of illness, and 14 percent less likely to expect to receive a bachelor's degree. Thus, I expect that school absenteeism, behavior problems, and low educational expectations are potential causal pathways influencing the academic performance of obese students during childhood.

Based on a recent search of the literature, this is the first empirical paper using propensity score matching and sensitivity analysis to examine the causal impact of childhood obesity on academic achievement. The strength of this study lies in three aspects that confirm the validity of obesity penalty estimates. First, results from propensity-score-matching models reveal a causal link between obesity and poor school performance. A relatively comprehensive set of determinants improved the validity of matching estimates in the face of bias from observable variables and reverse causality. In particular, unlike previous studies, I controlled for measures of nutrition, exercise level, and neighborhood safety. Covariates that occurred before treatment have strong explanatory power to predict both the treatment and the outcomes—they explain three fifths of the variation in reading test scores and more than half of the variation in math scores; the impact of the covariates on the probability of becoming obese is statistically significant and substantial in practice (appendix Table A1). Overall, these strategies yield valid matching estimates. Second, I explicitly evaluated the ways selection bias affects the validity of obesity penalties via Rosenbaum bounds. My methodological contribution is to show how large selection on unobserved variables would need to invalidate the entire obesity effect. It shows that the matching estimates remain valid unless an unobserved variable increases the chance of being obese by at least 20%. With a Rosenbaum bounds test, I provide a lower-bound estimate of the obesity effect on reading and math scores. Third, the causal link is further reinforced by the observation of striking gender differentials in obesity penalties, and a dose-response relationship between the degree of obesity and lower test scores. Additionally, matching estimates are insensitive to the specifications of matching regimes, the imputation of missing values, and analysis with alternative specifications of normal weight.

The obesity penalty for academic achievement during childhood has important implications for social stratification at both the individual and the population level. At the individual level, if early onset of obesity negatively affects cognitive development, when obese children grow up to be obese adults (Whitaker et al. 1997), they will likely acquire fewer skills that are highly valued in the labor market and therefore will earn less. A growing body of research has demonstrated that weight accounts for a sizable portion of earning inequality in adulthood, net of intelligence and family socioeconomic background (Judge and Cable 2004; Loh 1993). At the population level, if rates of obesity continue to rise from childhood to adulthood in the United States, and the population consists of a large proportion of obese adults with lower levels of productivity, the overall competitiveness of the labor force will be severely compromised in the near future. Thus, policies that effectively promote physical education and reduce weight stigma are sorely needed to prevent the deleterious impacts of childhood obesity.

This study generates need for future research to advance the scholarly understanding of the causal effects of childhood obesity on academic achievement. First, while propensity score matching and sensitivity analyses provide a way to assess the selection bias, these methods do not completely alleviate biases related to unmeasured variables. Recognizing that the process of becoming obese is not random in observational data, researchers should collect data and control certain measures that are unavailable in the ELCS-K data, such as intelligence, biomarkers, and obesity-related genetic factors. Norton and Han (2008) advanced the field by examining the association between adult obesity and wages and including obesity-related genetic factors, yet further data and studies are needed to improve the scholarly understanding of this relationship. Second, public perceptions of obesity vary by geographic concentration and may change with the rising prevalence of obesity. A recent report shows substantial geographic variability of childhood obesity at the county level (Centers for Disease Control and Prevention 2010). Discovering how these patterns lessen (or elevate) obesity penalties will require further investigation. Third, future studies should also consider potential racial differences in the obesity gap in academic achievement. African American and Hispanic children face less weight stigma within their communities, probably due to disproportionately high prevalence of obesity(Puhl and Heuer 2009), yet they possess fewer financial resources to fight obesity and have reduced educational expectations. Thus, it is crucial to explore racial differentials in obesity penalties with respect to academic achievement. Fourth, it is important to recognize that medical researchers may consider weight status as a continuous or ordinal variable, rather than a binary variable (Imai and van Dyk 2004). Future application of propensity score matching to multiple treatments will advance our understanding of the complex process generating obesity penalty in school performance.

References

Abadie, Alberto, David Drukker, Jane L. Herr, and Guido W. Imbens 2001. " Implementing Matching Estimators for Average Treatment Effects in Stata." *The Stata Journal* 1:18.

Angrist, Joshua, and Jinyong Hahn. 2004. "When to control for covariates? Panel asymptotics for estimates of treatment effects." *Review of Economics and Statistics* 86(1):58-72.

Averett, Susan L., and David C. Stifel. 2010. "Race and gender differences in the cognitive effects of childhood overweight." *Applied Economics Letters*.

Ball, Kylie, David Crawford, and Justin Kenardy. 2004. "Longitudinal relationships among overweight, life satisfaction, and aspirations in young women." *Obesity Research* 12(6):1019-30.

Bonilla, Sheila, Sarah Kehl, Kenny. Y. C. Kwong, Tricia Morphew, Rital Kachru, and Craig A. Jones. 2005. "School absenteeism in children with asthma in a Los Angeles inner city school." *Journal of Pediatrics* 147(6):802-06.

Bub, Kristen. L., Kathleen McCartney, and John. B. Willett. 2007. "Behavior problem trajectories and first-grade cognitive ability and achievement skills: A latent growth curve analysis." *Journal of Educational Psychology* 99(3):653-70.

Campbell, Susan. B., Susan Spieker, Margaret Burchinal, Michele. D. Poe, and The NICHD Early Child Care Research Network. 2006. "Trajectories of aggression from toddlerhood to age 9 predict academic and social functioning through age 12." *Journal of Child Psychology and Psychiatry* 47(8):791-800.

Cawley, John. 2004. "The impact of obesity on wages." *Journal of Human Resources* 39(2):451-74.

Cawley, John. 2010. "The Economics Of Childhood Obesity." *Health Affairs* 29(3):364-71.

Centers for Disease Control and Prevention. 2010. "Geographic Patterns Among Low-Income, Preschool-Aged Children 2006-2008."

Cesur, Resul, and Inas Rashad Kelly. 2010. "From Cradle to Classroom High Birth Weight and Cognitive Outcomes." *Forum for Health Economics & Policy* 13(2):2.

Crandall, Christian S. 1991. "Do heavy-weight students have more difficulty paying for college?" *Personality and Social Psychology Bulletin* 17(6):606-11.

Crandall, Christian S. 1995. "Do parents discriminate against their heavyweight daughters?" *Personality and Social Psychology Bulletin* 21(7):724-35.

Crosnoe, Robert. 2007. "Gender, obesity, and education." *Sociology of Education* 80(3):241-60.

Crosnoe, Robert, and Chandra Muller. 2004. "Body mass index, academic achievement, and school context: Examining the educational experiences of adolescents at risk of obesity." *Journal of Health and Social Behavior* 45(4):393-407.

Currie, Janet. 2009. "Healthy, Wealthy, and Wise: Socioeconomic Status, Poor Health in Childhood, and Human Capital Development." *Journal of Economic Literature* 47(1):87-122.

Daniels, Stephen R. 2006. "The consequences of childhood overweight and obesity." *Future of Children* 16(1):47-67.

Datar, Ashlesha, and Roland Sturm. 2006. "Childhood overweight and elementary school outcomes." *International Journal of Obesity* 30(9):1449-60.

Datar, Ashlesha, Roland Sturm, and Jennifer L. Magnabosco. 2004. "Childhood overweight and academic performance: National study of kindergartners and first-graders." *Obesity Research* 12(1):58-68.

Ding, Weili, Steven F. Lehrer, J. Niels Rosenquist, and Janet Audrain-McGovern. 2009. "The impact of poor health on academic performance: New evidence using genetic markers." *Journal of Health Economics* 28(3):578-97.

DiPrete, Thomas. A., and Markus Gangl. 2004. "Assessing bias in the estimation of causal effects: Rosenbaum bounds on matching estimators and instrumental variables estimation with imperfect instruments." Pp. 271-310 in *Sociological Methodology, Vol 34*. Weinheim: Wiley-V C H Verlag Gmbh.

Falkner, Nicole. H., Dianne Neumark-Sztainer, Mary Story, Robert. W. Jeffery, Trish Beuhring, and Michael. D. Resnick. 2001. "Social, educational, and psychological correlates of weight status in adolescents." *Obesity Research* 9(1):32-42.

Fan, Xitao, and Michael Chen. 2001. "Parental involvement and students' academic achievement: A meta-analysis." *Educational Psychology Review* 13(1):1-22.

Farkas, George. 2003. "COGNITIVE SKILLS AND NONCOGNITIVE TRAITS AND BEHAVIORS IN STRATIFICATION PROCESSES." *Annual Review of Sociology* 29(1):541-62.

Fikkan, Janna, and Esther Rothblum. 2011. "Is Fat a Feminist Issue? Exploring the Gendered Nature of Weight Bias." *Sex Roles*:1-18.

Finkelstein, Eric. A., Christopher J. Ruhm, and Katherine. M. Kosa. 2005. "Economic causes and consequences of obesity." *Annual Review of Public Health* 26:239-57.

Geier, Andrew B., Gary D. Foster, Leslie G. Womble, Jackie McLaughlin, Kelley E. Borradaile, Joan Nachmani, Sandy Sherman, Shiriki Kumanyika, and Justine Shults. 2007. "The relationship between relative weight and school attendance among elementary schoolchildren." *Obesity* 15(8):2157-61.

Hansen, Ben B. 2004. "Full matching in an observational study of coaching for the SAT." *Journal of the American Statistical Association* 99(467):609-18.

Ho, Daniel E., Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2007. "Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference." *Political Analysis* 15(3):199-236.

Imai, Kosuke, Gary King, and Olivia Lau. 2007. "ls: Least Squares Regression for Continuous Dependent Variables." in *Zelig: Everyone's Statistical Software*, edited by Kosuke Imai, Gary King, and Olivia Lau.

Imai, Kosuke, and David A van Dyk. 2004. "Causal Inference With General Treatment Regimes." *Journal of the American Statistical Association* 99(467):854-66.

Judge, Sharon, and Lisa Jahns. 2007. "Association of overweight with academic performance and social and Behavioral problems: An update from the early childhood longitudinal study." *Journal of School Health* 77(10):672-78.

Judge, Timothy. A., and Daniel. M. Cable. 2004. "The effect of physical height on workplace success and income: Preliminary test of a theoretical model." *Journal of Applied Psychology* 89(3):428-41.

Jyoti, Diana F., Edward A. Frongillo, and Sonya J. Jones. 2005. "Food Insecurity Affects School Children's Academic Performance, Weight Gain, and Social Skills." *Journal of Nutrition* 135(12):2831-39.

Kaestner, Robert, and Michael Grossman. 2009. "Effects of weight on children's educational achievement." *Economics of Education Review* 28(6):651-61.

Keele, Luke J. . 2009. "Rbounds: An R Package for Sensitivity Analysis with Matched Data."

Lerner, Richard M., Stuart A. Karabenick, and Murray Meisels. 1975. "Effects of age and sex on development of personal-space schemata towards body build." *Journal of Genetic Psychology* 127(1):91-101.

Li, Yanfeng, Qi Dai, James. C. Jackson, and Jian Zhang. 2008. "Overweight is associated with decreased cognitive functioning among school-age children and adolescents." *Obesity* 16(8):1809-15.

Loh, Eng Seng 1993. "The economic-effects of physical apperance." *Social Science Quarterly* 74(2):420-38.

McLeod, Jane. D., and Karen Kaiser. 2004. "Childhood emotional and behavioral problems and educational attainment." *American Sociological Review* 69(5):636-58.

Miller, Carol T., and Kathryn T. Downey. 1999. "A Meta-Analysis of Heavyweight and Self-Esteem." *Personality and Social Psychology Review* 3(1):68-84.

Miller, Jennifer, John Kranzler, Yijun Liu, Ilona Schmalfuss, Douglas W. Theriaque, Jonathan J. Shuster, Ann Hatfield, O. Thomas Mueller, Anthony P. Goldstone, Trilochan Sahoo, Arthur L. Beaudet, and Daniel J. Driscoll. 2006. "Neurocognitive findings in Prader-Willi syndrome and early-onset morbid obesity." *The Journal of Pediatrics* 149(2):192-92.

Morgan, Stephen L., and Christopher Winship. 2007. *Counterfactuals and causal inference: methods and principles for social research*: Cambridge University Press.

Morgan, Stephen. L., and David. J. Harding. 2006. "Matching estimators of causal effects - Prospects and pitfalls in theory and practice." *Sociological Methods & Research* 35(1):3-60.

Norton, Edward C., and Euna Han. 2008. "Genetic information, obesity, and labor market outcomes." *Health Economics* 17(9):1089-104.

Ogden, Cynthia L., Margaret D. Carroll, Lester R. Curtin, Molly M. Lamb, and Katherine M. Flegal. 2010. "Prevalence of High Body Mass Index in US Children and Adolescents, 2007-2008." *JAMA: The Journal of the American Medical Association* 303(3):242-49.

Perez-Chada, Daniel , Santiago Perez-Lloret, Alejandro J. Videla, Daniel Cardinali, Miguel A. Bergna, Mariano Fernandez-Acquier, Luis Larrateguy, Gustavo E. Zabert, and Christopher Drake. 2007. "Sleep disordered breathing and daytime sleepiness are associated with poor academic performance in teenagers. A study using the pediatric daytime sleepiness scale (PDSS)." *Sleep* 30(12):1698-703.

Procter, Kimberley L. 2007. "The aetiology of childhood obesity: a review." *Nutrition Research Reviews* 20(1):29-45.

Puhl, Rebecca M., and Chelsea A. Heuer. 2009. "The Stigma of Obesity: A Review and Update." *Obesity* 17(5):941-64.

Raghunathan, Trivellore E., James M. Lepkowski, John. Van Hoewyk, and Peter. Solenberger. 2001. "A multivariate technique for multiply imputing missing values using a sequence of regression models " *Survey Methodology* 27(1):10.

Rimm, Sylvia B., and Eric Rimm. 2004. *Rescuing the emotional lives of overweight children: what our kids go through-- and how we can help*: Rodale.

Rosenbaum, Paul R. 1989. "Optimal matching for observational studies." *Journal of the American Statistical Association* 84(408):1024-32.

Rosenbaum, Paul R. 2002. *Observational studies*: Springer.

Rosenbaum, Paul. R. 2005. "Sensitivity Analysis in Observational Studies." Pp. 1809-14 in *Encyclopedia of Statistics in Behavioral Science*, edited by Brian S. Everitt and David C. Howell. Chichester: John Wiley & Sons, Ltd.

Sabia, Joseph J. 2007. "The effect of body weight on adolescent academic performance." *Southern Economic Journal* 73(4):871-900.

Schwimmer, Jeffrey B., Tasha M. Burwinkle, and James W. Varni. 2003. "Health-related quality of life of severely obese children and adolescents." *Jama-Journal of the American Medical Association* 289(14):1813-19.

Sekhon, Jasjeet. 2009. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R." *Journal of Statistical Software* In Press.

Shore, Stuart M., Michael L. Sachs, Jeffrey R. Lidicker, Stephanie N. Brett, Adam R. Wright, and Joseph R. Libonati. 2008. "Decreased scholastic achievement in overweight middle school students." *Obesity* 16(7):1535-38.

Smith, Jeffrey A., and Petra E. Todd. 2005. "Does matching overcome LaLonde's critique of nonexperimental estimators?" *Journal of Econometrics* 125(1-2):305-53.

Strauss, Richard S. 2000. "Childhood obesity and self-esteem." *Pediatrics* 105(1):art. no.-e15.

Tang-Peronard, J. L., and B. L. Heitmann. 2008. "Stigmatization of obese children and adolescents, the importance of gender." *Obesity Reviews* 9(6):522-34.

van Buuren, Stef, and Karin CGM Oudshoorn. 2000. *Multivariate Imputation by Chained Equations. MICE V1.0 User's manual*: TNO Prevention and Health.

Wang, Youfa, and May A. Beydoun. 2007. "The obesity epidemic in the United States - Gender, age, socioeconomic, Racial/Ethnic, and geographic characteristics: A systematic review and meta-regression analysis." *Epidemiologic Reviews* 29:6-28.

Whitaker, Robert C., Jeffrey A. Wright, Margaret S. Pepe, Kristy D. Seidel, and William H. Dietz. 1997. "Predicting obesity in young adulthood from childhood and parental obesity." *New England Journal of Medicine* 337(13):869-73.

Table 1. Summary statistics for two sample t-tests comparing obese to normal weight children before matching for the primary sample (N=2,631)

	Mean		KS Bootstrap p-value[a]	
Variables	Treatment[b]	Control	Before Matching	After Matching
Reading (SD)	-0.17	0.18		
Girls	-0.14	0.29		
Boys	-0.18	0.08		
Math (SD)	-0.06	0.23		
Girls	-0.17	0.17		
Boys	0.01	0.28		
Girl	0.40	0.49	0.000	0.655
Age	11.07	11.08	0.520	0.377
Birth Weight (pounds)	7.66	7.39	< 2.22e-16	0.047
Free or Reduced Priced Lunch	0.36	0.21	0.000	0.157
Weekly Soda Consumption(times)	6.01	5.94	0.307	0.560
Weekly 20-minute Intensive Exercise (days)	3.79	4.14	< 2.22e-16	0.330
Weekly TV viewing (hours)	7.62	6.36	< 2.22e-16	0.502
Family Income (log)	10.54	10.88	< 2.22e-16	0.045
Mother's Years of Schooling	13.12	14.21	< 2.22e-16	0.211
Father's Occupation(% college graduates)	0.18	0.29	< 2.22e-16	0.383
Unsafe Neighborhood	0.20	0.15	0.010	1.000
Public School	0.86	0.78	0.000	0.317
Reading Scores at Kindergarten	28.25	31.70	< 2.22e-16	0.086
Age^2	122.75	122.97	0.520	0.377
Age×Birth Weight	84.84	81.88	< 2.22e-16	0.044
Mother's Education ^{2}	178.88	208.28	< 2.22e-16	0.211
Public School×Age	9.56	8.59	< 2.22e-16	0.416
Public School×Mother's Education	11.23	10.80	< 2.22e-16	0.24
Public School×Father's Occupation	0.14	0.21	< 2.22e-16	0.479

Source:The Early Childhood Longitudinal Study-Kindergarten Eighth Grade Public-Use Data (2006)

Note: [a]The Kolmogorov-Smirnov(KS) Bootstrap p-value measures the balance of the distribution of a continuous covariate between the treatment and the control group in genetic matching (Sekhon 2011). A KS Boot of tests p-value is equal to the T-test p-value for a dummy covariate. In both cases, a p-value below 0.05 indicates statistical significance.

[b]The treatment group is obese children (≥95th percentile) and the control group is normal-weight children (5th-75th percentile).

Table 2. Estimated Average Treatment Effect (ATE) of obesity on eighth graders' test scores in 2006 for the primary sample

	OLS[a]	Genetic Matching[b]		
	Obese (≥95th) vs. Normal (5th-75th)	Obese (≥95th) vs. Normal (5th-85th)	Obese (≥95th) vs. Normal (5th-75th)	Extreme Obese (≥97th) vs. Normal (5th-75th)
Reading	-0.105**	-0.138**	-0.170***	-0.229***
	(0.048)	(0.063)	(0.054)	(0.079)
Math	-0.062	-0.149***	-0.157**	-0.228***
	(0.039)	(0.051)	(0.056)	(0.066)
N	2,631	1,236	1,224	794

Source: The Early Childhood Longitudinal Study-Kindergarten Eighth Grade Public-Use Data (2006)

Note:

[a] The OLS regressions for reading and math control the same covariates as those used in the genetic matching (gender, age, birth weight, reduced lunch, soda consumption, intensive exercise, TV viewing, low income, mother's education, father's occupation, neighborhood safety, school type, age², age*birth weight, momed², public*age, public*college, and public*momed,).

[b] Genetic matching achieves the best balance in the distributions of covariates between treatment and control group (Sekhon2011).

Table 3. Rosenbaum Bound Sensitivity Test for the Obesity Penalty in Test Scores for the Primary Sample (N=1,236)

	Rosenbaum Bounds[a]				Hidden Bias Equivalent[b]
	Gamma (Γ)	Lower. Bound HL Est.	Upper. Bound HL Est.	p-values for Wilcoxon Signed-rank Test	Birth Weight (Ounces)
Reading					
	1.00	-0.15	-0.15	0.00	0.00
	1.05	-0.25	-0.05	0.00	3.36
	1.10	-0.35	0.05	0.01	6.56
	1.15	-0.45	0.15	0.03	9.60
	1.20	-0.55	0.25	0.07	12.48
	1.25	-0.65	0.35	0.15	15.36
Math					
	1.00	-0.16	-0.16	0.00	0.00
	1.05	-0.26	-0.06	0.00	3.36
	1.10	-0.36	0.04	0.01	6.56
	1.15	-0.46	0.14	0.03	9.60
	1.20	-0.56	0.24	0.08	12.48
	1.25	-0.66	0.34	0.16	15.36

Source: The Early Childhood Longitudinal Study-Kindergarten Eighth Grade Public-Use Data (2006)
Note:
[a] Column 1 of Table 3 reflects the assumption about endogeneity in the treatment assignment in terms of the odds ratio of the differential treatment assignment due to an unmeasured covariate. At each level of Γ, I calculate the lower and upper bounds of the Hodges-Lehmann point estimates of the treatment effect in the case of endogenous selection into treatment status, and the bounds for the p-critical value from the Wilcoxon signed rank test. By comparing the Rosenbaum bounds on treatment effects at different levels of Γ, I can evaluate the strength such unmeasured influences must have in order that the estimated treatment effects from propensity score matching would have arisen purely through random assignment.
[b] I calculate the Hidden Bias equivalent with the coefficients derived from logistic regression of obese on covariates, following DiPrete and Gangl (2004). For example, the hidden bias equivalent of $\Gamma = 1.20$ is 12.48 ounces (i.e. 0.78 pounds) increase in birth weight (i.e. $\log(\Gamma) / \beta(\text{birth weight}) = \log(1.20)/0.23 = 0.78$).

Table 4. Matching estimates of obesity penalty and sensitivity analysis for boys and girls (N=2,631).

		Estimates		Sensitivity Analysis	
		OLS[a] β	Genetic Matching ATT	Hodges-Lehmann Point Estimate	Wilcoxon Signed Rank P-Value
Boys					
	Reading	-0.037	-0.042	1.05	1.00
		(0.060)	(0.097)		
	Math	-0.087	-0.161**	1.10	1.10
		(0.053)	(0.077)		
	N	1405	750		
Girls					
	Reading	-0.140	-0.221**	1.15	1.35
		(0.069)	(0.095)		
	Math	-0.093	-0.214**	1.15	1.30
		(0.064)	(0.080)		
	N	1247	446		

Source: The Early Childhood Longitudinal Study-Kindergarten Eighth Grade Public-Use Data (2006)

Note:

[a] The interaction term of gender and obesity in the OLS regressions of reading and math for the entire sample appears to statistically significant.

Figure 1. The quadratic relationships between the BMI z-score and test scores at eighth grade in 2006.

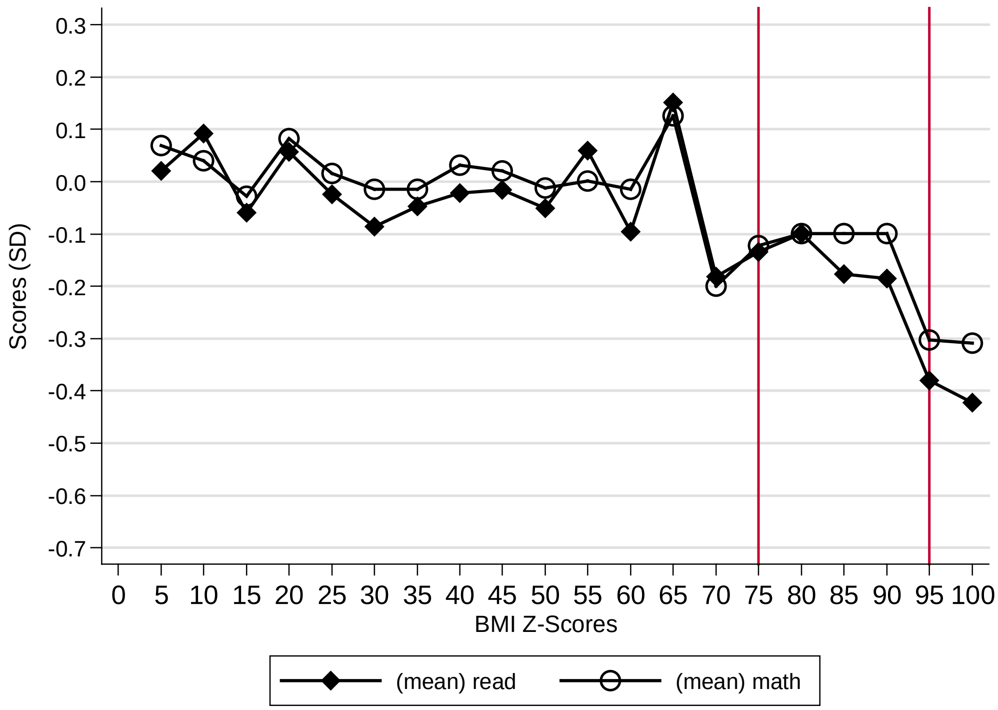


Figure 2. The estimated obesity penalties in reading IRT scale scores (SD) across eight matching regimes.

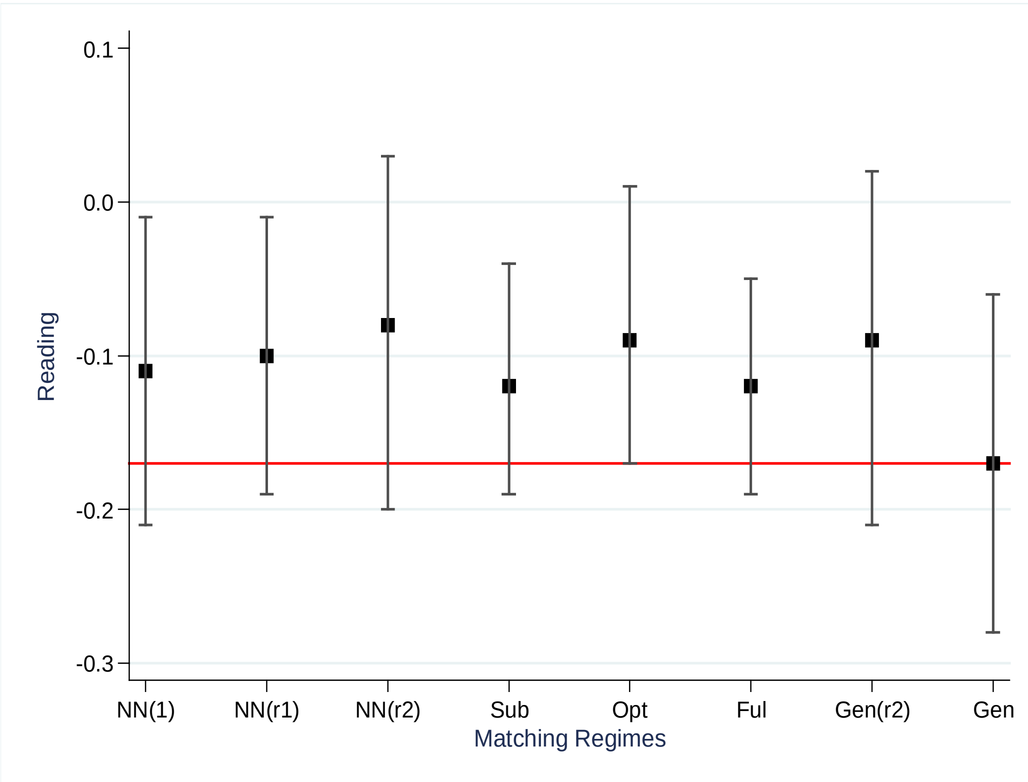
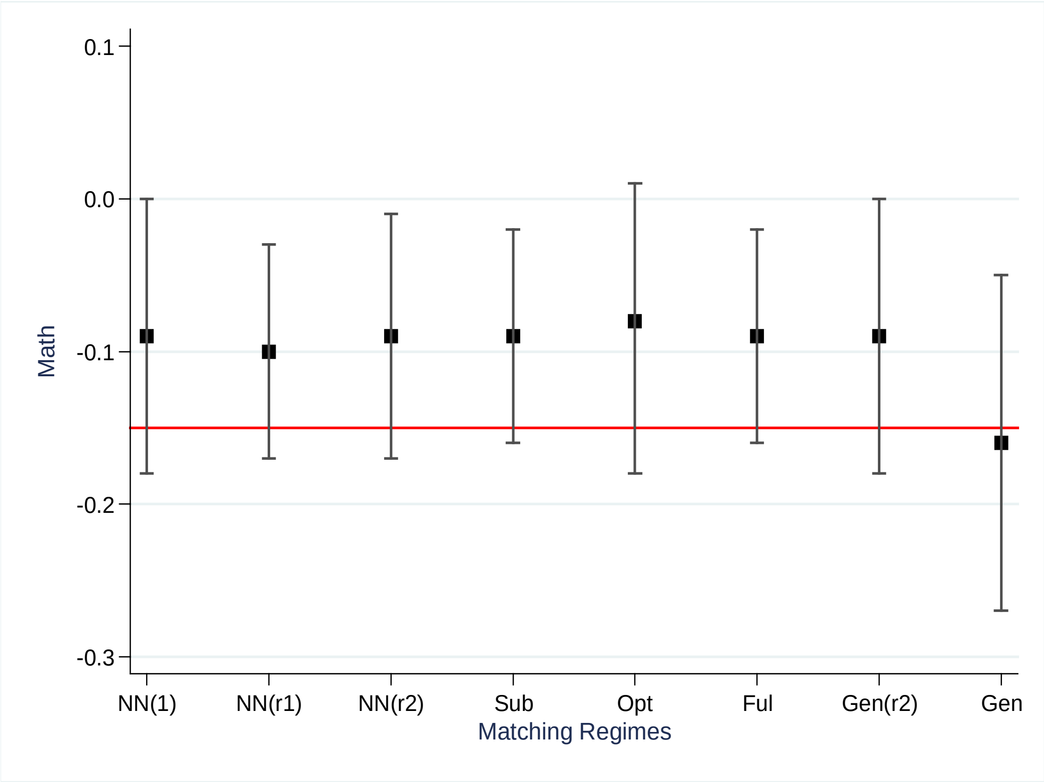


Figure 3. The estimated obesity penalties in math IRT scale scores (SD)across eight matching regimes.



Appendix Table A1. Logistic regression of obese for the case-complete sample.

Explanatory variables	Obese at fifth/eighth grade (≥95th percentile BMI)
Girl	-0.368***
	(0.101)
Age	0.057
	(0.144)
Birth Weight (lbs)	0.223***
	(0.041)
Free or Reduced Priced Lunch	0.213
	(0.131)
Weekly Soda Consumption (times)	-0.011
	(0.007)
Weekly Intensive Exercise (days)	-0.093***
	(0.025)
Weekly TV Viewing (hours)	0.062***
	(0.014)
Family Income (log)	-0.260***
	(0.082)
Mother's Years of Schooling	-0.053**
	(0.024)
Father's Occupation (% college graduates)	-0.914***
	(0.227)
Neighborhood Unsafe to Play	0.032
	(0.134)
Public School	0.269*
	(0.139)
Living in North	0.185
	(0.126)
Reading IRT Scale Score at Kindergarten	-0.023***
	(0.006)
Constant	0.732
	(1.861)
Observations	2652
Log-likelihood	-1307

Source: The Early Childhood Longitudinal Study-Kindergarten Eighth Grade Public-Use Data (2006) .
Note: Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Appendix Table A2. OLS and matching estimates of obesity penalties for the imputed sample[a].

		Estimates			Sensitivity Analysis	
		OLS	Full Matching	Genetic Matching	Hodges-Lehmann Point Estimate	Wilcoxon Signed Rank P-Value
		β	ATE	ATE		
Whole						
	Reading	-0.087**	-0.085**	-0.127**	1.10	1.20
		(0.031)	(0.029)	(0.050)		
	Math	-0.042	-0.067*	-0.117**	1.10	1.20
		(0.030)	(0.030)	(0.044)		
	N	4,460	4,460	2,226		
Girls						
	Reading	-0.149***	-0.146***	-0.187***	1.30	1.30
		(0.045)	(0.041)	(0.070)		
	Math	-0.065	-0.073	-0.098	1.05	1.05
		(0.045)	(0.044)	(0.068)		
	N	2,153	2,153	926		
Boys						
	Reading	-0.049	-0.046	-0.074	1.05	1.00
		(0.043)	(0.041)	(0.065)		
	Math	-0.032	-0.067	-0.108	1.10	1.10
		(0.039)	(0.039)	(0.057)		
	N	2,307	2,307	1,340		

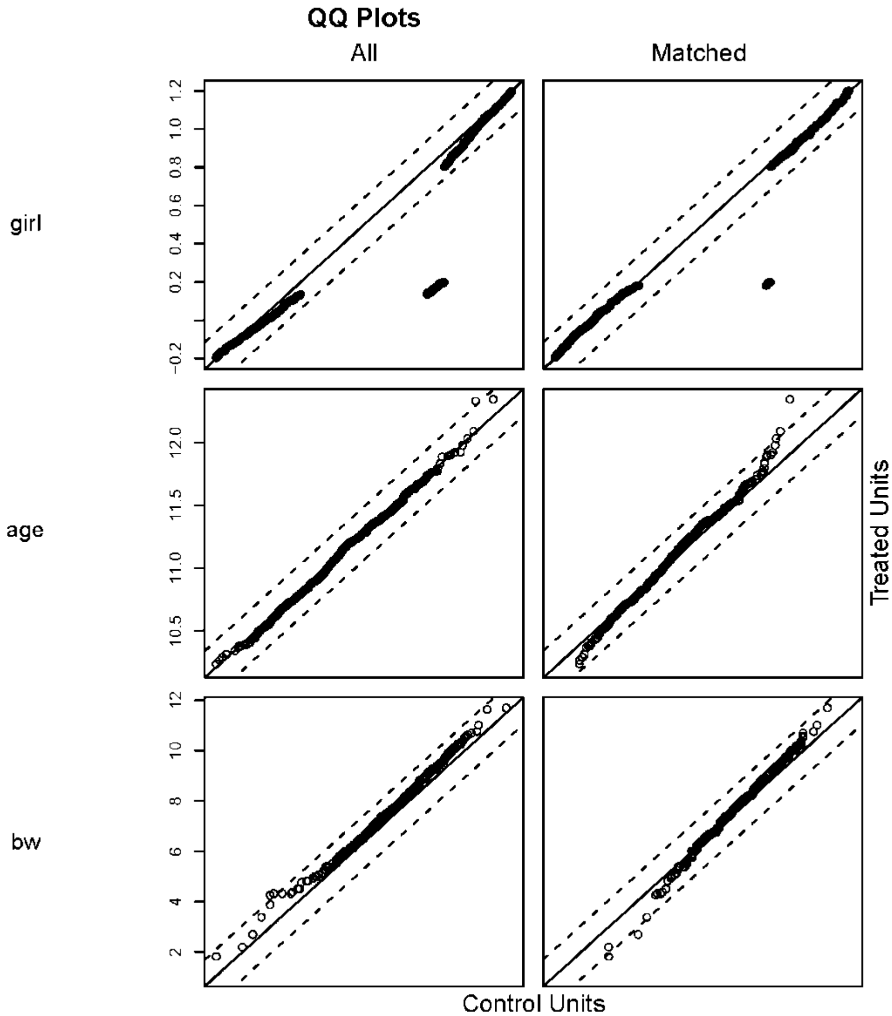
Source: The Early Childhood Longitudinal Study-Kindergarten Eighth Grade Public-Use Data (imputed with ICE) .

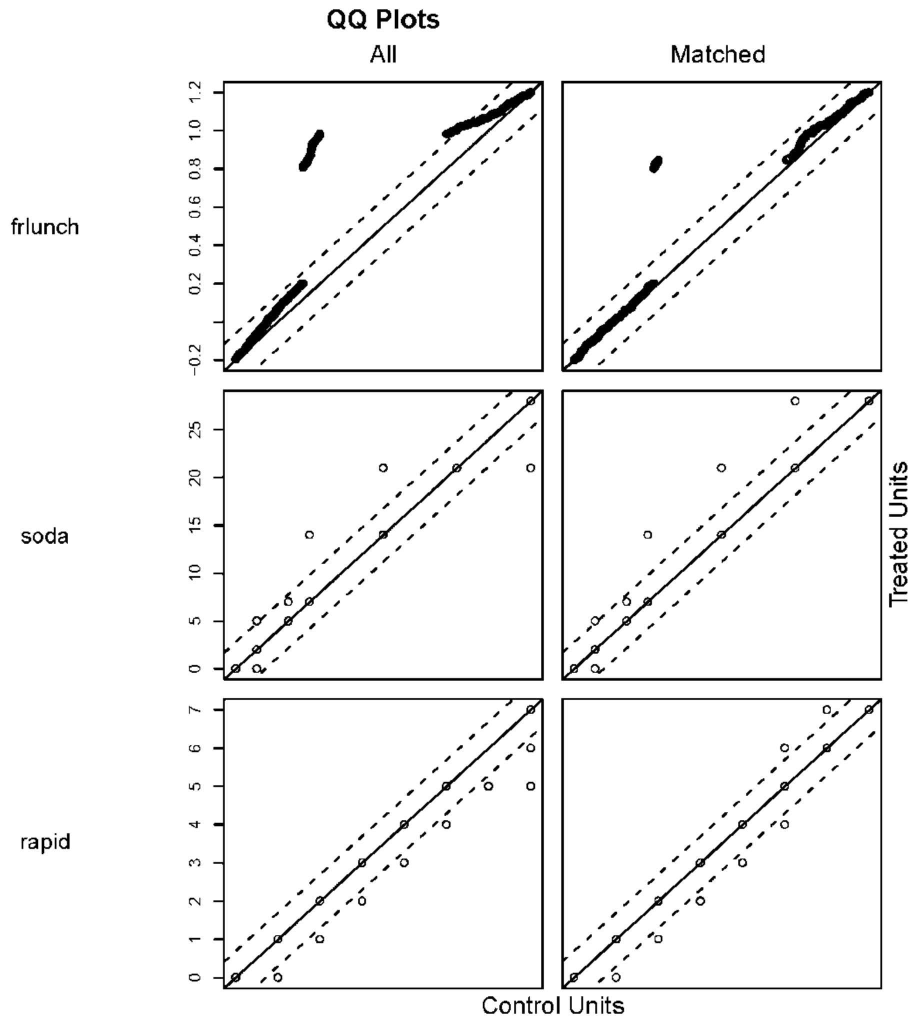
Note:

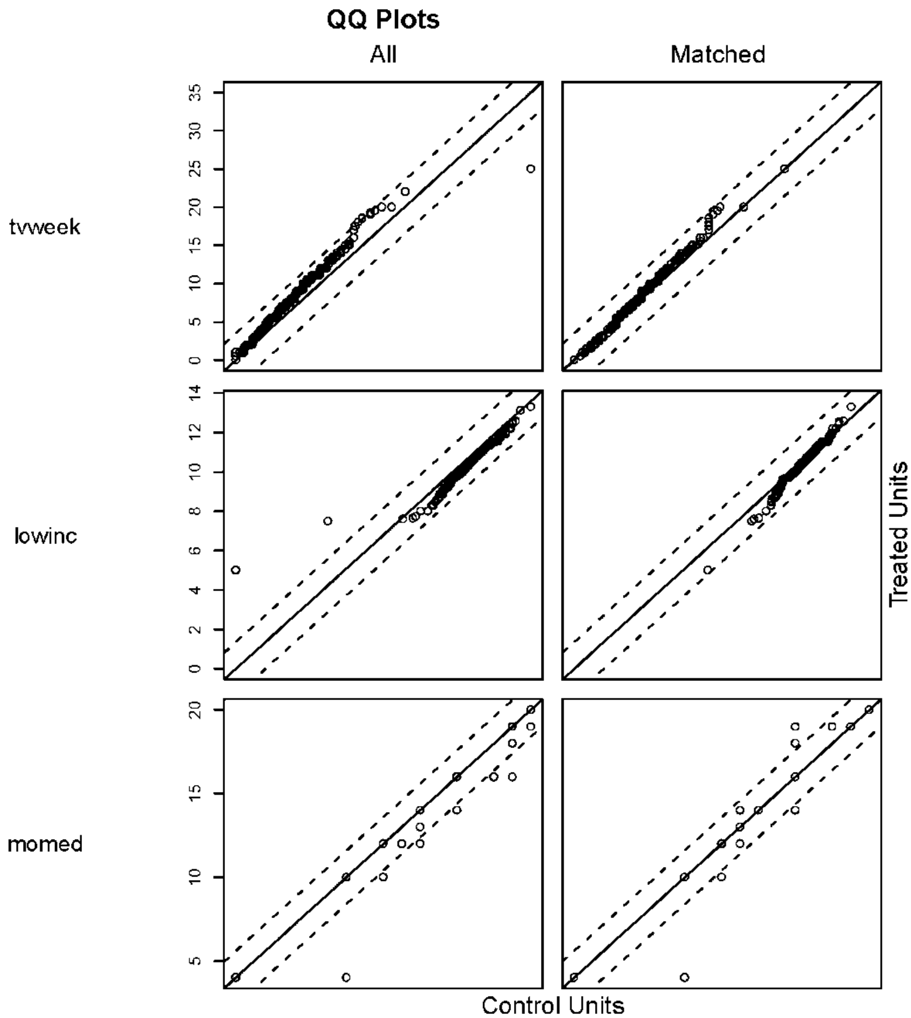
[a] I use the ICE procedure to fill in missing values in the covariates and generate five imputed datasets. I use the Rubin's correction to calculate the average estimates and standard errors across five imputed samples.

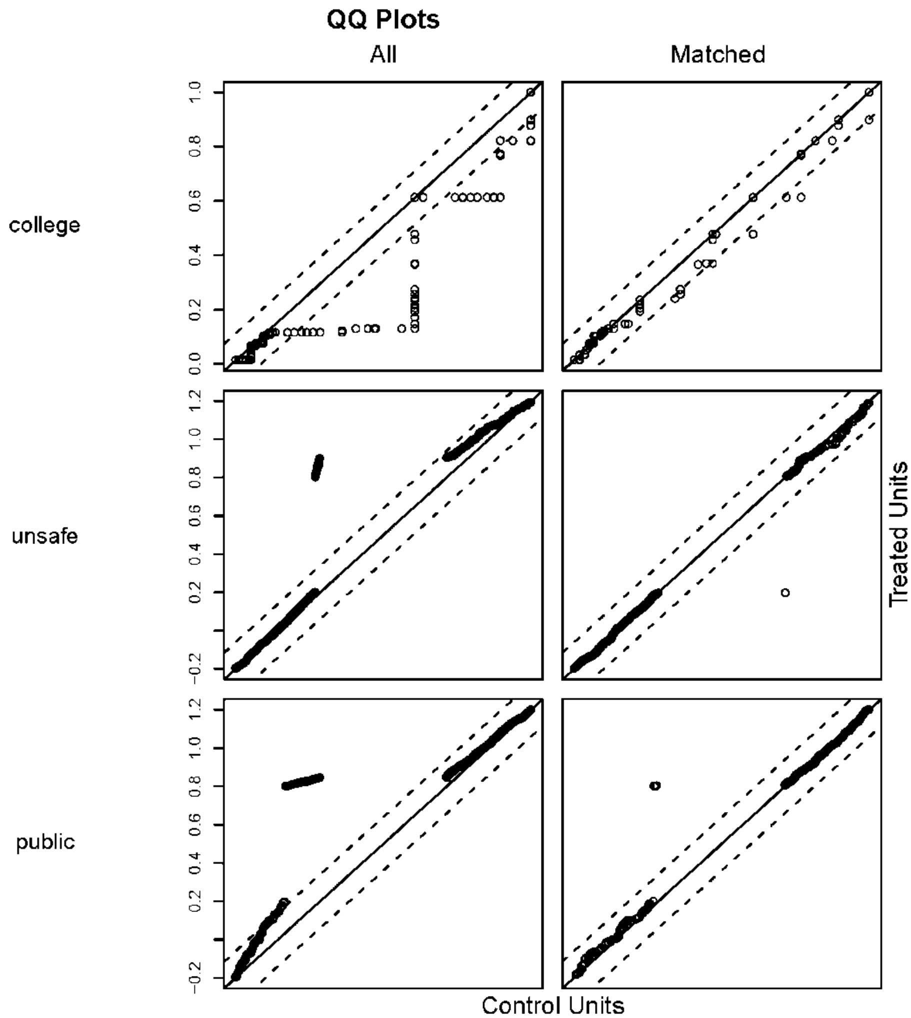
[b] Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

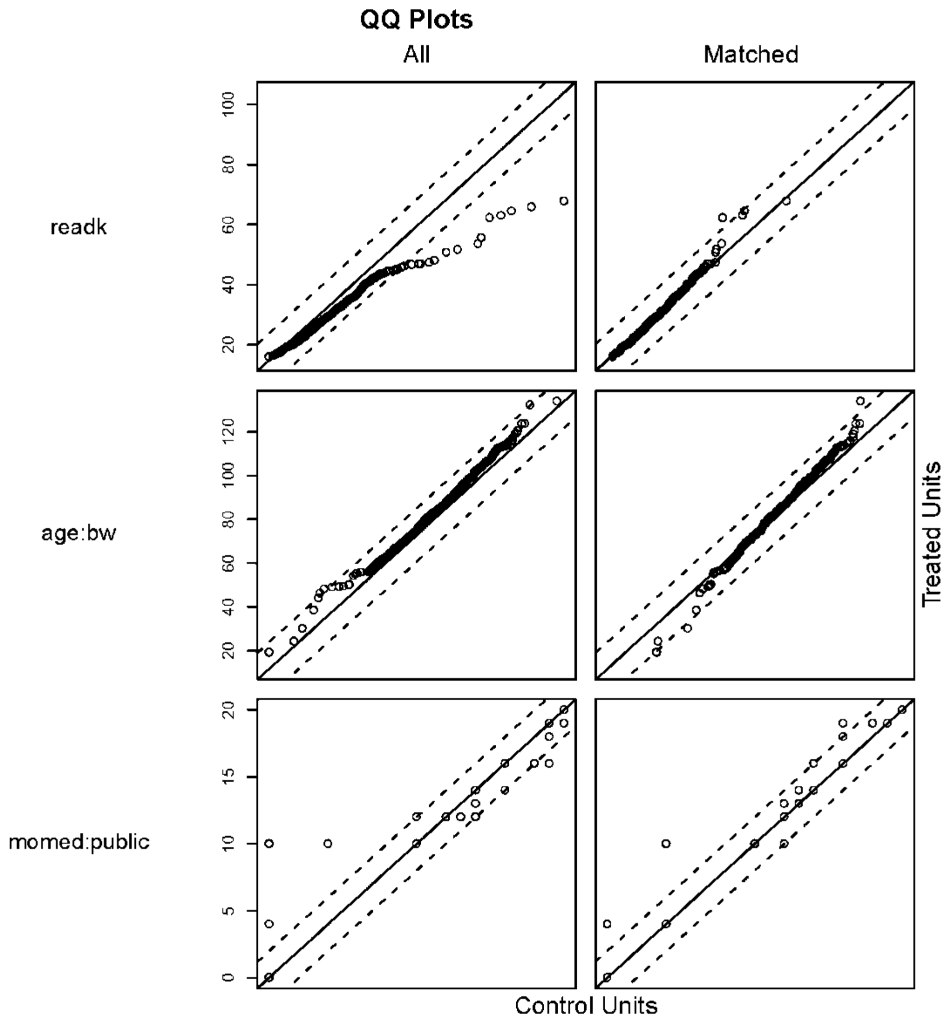
Appendix Figure 1. The Quantile-Quantile plot of covariates in the genetic matching.

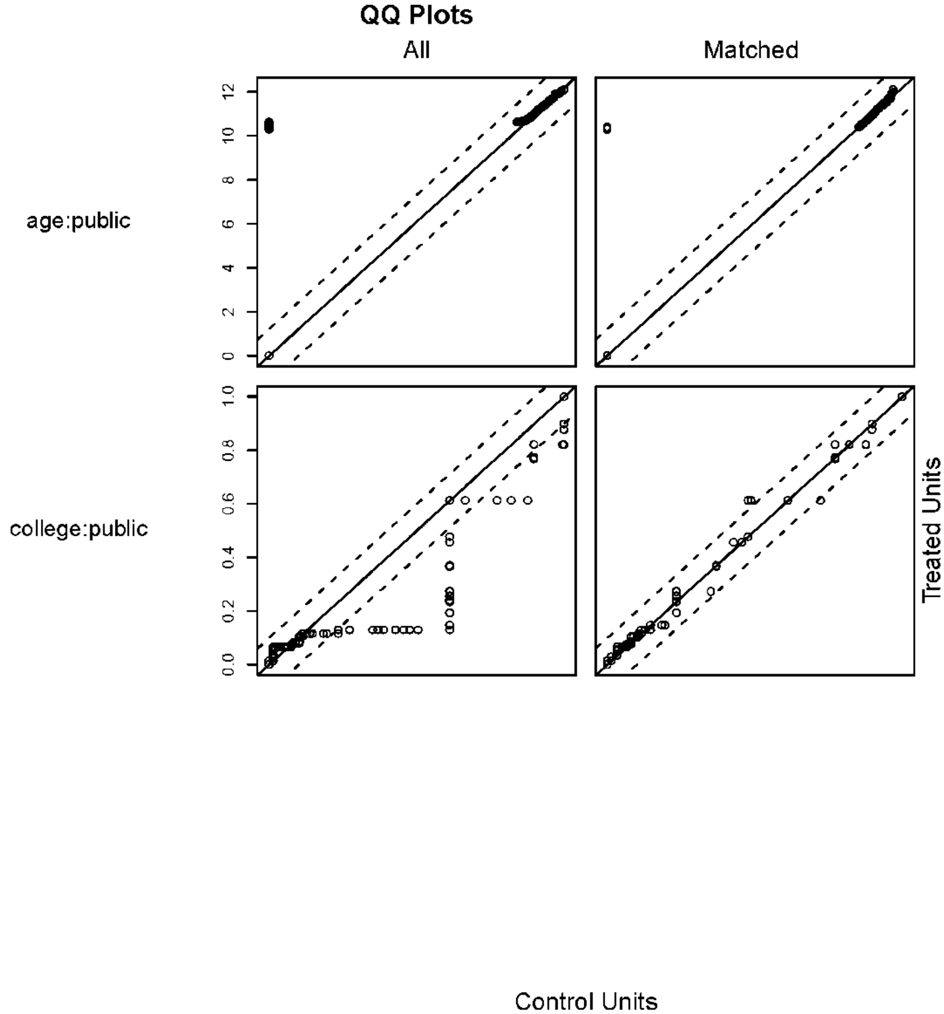



QQ Plots
All
Matched
frlunch
soda
rapid
Treated Units
Control Units
−0.2
0.2
0.4
0.6
0.8
1.0
1.2
0
5
10
15
20
25
0
1
2
3
4
5
6
7


QQ Plots
All
Matched
tvweek
lowinc
momed
Treated Units
Control Units


QQ Plots
All
Matched
college
unsafe
public
Treated Units
Control Units


QQ Plots
All
Matched
readk
age:bw
momed:public
Treated Units
Control Units
20
40
60
80
100
120
0
5
10
15



QQ Plots
All
Matched
age:public
college:public
0
2
4
6
8
10
12
0.0
0.2
0.4
0.6
0.8
1.0
Treated Units
Control Units