Impacts of classifying New York City students as overweight

Douglas Almond[a,b,1], **Ajin Lee**[c], **and Amy Ellen Schwartz**[d,e]

[a]Department of Economics & School of International and Public Affairs, Columbia University, New York, NY 10027; [b]National Bureau of Economic Research, Cambridge, MA 02138; [c]Department of Economics, Columbia University, New York, NY 10027; [d]Maxwell School, Center for Policy Research, Syracuse University, Syracuse, NY 13244; and [e]Institute for Education and Social Policy, New York University Steinhardt School of Culture, Education, and Human Development, New York, NY 10012

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved January 29, 2016 (received for review September 19, 2015)

US schools increasingly report body mass index (BMI) to students and their parents in annual fitness "report cards." We obtained 3,592,026 BMI reports for New York City public school students for 2007–2012. We focus on female students whose BMI puts them close to their age-specific cutoff for categorization as overweight. Overweight students are notified that their BMI "falls outside a healthy weight" and they should review their BMI with a health care provider. Using a regression discontinuity design, we compare those classified as overweight but near to the overweight cutoff to those whose BMI narrowly earned them a "healthy" BMI grouping. We find that overweight categorization generates small impacts on girls' subsequent BMI and weight. Whereas presumably an intent of BMI report cards was to slow BMI growth among heavier students, BMIs and weights did not decline relative to healthy peers when assessed the following academic year. Our results speak to the discrete categorization as overweight for girls with BMIs near the overweight cutoff, not to the overall effect of BMI reporting in New York City.

childhood obesity | New York City | regression discontinuity design | Fitnessgram | BMI

Obesity often emerges early in childhood. Among 7,738 US children, eighth graders were four times as likely to be obese if they were overweight in kindergarten (1). Parents can be surprisingly uninformed about overweight and obesity status of their children. Sixty-one percent of parents in San Diego correctly identified whether their child was overweight (2). "Obliviobesity" among US parents may be growing over time (3, 4). On the other hand, US school districts and states have begun distributing annual fitness and body mass index (BMI) "report cards" to students and parents. Such personalized informational interventions have appeal in economics because they can be relatively inexpensive, particularly compared with traditional programs that include the delivery of costly health services. As individual dietary and exercise habits are being established during childhood, it has been argued that obesity surveillance, reporting, and prevention interventions should likewise begin early.

Opponents of BMI reporting argue that informing children that they are "fat" can be stigmatizing, hurt their self-esteem, and even encourage bullying. Such unintended reactions may prompt a cascade of behavioral responses that do not improve health (5). Additionally, BMI (weight divided by height squared) is routinely criticized as a metric of fitness. Whether BMI report cards are an effective tool for helping to reduce obesity is not obvious a priori. Large-scale empirical analyses are now feasible thanks to expanded collection of BMI data, data generated for the administrative purpose of issuing BMI report cards.

Fitnessgrams were adopted by New York City's public schools in 2007–2008, reporting each student's BMI alongside categorical BMI designations. Specifically, each student's BMI is classified and reported to be "underweight," "healthy," "overweight," or "obese." Categorizations are assigned using the students' most recently recorded BMI *vis à vis* the age- and sex-specific BMI cutoffs from the Centers for Disease Control (CDC). By these national criteria (based on 1970s National Health and Nutrition Examination Survey data), over one-third of New York City public school students are either overweight or obese. Report cards for NYC students with an overweight BMI state, "Your BMI falls outside of a healthy weight; please review your BMI with a health care provider."

Hard copy Fitnessgram reports are distributed to students in school in May. Here, we consider the discontinuous assignment of the overweight label and its associated health recommendations. As we will argue below, categorization helps isolate the causal effect of BMI information on subsequent BMI and weight.

Methods

Research Design. We compare those narrowly designated as overweight to those narrowly designated as having a healthy BMI. For students who were particularly close to the overweight threshold, overweight categorization has an arbitrary component because individual control over small movements in the recorded BMI is imperfect. Reasons for this imperfect "local" control include: day-to-day variation in weight; diurnal variation in weight; measurement error in recorded height; differences in weight when measured across different scales, and students not knowing in advance when the in-school Fitnessgram BMI assessment will occur. Additionally, whereas the adult BMI cutoffs for overweight (25) and obese (30) are widely known, cutoffs among children are not because they change with each age in months and seldom fall on round numbers. For example, an 11-y-old girl would be classified as overweight if her BMI exceeded 20.19667, whereas 1 mo later her threshold would be 20.26514 (i.e., the overweight threshold for a girl 11 y and 1 mo old). Clearly students have some degree of control over their recorded BMI. Our identifying assumption is that control over the exact BMI, as recorded for Fitnessgram assessment, is short of perfect. That said, we allow for a continuous relationship between baseline and subsequent BMI, an underlying process which we do not ascribe to the categorization.

Imprecise control over BMI would imply that observed and unobserved predetermined characteristics have continuous distributions in baseline BMI.

Significance

One third of US children are overweight. Childhood obesity is strongly associated with academic performance, predicts obesity in adulthood, and may be easier to modify than adult weight. Parents who are unaware their children are overweight may impede healthier weight ranges. We consider an informational intervention of New York City's obesity report cards using a regression discontinuity design framework. To a surprising degree, discontinuous assignment rules are commonplace in allocating program services and shaping clinical care decisions by health care providers. Regression discontinuity designs offer a relatively untapped methodological approach for isolating causal effects, particularly as they relate to health.

Author contributions: D.A., A.L., and A.E.S. designed research; performed research; contributed new reagents/analytic tools; analyzed data; and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: almond@columbia.edu.

That is, if individuals are unable to precisely sort on either side of the assigned threshold, treatment is essentially randomly assigned around the cutoff (6). As a consequence, the discontinuity in the dependent variable at the threshold captures the causal effect of the overweight label. To empirically test imprecise control over BMI, we examine whether the density of baseline BMI is continuous at the cutoff. If individuals can precisely control their BMI so as to receive the healthy label, we might expect to see "too many" students just below the cutoff. We examine the density and confirm that observations are not "heaped" near the threshold. Moreover, we can also evaluate whether BMI control appears imperfect by assessing whether individuals are similar in observable characteristics as the distance from the threshold goes to zero. Although it is impossible to test whether their unobservable factors jump at the threshold (e.g., the unobserved, individual BMI progression before entry into our panel data), incomplete control of recorded BMI described above would likewise lead such unobservables to be continuous across the threshold. This is not to suggest we think key unobservables are uncorrelated with BMI. Rather, the key issue is whether such factors are discontinuous at the threshold.

Data Access and Study Population. New York City's Department of Education (NYCDOE) has a standardized application procedure through which researchers request access to deidentified microdata. For more information, see schools.nyc.gov/Accountability/data/DataRequests.htm. Study and data handling procedures were approved by both the New York University and NYCDOE Institutional Review Boards. No individual consent was required for this analysis of deidentified administrative data.

For the 2007–2008 to 2011–2012 academic years, we obtained 3,692,026 BMI measurements on New York City public school students. Records for individual students are linked longitudinally using a unique, encrypted student identifier. Here, we focus on how female students' subsequent BMI and weight respond to an overweight designation in the previous academic year. We restrict estimation to the subsample of girls whose baseline BMI fell within 0.5 SDs (roughly 2 BMI units) of the overweight threshold; 442,408 records met these selection criteria. As a consequence of imprecise BMI control, we expect (and indeed find) the baseline characteristics of the sample just above and below their respective overweight threshold to be similar after allowing for a BMI trend. We test whether the results are sensitive to choices of bandwidth other than 0.5 SDs in *Alternative Bandwidths* below.

Econometric Model. We estimate the effect of the overweight label on future BMI and weight in a regression discontinuity design (RDD) framework:

$$\begin{aligned}\mathrm{BMI}_{ia,t+1} = \beta_0 + \beta_1 \cdot \mathrm{overweight}_{iat} + f(\mathrm{BMI}_{iat} - c_a) + \\ \phi_a + \phi_a \cdot f(\mathrm{BMI}_{iat} - c_a) + \epsilon_{iat},\end{aligned}$$

where β_1 is the coefficient of interest. Overweight$_{iat}$ equals 1 when the BMI is greater than the overweight threshold for student i with age in months a in year t, and 0 otherwise. We control for a linear function $f(\cdot)$ of the baseline BMI in year t: BMI$_{iat}$, centered at zero for the relevant age-specific cutoff c_a. We allow for the slopes to vary across the threshold by including the interaction term between $(\mathrm{BMI}_{iat} - c_a)$ and overweight$_{iat}$. Specifically, $f(\mathrm{BMI}_{iat} - c_a) = \delta_1 \cdot (\mathrm{BMI}_{iat} - c_a) + \delta_2 \cdot (\mathrm{BMI}_{iat} - c_a) \cdot \mathrm{overweight}_{iat}$.

Thus, we do control for the (positive) correlation between baseline BMI and subsequent BMI. This accounts for a potential difference in the growth rates between girls below and above the threshold. Because our sample consists of girls of all ages in grades K–11 in the baseline year, we additionally control for age in months fixed effects ϕ_a. Overweight assignment is defined using the 85th percentile for 1970s US children of the same age and sex. Thus, we compare girls just below the overweight threshold within the same age in months group to those just above. We also flexibly allow for the slopes in BMI to vary for each age by including the interaction terms between ϕ_a and $f(\mathrm{BMI}_{iat} - c_a)$. We additionally examine weight$_{ia,t+1}$ as a dependent variable. In the tables, we report the estimated β_1 s and their confidence intervals.

Results

We find small effects of being labeled overweight in New York City schools on female students' subsequent BMI and weight. These impact estimates are precise, ruling out nearly all "beneficial" effects of being classified as overweight on subsequent BMI and weight. The results are similar when we examine the effects of being classified as obese. We perform an identical analysis for males, and do not find any effects.

These findings are summarized in Table 1 below. Each row reports estimates from a separate regression of the dependent variable on the overweight categorization controlling for baseline

Table 1. Effects of the overweight label

Dependent variable	Mean value below cutoff*	Overweight label†, 95% CI	*P* value
Subsequent health outcome			
BMI next year	19.84	0.028 (0.002,0.054)	0.04
Weight next year	95.81	0.169 (0.002,0.336)	0.05
Baseline characteristic			
Grade group			
K-3	0.36	0.000 (-0.001,0.001)	0.94
4–5	0.16	−0.001 (-0.002,0.001)	0.59
6–8	0.25	−0.000 (-0.002,0.002)	0.83
9–12	0.24	0.001 (-0.001,0.002)	0.36
Ethnic group			
Asian	0.16	0.000 (-0.003,0.004)	0.79
Black	0.29	−0.001 (-0.005,0.004)	0.77
Hispanic	0.40	−0.002 (-0.008,0.003)	0.37
White	0.15	0.003 (-0.001,0.006)	0.15
PACER score‡	23.44	−0.054 (-0.230,0.123)	0.55
Free or reduced-price lunch§	0.85	−0.001 (-0.005,0.002)	0.43
Math scale score¶	683.40	−0.004 (-0.436,0.427)	0.98
Special education	0.08	−0.002 (-0.005,0.001)	0.17

*The sample below the overweight cutoff is the "untreated" control group in our RDD.

†We regress each dependent variable on the binary indicator of the overweight label controlling for baseline BMI, age in months fixed effects, and interaction terms between age in months fixed effects and baseline BMI. We allow for the slopes in BMI to vary below and above the threshold. The reported estimates are coefficients on the binary indicator, measuring the size of the discontinuity at the threshold.

‡PACER stands for Progressive Aerobic Cardiovascular Endurance Run. Students run laps between two points in a certain amount of time, and the score is the number of laps completed.

§Students are eligible for free lunch if their guardians earn less than 130% federal poverty level and reduced-price lunch if their guardians earn less than 185% federal poverty level.

¶Students in grades 3–8 take the State Mathematics test each spring. The number of correct answers is converted into a scale score, which makes it possible to compare performance on the test across different grades.

BMI and age in months. Being labeled as overweight has little and, if anything, a small positive effect on next year's BMI [an increase of 0.03 BMI units; 95% confidence interval (CI), 0.002–0.054; P value = 0.04 in row 1] and next year's weight (an increase of 0.17 lb.; 95% CI, 0.002–0.336; P value = 0.05 in row 2). This suggests that the overweight label has no beneficial effects compared with assignment with the healthy label (among students with very similar baseline BMIs).

The subsequent rows of Table 1 confirm that baseline characteristics are not statistically different across the overweight threshold. This is consistent with our identifying assumption that individuals are incapable of precisely manipulating their BMI around the cutoff. Thus, focusing empirical comparisons on those just below and just above the threshold furnishes a "natural experiment" that helps balance the other determinants and correlates of BMI.

In additional analyses, we consider the subgroup of girls who had not previously been notified they were overweight by Fitnessgram. For them, overweight categorization was potentially more salient. Being newly categorized as overweight indeed has a larger impact on subsequent BMI (roughly double the average effect reported in Table 1). Still, the effect magnitude remains modest: a 0.07 BMI unit increase on a mean of roughly 20 (95% CI, 0.001–0.136; $P = 0.05$). Second, we consider heterogeneity by grade and find that the oldest children in our sample—girls sent report cards at the end of their junior year—show the largest response. For seniors, we estimate BMI increased 0.27 units (on a mean of 23.6; 95% CI, 0.08–0.46; $P = 0.005$) and weight increased 1.35 lb. (on a mean of 136.3 lb.; 95% CI, 0.11–2.60; $P = 0.03$) due to overweight labeling at the end of the junior year. Finally, we examine whether the overweight label affects subsequent academic performance (measured by the State Mathematics and English Language Arts test scores) of female students, but find no effects on these performance measures.

Alternative Bandwidths. We consider whether our results are sensitive to the choice of bandwidth, i.e., the size of the window around the overweight cutoff used for analysis. If the window is too narrow the estimates will be imprecise and control for the running variable may be too flexible and "overfit." On the other hand, observations far from the threshold should not drive discontinuity impact estimates. We reestimate our regression equation using subsamples from 0.1 SD up to 1 SD from the cutoff (in 0.1-SD increments). Fig. 1*A* plots the RDD estimates for subsequent BMI along with the 95% confidence intervals for each bandwidth. Fig. 1*B* plots the estimates for subsequent year's weight. The bandwidth used in Table 1 regressions is 0.5 SD. As expected, estimates of β_1 are imprecise for the smallest bandwidths. Estimates of β_1 are surprisingly stable around 0.03 for bandwidths from 0.3 SD to 0.8 SD. Estimates start to rise for wider windows, potentially due to the bias arising from observations far away from the threshold. However, the point estimates do not fall below 0.03 for all bandwidths, suggesting throughout little or no beneficial effect of the overweight label. Estimates for next year's weight are generally imprecise, but the point estimates are positive and stable for 0.2 SD and above. Again, the estimates seem to increase when we use larger windows.

Placebo Percentiles. We do not expect effects at placebo BMI percentiles that do not affect the categorizations and health recommendations reported by Fitnessgram. In addition to the 85th percentile that determines overweight status, CDC reports the following "closest" percentiles: the 50th, 75th, 90th, and 97th [www.cdc.gov/growthcharts/html_charts/bmiagerev.htm]. As with the actual overweight cutoff, these thresholds vary depending on age in months and sex of the child. Just 31% of NYC students are below the 50th percentile and 11% are above the 97th percentile, as CDC set these percentiles using data on US children

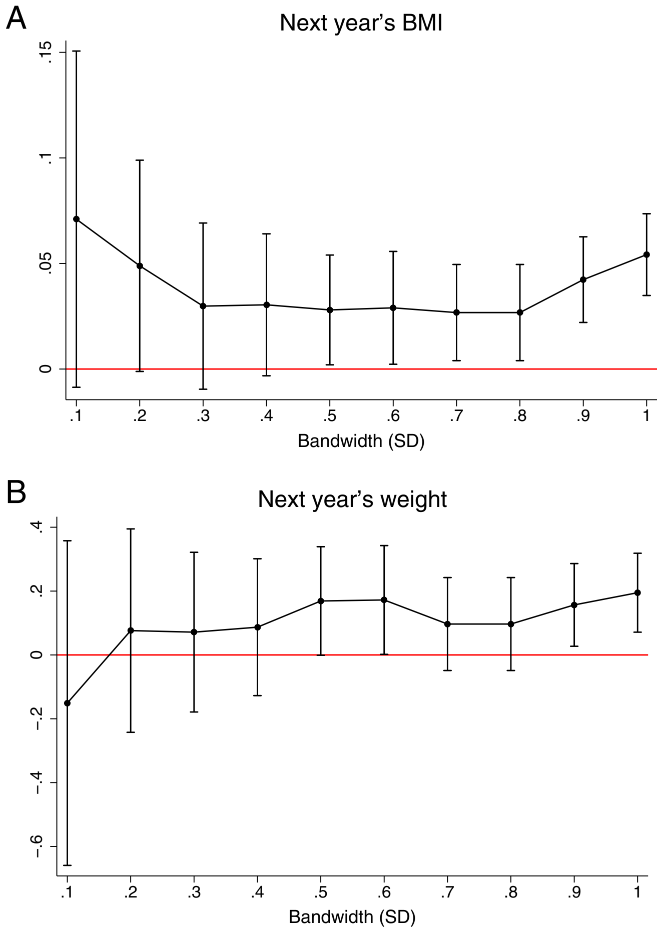

Fig. 1. Regression discontinuity estimates for alternative bandwidths of next year's BMI (*A*) and weight (*B*).

during the 1970s. We estimate our regression equation treating each percentile as the "overweight cutoff." Around each placebo cutoff, we include observations within 0.5 SDs, i.e., the same bandwidth as Table 1.

As expected, the placebo percentiles (50th, 75th, 90th, and 97th) do not yield statistically significant "effects" on subsequent BMI. These nonresults suggest $f(\cdot)$ adequately captures underlying (continuous) BMI trends that can vary with baseline BMI. When we instead consider weight as the outcome, again the 75th, 90th, and 97th percentiles generate estimates for β_1 that are smaller in absolute value and not statistically distinguishable from zero. We do estimate a statistically significant false positive for the 50th percentile on subsequent weight in pounds. With a 95% CI, we expect 5% false positives/false negatives: we obtained 1 in 8.

Discussion

To a surprising degree, discontinuous assignment rules are common in allocating both program services, e.g. ref. 7, and clinical care. Diabetes diagnoses are frequently based in part on a threshold fasting glucose level and hypertension diagnoses based on a threshold blood pressure level (8). Newborn infants weighing below 1,500 g and thereby "very low birth weight" receive discontinuously more neonatal care and have lower infant mortality than slightly heavier infants (9). Outside of economics, analysis of the RDD offers a relatively untapped econometric methodology for isolating causal effects, particularly as they relate to health (10). Practice guidelines and even "rules of thumb" can constitute "silent adjudicators of clinical practice" (8). At present, the discreteness inherent in both explicit guidelines and tacit rules of thumb are seldom leveraged by researchers, despite the growing availability of administrative data on health. Whereas in our application the overweight label is fully determined by BMI for a given age and gender, this need not be the case. As is more common, treatment assignment in RDD analyses can be multifactorial, as in the above examples of diabetes and hypertension diagnosis. Thus, valid RDDs can be either "sharp" like ours or "fuzzy" (6).

Our BMI results suggest little or no benefit of overweight notification among female students in New York City. These results apply to the subpopulation of female public school students in New York City with a high probability of being near the overweight cutoff. The results are similar when we examine the effects of obesity classification. We perform an identical analysis for males, and do not find any significant effects. Unfortunately, our data do not reveal to what extent the BMI report cards are actually read and "processed" by the student and their parents. In Mexico, an experimental distribution of BMI report cards did improve parents' knowledge of their child's BMI, but likewise no BMI reductions were detected (11). In New York, failure to read and process BMI report cards could attenuate the effect of categorization on subsequent BMI. However, to the extent that we find significant effects, our findings suggest that indeed these report cards were read by an important subpopulation.

We believe our RDD approach isolates the effect of having BMI above versus below the categorization threshold. This effect could be generated either by: (*i*) those below the threshold with healthy weights responding to categorization by reducing their weights, or (*ii*) those being labeled as overweight and advised to consult with a health care provider increasing their weight (or some combination of both responses). A priori, we find it more likely that those classified as overweight responded, reinforced by observation that the largest response is found among teen girls, who might respond less constructively to the overweight categorization absent additional supportive services. Future work might consider whether electronic distribution of BMI report cards yields comparably small impacts and whether the effect of BMI labels likewise remains small.

ACKNOWLEDGMENTS. We are grateful to Siddhartha Aneja and Meryle Weinstein for assistance with the data. We thank the Office of School Wellness Programs and the Research and Policy Support Group at the New York City Department of Education for providing requisite microdata, and the National Institutes of Health (5 R01 HD070739) and National Science Foundation (SES-0847329) for financial support.

1. Cunningham SA, Kramer MR, Narayan KM (2014) Incidence of childhood obesity in the United States. *N Engl J Med* 370(5):403–411.
2. Huang JS, et al. (2007) Parental ability to discriminate the weight status of children: Results of a survey. *Pediatrics* 120(1):e112–e119.
3. Katz DL (2015) Oblivobesity: Looking over the overweight that parents keep overlooking. *Child Obes* 11(3):225–226.
4. Hoffman J (June 16, 2015) Blind to a child's obesity. *NY Times*, p D1.
5. Hunger JM, Tomiyama AJ (2014) Weight labeling and obesity: A longitudinal study of girls aged 10 to 19 years. *JAMA Pediatr* 168(6):579–580.
6. Lee DS, Lemieux T (2010) Regression discontinuity designs in economics. *J Econ Lit* 48(2):281–355.
7. Hainmueller J, Hangartner D, Pietrantuono G (2015) Naturalization fosters the long-term political integration of immigrants. *Proc Natl Acad Sci USA* 112(41):12651–12656.
8. McDonald CJ (1996) Medical heuristics: The silent adjudicators of clinical practice. *Ann Intern Med* 124(1 Pt 1):56–62.
9. Almond D, Doyle JJ, Jr, Kowalski AE, Williams H (2010) Estimating marginal returns to medical care: Evidence from at-risk newborns. *Q J Econ* 125(2):591–634.
10. Moscoe E, Bor J, Bärnighausen T (2015) Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: A review of current and best practice. *J Clin Epidemiol* 68(2):122–133.
11. Prina S, Royer H (2014) The importance of parental knowledge: Evidence from weight report cards in Mexico. *J Health Econ* 37:232–247.