DOES INCREASING INCOME LEAD TO GREATER CHILDHOOD OBESITY?

EVIDENCE FROM A NATURAL EXPERIMENT

Molly A. Martin
Pennsylvania State University

Elizabeth O. Ananat
Duke University

Kelly D. Davis
Oregon State University

Diane K. McLaughlin
Pennsylvania State University

September 25, 2015

Submission for the 2016 Annual Meetings of the Population Association of America
Does increasing income lead to greater childhood obesity?
Evidence from a natural experiment

Abstract

Many scholars anticipate family income to protect youth from the risks of becoming overweight or obese, but observational research findings do not support this hypothesis. We speculate that unobserved parental traits, however, could suppress the true association between income and children's risks of being overweight or obese. We take advantage of an emerging natural experiment – the extraction of natural gas in the Marcellus Shale geological formation – to study whether increasing family income affects children's risks of being overweight or obese. Using predicted Marcellus Shale income as an instrument, we will conduct difference-in-difference analyses with multiple comparison groups to isolate the causal effect of family income increases on rates of overweight and obesity for children in (a) elementary school and (b) middle- or high school. This natural experiment offers an unprecedented opportunity to answer interesting and important population health questions about the effect of family income on children's weight.
Extended Abstract

When scholars identify the risk factors for childhood obesity, many expect that income plays a key role, noting that poor children are more likely to be overweight or obese. Yet while this descriptive pattern is accurate, family income and poverty are rarely significant predictors of children’s weight in multivariate models (Goodman et al. 2003; Gordon-Larsen et al. 2003; Martin 2008; Martin et al. 2012; Wang and Zhang 2006; Zhang and Wang 2007). Yet we can imagine a series of unobserved parental traits that could suppress a significant association between family income and childhood obesity in these observational data. Thus, we leverage a natural experiment to assess the causal relationship between increasing family income and childhood obesity.

We take advantage of an emerging natural experiment – the extraction of natural gas in the Marcellus Shale geological formation – to study how increases in family income affect children’s risk for obesity and overweight. Before recent technological advances, deep gas was not extractable and had no economic value; now, families who own mineral rights for land located above the Shale can receive payments upon signing a lease and royalties when gas is sold. There are four exogenous sources of variation in expected Marcellus Shale income: the quality of the shale, differences in state policy, the distance and ease of travel on existing highways to hubs of Shale development, and the historical property rights in the area. Among these, the key factors are the quality of the shale and state policy. Within the Marcellus Shale some areas are more economically viable because of the quality of the shale and, thus, profitability of each well. In addition, Pennsylvania (PA) allows drilling, while New York (NY) does not. As a result, some families living above the Marcellus Shale will get rich; others will receive nothing. We exploit these differences
to eliminate the confounding of unobserved parental traits to identify the true effect of increasing income on childhood obesity.

Background

Scholars theorize that greater family income can protect children against the risks of childhood obesity. Family income signals the family’s relative power to purchase goods and services related to health depending on prices and increasingly healthy options are frequently more costly. For example, as a result of federal agricultural policy, “healthy” food is generally more expensive relative to “bad” food, meaning highly-processed foods, high in sugar, and deplete of vital nutrients (Drewnowski and Specter 2004). In addition, the costs of participating in physical activity are increasing for families as schools divest from organized sports (McNeal 1998) and parents must privately fund their children’s participation in community- or school-based organized physical activities, like sports and dance. As such, a common expectation is that greater family income reduces the risk of childhood obesity because it increases their ability to purchase goods and services that help children maintain a healthy weight (Cawley 2004).

Increases in family income could also indirectly reduce childhood obesity by reducing chronic stress within the family. Economic distress can strain both parents’ and children’s ability to manage the functions of daily life. Economically distressed families exhibit more chaos in the home (Evans et al. 2005), and reduced parental involvement with their children (Brody, Stoneman, and Flor 1995; Gershoff et al. 2007), and prior research finds consistent family routines and parental involvement encourage healthy-weight related behaviors. In addition, prior research finds chronic stress is associated with
associated with increased abdominal fat, weight gain and a greater likelihood of becoming overweight or obese (Anagnostis et al. 2009; Björntorp 2001; Fraser et al. 1999).

It is important to note, however, that greater income can also facilitate the purchase of high-cost goods that could lead to weight gain (e.g., cable television, video games) by reducing children's physical activity (Cawley 2004). Thus, it is possible that greater family income could increase the risks of childhood obesity. It depends on the relative balance of purchases that promote a healthy weight and those that create foster inactivity.

The empirical evidence is generally mixed and generally does not comport with the dominant theoretical hypothesis of a negative relationship between family income and childhood weight. In fact, only one nationally-representative study finds a significant negative association between family income and adolescent’s weight (Goodman 1999). Other studies find a significant, negative association but only for narrowly-defined population subgroups (e.g, African-American females, Hispanic boys) (Balistreri and Van Hook 2009; Goodman, Slap, and Huang 2003; Gordon-Larsen, Adair, and Popkin 2003; Kimm et al. 1996; Miech et al. 2006; Zhang and Wang 2007), although the theoretical expectations anticipate a more universal pattern, with effects for all children. Interestingly, one study even finds a positive association between family income and adolescent weight (Haas et al. 2003). By and large, however, most studies find no statistically significant association between family income and children’s weight (Goodman et al. 2003; Gordon-Larsen et al. 2003; Martin 2008; Martin et al. 2012; Wang and Zhang 2006; Zhang and Wang 2007).

A key challenge, however, in this field is that family income is correlated with a host of parental traits that are difficult to measure and statistically control for in multivariate
models, such as parents’ cognitive abilities, personality traits, habits, and values. Such factors can introduce significant bias to estimated associations in observational data. We know of no prior studies that use an experimental or quasi-experimental approach to eliminate biases due to unobserved heterogeneity.

Theoretically, we anticipate several unobserved parental traits could lead to underestimates of the “true” effect of family income. In particular, we hypothesize that parents’ relative organizational skills, sense of diligence, and ability to stick to routines could suppress the estimated relationship between family income and children’s weight. These “soft skills” would likely increase family income via increases in their parents’ likelihood of employment and wage rates, while reducing the risks for childhood obesity via the creation of healthy family routines for children’s meals and physical activity, and reductions of chaos in the home, which could lead children to snack more, engage in more screen time, and have irregular bedtimes and sleep habits. These traits and their hypothesized implications are difficult to capture, but if they are theoretically relevant, then they could create long-term risks for childhood obesity as these daily risks accumulate. Therefore, we hypothesize that unobserved heterogeneity could downwardly bias the estimated effect of family income for childhood obesity.

It is important to also consider the likely timing of income effects. Given the various biological mechanisms intent on regulating the balance of energy input and output and, thus weight, it often takes years to accumulate weight to the point of becoming overweight or obese. Thus, we do not expect a change in income in year T to lead to significant shifts in overweight or obesity in year T+1. Instead, we anticipate that it takes numerous years for income effects to accumulate and model the effects of approximately four years of income
changes (between 2006 and 2010) to predict later overweight and obesity (measured in 2010).

Research Design

We seek to eliminate the effect of these unobserved parental traits by using a quasi-experimental design well-validated in many other natural experiments – a difference-in-differences approach – to exploit features of the Marcellus Shale development. We take advantage of an emerging natural experiment occurring across Pennsylvania (PA) and New York (NY). Due to new drilling technology that combines hydraulic fracturing with horizontal drilling, natural gas trapped within the Marcellus Shale formation (see Figure 1) thousands of feet below the surface is now profitable to extract. Gas companies are now spending billions of dollars to drill, extract, process, and transport the natural gas from the Marcellus Shale to market. Families who own their property’s mineral rights can sign a lease to allow gas companies to drill and extract the gas beneath their property, and so potentially receive large income gains. The lease establishes the royalty rate (i.e., the percentage of the profit the mineral rights owner receives upon the sale of the gas from their parcel). Reported royalty rates vary (primarily due to differences in when leases were signed), but the PA state-mandated minimum is 12.5% and average is 13.4% (Ward et al. 2011). The royalties families receive reflect the contracted royalty rate, the price of natural gas upon its sale, and the per acre volume of gas extracted (which reflects the Shale’s quality beneath a parcel). Lease (i.e., bonus) payments are contracted to arrive either in one or multiple payments, while royalties are scheduled to arrive repeatedly. A relatively high percentage of landowners in Southwestern PA do not possess the mineral rights for
their property due to the historical development of coal in the area. For many properties there, the mineral rights were detached from the property decades ago when coal companies bought the mineral rights from landowners. Thus, holding the amount of local shale development constant, families in some areas of PA affected by prior coal or shallow-well gas extraction are less likely to experience income gains than families in other areas of PA.

The first productive Marcellus Shale well was drilled in southwestern PA in 2005, yet the pace of drilling did not accelerate until 2008 when estimates were released that the Marcellus contains 50 trillion cubic feet of recoverable gas (i.e., enough to supply the nation for two years) (Engelder and Lash 2008); estimates were later upgraded to nearly 500 trillion cubic feet) (Engelder 2009). Before 2005, families who purchased land did not consider Marcellus Shale gas in their purchase decision.

There are four exogenous sources of variation in the amount of income families might receive from Marcellus Shale gas development. First, there is variation because of the geological quality of the shale. The Core of the Marcellus Shale formation (“A” and “C” in Figure 1) is more economically viable for production given the known depth, thickness, porosity, thermal maturity, and silica content of the shale (Dell, Lockshin, and Gruber 2008). Second, there is exogenous variation in the amount of income families living in the Core can receive because of state policy. NY currently has a statewide moratorium on Marcellus Shale drilling; PA does not. Natural gas companies reported paying $2.07 billion in Marcellus Shale lease and royalty payments in PA in 2010 (Considine, Watson, and Blumsack 2011). Therefore, we define our “treatment” and “comparison” school districts based on the intersection of these two factors. Third, the original location of communities
relative to the cities that eventually became the hubs of Marcellus Shale economic development – Williamsport, PA in northeastern Pennsylvania and Canonsburg, PA in southwestern PA. Their proximity and ease of access via existing state and federal highways made some areas easier for companies to develop. Fourth, there is regional variation in the share of properties that had mineral rights detached during prior coal and shallow-well gas extraction, which causes families to be ineligible for royalty income when gas is extracted from that parcel.

We will utilize a difference-in-differences (DID) technique to measure the predicted effect of average royalty income (i.e., our “treatment”) on average measures of child well-being over specified periods of time. Using the labels in Figure 1, our initial difference-in-differences model will examine \((A-B) - (C-D)\). By including NY districts above the Core of the Shale and non-Core NY districts, any underlying differences between Core and non-Core districts in PA and NY are implicitly controlled for, even when these variables have not been observed. Assuming that NY and PA districts above the Shale’s Core would have had parallel trends over time in child well-being in the absence of Shale development in PA, the observed differences in child outcomes in NY between those above the Core and non-Core can be interpreted as the change in child well-being that Core PA districts would have experienced (relative to non-Core PA districts) had Core PA districts not experienced the increase in average royalty income. Later models will also consider the initial and annual proximity of a school district to the nearest natural gas pipeline to identify additional exogenous variation in royalty income using a difference-in-difference-in-differences analysis. Finally, some models will separately examine the Core PA areas where few
families have mineral rights in order to identify effects of Shale development that do not come through increased income.

We recognize that Marcellus Shale development may affect local children through channels other than income (e.g., via increased truck traffic, disruptions to family routines, or water contamination). Therefore, we aim to isolate the effect of income by separately measuring outcomes in areas where, historically, few PA families hold mineral rights (and therefore experience other effects from Marcellus development but not income effects) and the effects in areas where a larger share of PA families hold mineral rights that can be leased for Marcellus Shale gas extraction (and therefore experience these other effects plus income effects).

We think this interesting natural experiment can offer critical information on a long-standing question regarding the causal impacts of increasing income on children’s risks for overweight and obesity. Although we do not have an a priori reason to anticipate risks varying across the stages within childhood, the data permit us to explore whether these risks differ for children in elementary school relative to children in middle or high school.

Data and Methods

To leverage this natural experiment, we integrate comprehensive administrative datasets covering the PA and NY populations living above the Marcellus Shale with geological and natural gas industry information and Census data from several secondary sources about children and their families living in PA and NY. We use data aggregated to the school-district level instead of counties to have more refined geographic measures and a sufficient number of cases across Marcellus conditions in PA and NY. Furthermore,
children’s obesity data are collected in the schools and provided by PA and NY departments of education at the school-district level.

**Measures**

*Children’s Overweight and Obesity.* The PA and NY Departments of Health collect data on the percentage of public school students who are overweight or obese based on their body mass index (BMI) and the CDC standards for evaluating children’s weight-for-age by sex (Ogden et al. 2002). In PA, trained professionals collect children’s BMI in school. PA data are available annually for grades K-12 for years 2007/08 through 2010/11. The data are grouped as grades K-6 (i.e., elementary grades) and grades 7-12 (middle and high school grades). Our analyses will focus on the 2010/11 data.

In NY, data on children’s weight-for-age are collected by schools from students’ submitted health certificate forms, which are mandatory for children in grades K, 2, 4, 7 and 10. The health certificate forms are usually completed by the child’s physician after conducting a physical exam. School district reports for elementary grades (i.e., K, 2 and 4) and middle- and high-school grades (i.e., 7 and 10) are available for 2010-12 (where academic years 2010/11 and 2011/12 are combined).

*Family Income.* We have annual income tax data aggregated to the school district level for PA from 2006 through 2010. The data provide dollars of taxable income for adjusted gross income; gross compensation; business expenses; taxable compensation; interest; dividends; net profits; sale of property; estate or trust; gambling or lottery; and rents, royalties, patents, and copyrights. We will use these data to predict average total household income and average total royalty income per PA school district per year.
For NY, we have the annual adjusted gross income for each school district since 2005. Although it would be preferable to have school district-level data on taxable income across the various sources, we do not expect NY residents to experience a rise in royalty income over this period given the statewide ban on drilling.

*Geology.* Figure 1 shows the digitized map of the geological and initially predicted economic limits of the Marcellus Shale with school district boundaries. To externally validate this geological and economic Marcellus map, we have latitude and longitude of all PA Marcellus well permits filed in or before 2014 and we have created time-varying, school-district counts of well permits. We have a 2005 map of existing highways in PA, which we will use to predict royalty income based on initial (i.e., exogenous to the development of Marcellus Shale) distance by highway to the closest economic hub of Shale development. The GIS analyst has also digitized maps documenting the historical extraction of coal and shallow-well gas in PA.

*Control variables.* In some models we will add controls for the legacy of “shallow” natural gas wells and coal mining for the community members’ ownership of the mineral rights, gas companies’ violations of environmental regulations, population size, and proportion of housing in the district rented versus owned.

**Analytic Approach**

We estimate difference-in-differences models, comparing (a) differences between children in PA districts sited above geologically-rich Marcellus Shale areas and children in PA districts above geologically-poor Marcellus Shale areas relative to (b) differences between children in NY districts sited above geologically-rich shale areas with children in NY districts sited above geologically-poor shale areas.
We will use a two-sample instrumental variable approach and a differences-in-differences regression framework to determine the effect of income on childhood overweight and childhood obesity. We can view school districts above the PA Core of Marcellus experiencing income gains as having received a quasi-random “treatment” and compare them to other PA and NY districts above the Marcellus Shale. Our system of equations can be represented as follows:

(1) \[ ALRI_{ys} = \text{Year} + \text{State} + \text{Core} + \text{State*Core} + \text{State*Year} + \text{Year*Core} + \text{State*Year*Core} + V_{ys} + E_{ys} \]

(2) \[ \text{Obesity}_{ys} = \text{Year}_{y} + \hat{ALRI}_{ys} + S_{ys} + V_{ys} + E_{ys} \]

Equation (1) generates our measure of predicted average local royalty income in each school district in each year, \( ALRI_{ys} \). This equation uses a difference-in-difference model and school district-aggregated income tax data to determine how Marcellus Shale quality and state of residence predict average local royalty income (\( ALRI_{ys} \)). For example, living in the PA Core in years after drilling begins could cause a $10,000 increase in average local royalty income.

Equation (2) estimates the effect of this predicted change in average local royalty income on local outcomes. \( \text{Obesity}_{ys} \) is the prevalence of obesity in school district \( s \) in year \( y \). We will also predict the prevalence of overweight (\( \text{Overweight}_{ys} \)) in school district \( s \) in year \( y \). \( S \) is a school district fixed effect (i.e. a set of dichotomous indicators, one for each district, that control for persistent observed and unobserved differences between districts (e.g., rurality, religiosity), that may affect outcomes). In some specifications, we will also include \( V_{ys} \), a set of control variables, including size and composition of the local and school populations, as additional controls in both equations (1) and (2).
To continue our example, PA Core school districts in years after drilling begins might experience an increase in obesity rates of 0.15 standard deviations. We would then make the link to assess the causal effect of increasing income given that Marcellus Shale quality and PA/NY drilling policy are exogenous to school district choices. In this example, a $10,000 increase in average income causes a 0.15 SD increase in obesity rates for youth in these districts; the estimated effect of a $10,000 increase in $\text{ALRI}_{ys}$ in equation (2) would be 0.15.

In some specifications, we will generate $\text{ALRI}_{ys}$ using a difference-in-difference-in-differences technique to predict average local royalty income based not only on shale quality, year, and state but also proximity to Marcellus Shale economic development hubs (i.e., Williamsport in northeastern PA and Canonsburg in southwestern PA), as follows:

\[(3) \quad \text{ALRI}_{ys} = \text{Year} + \text{State} + \text{Core} + \text{State} \times \text{Core} + \text{State} \times \text{Year} + \text{Year} \times \text{Core} + \text{State} \times \text{Year} \times \text{Core} + \text{HubDistance} \times (1 + \text{Year} + \text{State} + \text{Core} + \text{State} \times \text{Core} + \text{State} \times \text{Year} + \text{Year} \times \text{Core} + \text{State} \times \text{Year} \times \text{Core}) + \text{V}_{ys} + \text{E}\]

The approach in equation (3) will allow us to make more powerful and differentiated predictions of local royalty income, but at the risk that distance from Marcellus economic hubs may not be exogenous to unobserved time-varying local characteristics. Creating estimates of $\text{ALRI}_{ys}$ using both equations (1) and (3) allows us to tradeoff between predictive power and causal identification. We will also run falsification checks using equation (3) on a subsample of districts where either (a) most families with children lived in rental housing in 2000 or (b) the district is located in coal or shallow-well gas extraction areas where families are less likely to hold mineral rights. Families in these districts should have limited access to potential royalty income. As such, equations (1) and (3) should
generate a predicted increase in income ($\hat{ALRI}_{ys}$) that is not significantly different from zero for these districts. If so, then we can use this subsample as a control group for measuring the income effects from Marcellus; any changes in outcomes we see for this group would be attributable to other effects of Shale development, not to income effects.

As in all natural experiments, a variety of threats to validity of our causal inferences exist. For this analysis, we have already anticipated risks due to changes in non-royalty income (e.g., changes in labor income), declines in environmental quality (e.g., water contamination), changes in property taxes and assessed property values, and selective in- and out-migration from these local communities. Fortunately, we can leverage our analytic strategy to minimize these threats and detect the influence of any remaining threats.

**Preliminary Results**

We are still compiling and appropriately classifying school-district data according to the area’s geological conditions. Most importantly, we are still refining the income tax data into a usable form. We have, however, created numerous geological and geographic indicators and merged those data with the Pennsylvania and New York school district obesity data. Examining these preliminary data, we can see that school districts with a higher percentage of their land area above the Marcellus core also have a higher rate of childhood obesity for both elementary school students ($r = .22, p<.001$) and middle school and high school students ($r = .17, p<.001$) in Pennsylvania. But in New York the percentage of a district above the Marcellus Shale core is only significantly related to the rate of obesity among middle- and high-school student ($r = .08, p<.05$). Rates of overweight are not correlated with presence above the Marcellus Shale core.
Future Directions

Before PAA in Spring 2016, we will complete the full analyses described above, as well as perform the various robustness checks outlined. In addition, we intend to merge and harmonize National Center for Education Statistics Common Core data regarding school districts to incorporate more control variables for observable differences across districts. We are excited about this research and look forward to pursuing this exciting research question.
REFERENCES


