The Effect of the Fresh Fruit and Vegetable Program on Childhood Obesity[*]

Yiwei Qian,
PHD Candidate
Department of Economics
University of Arkansas
yqian@uark.edu

Rodolfo M. Nayga Jr.,
Professor and Tyson Chair in Food Policy Economics
Department of Agricultural Economics & Agribusiness
University of Arkansas Division of Agriculture
rnayga@uark.edu

Michael R. Thomsen,
Associate Professor
Department of Agricultural Economics and Agribusiness
University of Arkansas Division of Agriculture
mthomsen@uark.edu

Heather Rouse, PhD,
Assistant Professor, Center for Applied Research and Evaluation
Director of Health Policy Research, Arkansas Center for Health Improvement
University of Arkansas for Medical Sciences
HLRouse@uams.edu

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2013 AAEA & CAES Joint Annual Meeting, Washington, DC, August 4-6, 2013.

[*] This research was funded by the Agriculture and Food Research Initiative of the USDA National Institute of Food and Agriculture, grant number 2011-68001-30014. This work was also partly supported by the National Research Foundation of Korea (NRF-2011-330-B00074).

The Effect of the Fresh Fruit and Vegetable Program on Childhood Obesity

Abstract

This paper investigates how the Fresh Fruits and Vegetable Program (FFVP), a national program which provides funding for the distribution of free fresh fruits and vegetables to students in participating schools, affects childhood obesity. Using a panel data set, we combine matching methodology and difference-in-differences analysis to estimate the effect of the FFVP on childhood obesity outcomes. The results suggest that estimates of the FFVP effect are very sensitive to use of different matching methods. With the use of a stricter matching method, the estimate of the FFVP effect is negative and significant, indicating that the program reduces children's body mass index. Less strict matching methods yield opposite results.

Keywords: body mass index, childhood obesity, fresh fruit and vegetable program, matching, difference-in-differences

The Effect of the Fresh Fruit and Vegetable Program on Childhood Obesity

Obesity prevalence among children and adolescents in the United States has significantly increased during the past few decades. It is now a major health problem and poses a challenge for government, public health agencies and medical communities. Approximately 13 million U.S. children and adolescents are considered obese[1], with a body mass index (BMI) at or above the 95th percentile. Ogden et al. (2010) indicated that from 1980 to 2008, obesity rates nearly tripled — from 6.5% to 19.6% — for children aged 6 to 11 and more than tripled for adolescents age 12 to 19—from 5% to 18.0%. Obese adolescents have an 80% chance of becoming obese adults, which places them at greater risk for health problems throughout life (Guo and Chumlea 1999).

Epstein et al. (2001) indicated that increasing fruit and vegetable intake would decrease high-fat/high-sugar intake for children and their parents, and could be a useful approach to preventing childhood obesity. However, children and adolescents, and particularly those from low-income households, do not consume the recommended amounts of fruits and vegetables. The United States Department of Agriculture (USDA) guidelines recommend that children eat 6-13 serving of fruits and vegetables each day, but US children only eat 3.5 servings per day on average (Jamelske et al. 2008). Thus, strategies that encourage the consumption of healthier foods such as fruit and vegetables may be one way to address childhood obesity.

[1]Obesity is defined as body mass index (BMI) at or above the 95th percentile based on the 2000 Centers for Disease Control and Prevention BMI-for-age growth charts. Children with BMI between the 85th and 95th percentile are classified as overweight.

The USDA created the Fresh Fruit and Vegetable Program (FFVP) in 2002. This program is intended to increase fruit and vegetable consumption among students in the nation's poorest elementary schools by providing reimbursement to schools for offering fresh fruits and vegetables, free to students, throughout the school day and separately from lunch and breakfast meals. According to the USDA Food Nutrition Service (2010), the objectives of FFVP include: (1) to create healthier school environments by providing healthier food choices; (2) expand the variety of fruits and vegetables available to children; (3) increase children's fruit and vegetable consumption; and (4) make a positive difference in children's diets to impact their present and future health.

We focus this study on children in the state of Arkansas. Arkansas is an interesting case to study since it has one of the highest childhood obesity rates in the US. The National Survey of Children's Health indicated that in Arkansas, about 32.9% of 10-17 year old children were either obese or overweight in 2005 and this percentage increased to 37.5% in 2007[2]. Additionally, Arkansas was the first state to legislatively mandate the measurement and collection of BMI for every public school student starting in 2003 (Act 1220). Measured annually, these data provide a unique opportunity to study child weight status and the programs and policies designed to impact BMI.

Arkansas schools began participating in the FFVP during the 2008-2009 school year. The FFVP is primarily administered through the Arkansas Department of Education (ADE). Presently, for a school to participate in the FFVP, the school must also participate in the National School Lunch Program (NSLP) and at least 50 percent of

[2] Source: Childhood Obesity Action Network. State Obesity Profiles, 2008.

students must be eligible for the free and reduced lunches. This is to ensure that the program benefits low-income students who otherwise would have fewer opportunities to consume a variety of fruit and vegetables. All students in participating schools are provided fruits and vegetables. Schools are selected based on an application process and program funds are used to reimburse schools for providing fruit and vegetables as snacks at the rate of $50 to $75 per student per year (USDA Food and Nutrition Service, 2010). The average amount of funding per school during the 2008-2009 and 2009-2010 school years was $27,334 and $21,382[3], respectively. However, nearly twice as many schools participated in the 2009-2010 school year and so the decrease in average funding is not indicative of reduced reach of the program.

There is scant literature, however, on the effectiveness of the FFVP to reduce childhood obesity. Most of the studies on FFVP are focused on the program's impact on fruit and vegetable consumption. For example, Jamelske et al. (2008) surveyed 784 students who participated and 384 students who did not participate in the FFVP in Wisconsin and found that FFVP participants reported an increased willingness to eat fruits and vegetables compared to non-participants. Davis et al. (2009) surveyed 1,515 high school students who participated in the program and 1,377 high school students who did not participate and compared the fruit and vegetable intakes of both groups. Their results indicated that FFVP participants were more likely than non-participants to consume fruit, juice, and vegetables in amounts recommended by dietary guidelines. Ohri-Vachaspati, Turner and Chaloupka (2012) also conducted a study on 620 public

[3] Source: Arkansas Department of Education (ADE) Child Nutrition Unit.

elementary schools participating in the National School Lunch Program during 2009-2010 and found that FFVP participating schools were significantly more likely (odds ratio 2.07) to serve fresh fruit during lunch meals than FFVP non-participating schools.

Bartlett et al. (2013) evaluated the effect of FFVP on fruit and vegetable consumption and total energy intake for children. Using regression discontinuity, they estimated that the program increased average fruit and vegetable consumption of students in participating schools on FFVP days by approximately one-quarter of a cup per day. They also found no significant increase in total energy intake, which suggests that the increase in fruit and vegetable consumption replaced the consumption of other foods. Boukhris (2007) investigated FFVP participation in Texas and found that there was no significant difference between the FFVP schools and non-FFVP schools in fruit and vegetable expenditures in 2006, but in 2007 the FFVP schools had higher fruit and vegetable expenditures than non-FFVP schools.

Given the promising results of these past studies linking program participation to improvements in fruit and vegetable consumption, it would also be interesting to examine the effect of FFVP on childhood obesity. To our knowledge, no other study has evaluated this issue. In this paper, we use a unique panel dataset that includes measured body mass index (BMI) of school children in Arkansas. We employ difference-in-differences and matching methods to identify the effect of FFVP on children's BMI. Our results suggest that FFVP effects are sensitive to the use of matching methods, but when using stricter matching methods (i.e., matching methods that produce more balance), FFVP participation reduces children's BMI measures.

The next section describes the data sources and the variables used in the analysis. Section 3 discusses the empirical strategy we used to identify the effect of FFVP participation on children's BMI. Section 4 presents the results and describes their sensitivity to different matching methods. Finally, section 5 concludes and offers suggestions for future research.

Data

Data Sources

Our data come from three different sources. First, we use FFVP participation data from 2008-2010. These data were obtained from the Arkansas Department of Education (ADE) Child Nutrition Unit and include program participation status and funding information by school and year. There were 24 FFVP schools in the 2008-2009 school year and 47 FFVP schools in the 2009-2010 school year. Second, we use the Arkansas BMI dataset for 2007 to 2010. This is a unique panel dataset at the individual student level that includes child weight and height data collected by trained personnel in the public schools and maintained through legislative mandate at the Arkansas Center for Health Improvement (ACHI) (Justus et al. 2007). BMI is calculated as a ratio ([weight in pounds / (height in inches)^2] $\times$ 703) and then converted to age-gender specific z-scores according to the Centers for Disease Control and Prevention guidelines (CDC 2013). Measures used for this analysis included the BMI z-score[4], BMI percentile[5], gender, race, home language, and free or reduced lunch program participation status. Additionally, ACHI personnel geo referenced and interfaced these data with food store locations so that our final dataset provided measures of the food environment around the children's home and schools. Only children in even-numbered grades (kindergarten through 10th grade) were consistently measured across all years during the period of our study. For this reason, we include students in kindergarten, second, fourth, sixth and eighth grades in our study. Third, we used demographic characteristics data from the American Community Survey's (ACS) 2006-2010 five-year estimates. These include data on proportion of population by race, income level, education, work status and other neighborhood measures for the census block group of the child's residence. We use these as control variables in our models.

Variable Definitions and Descriptive Statistics

The choice of control variables for the matching and the regression models is an important consideration in our study. Matching is a "data hungry" technique in terms of the number of variables required to find matched groups. In our study, the control variables are based on the factors which are hypothesized to affect our outcome variable, children's BMI. Table 1 exhibits the description of the variables used in the analysis.

[4] BMI z-score is defined as a deviation of the value for an individual from the mean value of the reference population divided by the standard deviation for the reference population.
[5] BMI percentile is a value of a cumulative probability distribution of BMI z-score

One important factor for obesity is income level. Wang (2001) indicates that for 10-18 year old children in the US, the obesity and overweight rate is 32.7% for low-income households, 25.5% for middle-income households and 19% for high-income households. Casey et al. (2001) also analyzed data from 5,669 children (0-17 year old) from 3,790 households. They found that children in low-income families reported a higher obesity and overweight rate (46.7%) than children in high-income families (31.5%). Singh, Siahpush and Kogan (2010) analyzed obesity outcomes for more than 44,000 children from 2003-2007 and found that obesity prevalence for children below the poverty threshold was 27.4%, 2.7 times higher than the prevalence for children with family income exceeding 400% of the poverty threshold. One reason for the inverse relationship between obesity rates and income is that low-income communities often lack access to stores that sell fresh fruit and vegetables and have instead stores that sell foods low in nutritional value. Haynes-Maslow et al. (2013) identified 6 major community-level barriers to accessing fruits and vegetables. These are cost, transportation, quality, variety, the food environment, and societal norms on food. Their research showed that in lower income communities, access to fresh fruit and vegetables can be difficult because of the lack of affordable transportation options. Moreover, the quality and variety of fresh fruit and vegetables can be limited in lower income areas.

Given the findings from these past studies, we control for income using the student’s free and reduced lunch participation status. Additional income controls at the census block group level include the proportion of population below the poverty level, median household income, and median value of owner occupied housing units.

To measure and control for access to healthy foods, we computed the distance between the student's residence and the nearest large grocery store that contained a fresh produce department. Grocery stores and their locations in Arkansas, by year, were obtained from Dun and Bradstreet. We adopted the low access area criteria found in the USDA/ERS Food Desert Locator[6]. That is, students living in urban census block groups were classified as having low-access to healthy foods if their residence was more than one mile from a large supermarket. Students in rural block groups were classified as low access if this distance was greater than ten miles. Food access is also affected by transportation options and so controls are included for the proportion of population that uses public transport for commutes to work and for the proportion of families with no vehicle availability.

Educational level, working status and marital status of parents are also important factors for childhood obesity. For example, Nayga (2000) has shown that schooling can influence obesity outcomes. His results also suggested that health knowledge decreases the probability of an individual becoming obese. Singh, Siahpush and Kogan (2009) found that obesity prevalence for children with parents having less than 12 years of education was 30.4% in 2007, 3.1 times higher than the prevalence for children with parents with a college degree. Obesity prevalence also increased significantly among children from single-mother households from 18.9% in 2003 to 21.9% in 2007. Anderson, Butcher and Levine (2003) investigated whether children are more or less likely to be

[6] http://www.ers.usda.gov/data/fooddesert/about.html.

overweight if their mothers work and their results indicated that a child is more likely to be overweight if his/her mother worked more intensively.

We do not have information about the education level, working status, and marital status of parents of the students in our sample, but we are able to measure these for the neighborhood of the child's residence using census block group data from the American Community Survey. All these control variables are listed in Table 1.

The BMI data do, however, include some important individual-level control variables. These include age in months, gender, ethnicity (*White, Black, Hispanic, Native, or Asian*), language spoken at home, and free or reduced lunch participation status. There are 1,116 individuals who participated in FFVP for two school years in 14 participating schools and 62,868 individuals in 836 schools which did not participate in FFVP. We drop all of the individuals who participated in FFVP for only one school year because it is not likely that there would be a measureable effect during the first year of program implementation.

Table 2 presents the descriptive statistics of the variables for the entire sample and for those in the treatment group (i.e., students in schools which participated in FFVP). For the entire sample, the average BMI z-score is 0.696. Age range of the students in months was 58 months to 196 months old. 50.7 percent were male, 66.6 percent were White, and 20.8 percent were Black. The proportions of students who were Hispanic, Native, and Asian were 9.6 percent, 0.7 percent, and 1.8 percent, respectively. 44.9 percent of students in the sample participated in the free lunch program and 9.9 percent participated in reduced lunch program. 7.7 percent of students spoke Spanish, 62.1 percent lived in an urban area, and 37.9 percent lived in a rural area. 31.4 percent of students in the sample were considered to have low access to grocery stores.

For the treatment group, the average BMI z-score is 0.761, which is higher than the sample at large. This, however, is expected given the eligibility requirements of the FFVP as discussed previously (i.e., the school must participate in the NSLP and at least 50 percent of students must be eligible for the free and reduced lunch program). Indeed, 51 percent of students in the treated sample participated in free lunch program and 11.8 percent participated in the reduced lunch program. The treated group also had lower average values on the neighborhood income measures and on the food access measures, which is further evidence that FFVP is more likely to serve those students who come from low income areas and who lack sufficient access to grocery stores.

Methodology

A major concern in assessing the effect of FFVP is that FFVP participation by schools is not randomly assigned, so it is possible that schools self-selected into the program. Hence, the characteristics of FFVP participating schools could be quite different from those of non-participating schools. It is also possible that some unobserved factors could influence both FFVP participation and obesity outcomes (e.g., school health related preferences and programs, parental factors). The availability of panel data allows us to address these endogeneity issues, along with the use of a difference-in differences (DID) framework. To further alleviate concerns regarding the comparability of the treatment and control groups and to limit model dependence (Campbell et al. 2011; Islam 2011), we also use matching techniques prior to running our DID panel models. Heckman, Ichimura, and Todd (1997) concluded that DID matching helps control for heterogeneity in initial conditions and also controls for unobserved determinants of participation. Hence, we attempt to account for potential selection biases by combining matching, DID, and panel estimation methodologies in our analysis. We run our panel DID models using the unmatched sample and several matched samples resulting from different matching methods.

Our panel data includes individual student level observations from 2007-2010. Since FFVP in Arkansas started during the 2008-2009 school year, we use the 2007-2008 school year as period 1 (or the before period) and the 2009-2010 school year, the 2nd year of the FFVP implementation, as period 2 (or the after period). We then define the treatment group as those students who participated in FFVP in both 2008-2009 and 2009-2010 school years and the control group as those students who did not participate in FFVP from 2007-2010.

Matching

The main idea of matching is to find a group of control individuals that are similar to the treated individuals in all pre-treated characteristics. We use propensity score matching (PSM) and coarsened exact matching (CEM) to match the treated and control groups.

Propensity Score Matching (PSM)

Rosenbaum and Rubin (1983, 1985) introduced Propensity Score Matching (PSM) as a matching method to construct a statistical comparison group that is based on a model of the probability of participating in the treatment, T, conditional on observed characteristics, X, or the propensity score: P(X) = Pr(T = 1|X). To get the propensity score, first we run a standard logit model where the dependent variable is the treatment variable, which is *FFVP* participation, and the independent variables are a set of control variables.

One of the most frequently used matching techniques is *nearest-neighbor matching*, where each treatment unit is matched to the comparison unit with the closest propensity score. Rosenbaum and Rubin (1985) and Becker and Ichino (2002) introduced the structure of *nearest-neighbor matching.* Denote by *C(i)* the set of control units matched to the treated unit *i* with an estimated value of the propensity score of p_i and the control unit *j* with propensity score of p_j. The nearest neighbor matching sets can be calculated by:

$$C(i) = \min_{j} \left| p_i - p_j \right| .$$

Nearest-neighbor matching within n neighbors means that for each matched treated unit, there are n matched control units which have the n closest propensity scores. In our analysis, we choose the *nearest-neighbor matching* within 2 neighbors and within 3 neighbors.

The other matching algorithm we choose is *Mahalanobis matching.* Rosenbaum and Rubin (1985) introduced the structure of PSM based on the Mahalanobis distance.

The Mahalanobis distance is the distance between two *N* dimensional points scaled by the statistical variation in each component of the point. For example, if x_1 and x_2 are two points from the same distribution with covariance matrix, Σ, then the Mahalanobis distance can be expressed as:

$$D(x_1, x_2) = (x_1 - x_2)'\Sigma^{-1}(x_1 - x_2).$$

Mahalanobis matching has 3 steps: First, we calculate the propensity scores for every subject; second, the first subject from the treatment group is selected. Subjects in the control group whose propensity score is within the caliper (one quarter of standard deviation of the logit of the propensity score as suggested by Rosenbaum and Rubin 1985) are identified as initial matched candidates; Third, Mahalanobis distances are calculated between the subject in the treatment group and those initially selected subjects in the control group. The subject with the smallest distance to the subject in the treatment group is selected as a final matched candidate. The matched pair is then removed from the pool, and the process will repeat for the next subject in the treatment group.

In our study, we use *Mahalanobis matching* without calipers and *Mahalanobis matching* with calipers of 0.05, 0.075, and 0.1. The use of a caliper provides stricter matches because observations are matched only if their absolute distance in propensity scores is smaller than the caliper. Hence, a treated individual will remain unmatched if the nearest observation in the control group is outside of the bound set by the caliper.

Coarsened Exact Matching (CEM)

We also utilize a strict matching method called *coarsened exact matching* (CEM) (Iacus, King and Porro 2011; Iacus, King and Porro 2012; Blackwell, Iacus, King, and Porro 2009). The main motivation for CEM is that while exact matching always provides perfect balance, it typically produces few matches due to the curse-of-dimensionality. The idea of CEM is to temporarily coarsened each variable into substantively meaningful groups, and then exact match on these coarsened data. Afterwards, the original (uncoarsened) values of the matched data are retained.

CEM requires three steps: (1) Coarsen each of the original variables in X` to C(X`). (2) Apply exact matching to C(X`), which involves sorting the observations into strata, and each with unique values of C(X`). (3) Assign these strata to the original data, X`, and drop any observation whose stratum does not contain at least one treated and one control unit.

The advantage of CEM is obvious in that it generally provides stricter matching criteria compared to PSM and it also allows the analyst to add continuous variables as control covariates. For PSM, if a lot of continuous variables are used in the matching, it is possible that the matched samples have close propensity scores but not close values on these continuous variables. However, for the CEM, the value of every matching variable needs to be exactly the same (after coarsening). In our research, we let the coarsening algorithm cut the range of the continuous variable into equal intervals of length.

To summarize, our matching strategy includes the use of the following matching procedures: *nearest-neighbor matching* within 2 neighbors, *nearest-neighbor matching* within 3 neighbors, *Mahalanobis matching* without calipers, *Mahalanobis matching* with calipers of 0.05, 0.075, and 0.1 and *coarsened exact matching.*

The Imbalance Test

After conducting the matching between those in the control and treatment groups using the seven different matching methods discussed above, we need to test the degree of imbalance in the covariates in the two groups. The goal of measuring imbalance is to summarize the differences in the multivariate empirical distribution of the pretreatment covariates for the treatment group and matched control group. That is, we wish to assess how similar the control and treated groups are based on given characteristics. In our study, we choose the imbalance test introduced by Iacus, King and Porro (2011); i.e., the $\mathcal{L}_1$ statistic as a comprehensive measure of global imbalance.

To build this measure, Iacus, King and Porro (2011) obtained two multidimensional histograms by direct cross-tabulation of the covariates in the treated and control groups, given a choice of bins for each variable. Let $H(X_1)$ denote the set of distinct values generated by the bins chosen for variable X_1, i.e., the set of intervals into which the support of variable X_1 has been cut. Then, the multidimensional histogram is constructed from the set of cells generated as $H(X_1) \times \cdots \times H(X_k) = H(X) = H$. Set f and g as the relative empirical frequency distributions for the treated and control units, respectively and record the k-dimensional relative frequencies for the treated $f_{\ell_1 \ldots \ell_k}$ and control $g_{\ell_1 \ldots \ell_k}$ units. The measure of imbalance is the absolute difference over all the cell values:

$$\mathcal{L}_1(f,g) = \frac{1}{2}\sum_{\ell_1\ldots\ell_k\in H(X)} |f_{\ell_1\ldots\ell_k} - g_{\ell_1\ldots\ell_k}|.$$

The $\mathcal{L}_1$ measure offers an intuitive interpretation, for any given set of bins: if the two empirical distributions are completely separated (up to H), then $\mathcal{L}_1$= 1; if the distributions exactly coincide, which indicates perfect global balance, then $\mathcal{L}_1$= 0. In all other cases, $\mathcal{L}_1 \in (0, 1)$. If $\mathcal{L}_1$ = 0.7, then 30% of the area under the two histograms overlap. Thus, if we want to choose the best matching methodology, we need the $\mathcal{L}_1$statistic to be as low as it can be.

Difference-in-Differences Design

After matching, we run a difference-in-differences regression on these new matched samples. The DID equation is:

$$(1) \qquad Y_{it} = \alpha + \beta_1 d_{1it} + \beta_2 FFVP_{it} + \beta_3 DID_{it} + \delta X'_{it} + \epsilon_{it}$$

where Y_{it} denotes the outcome variables (i.e., BMI z-score and BMI percentile) for individual i at period t; d_{1it} is a dummy variable for the different periods and takes the value of 1 if observations are from period 2 and a value of 0 otherwise ; $FFVP_{it}$ is a dummy variable for FFVP participation and takes a value of 1 if the individual participated in FFVP for both school years and a value of 0 if she did not participate for both school years; DID_{it} is the DID interaction term; X'_{it} is a vector of control variables and ϵ_{it} is the error term.

Since FFVP participation($FFV\mathrm{P}_{\mathrm{it}}$) starts in second period, $\mathrm{FFVP}_{\mathrm{it}} = 1$ can only be observed in the second period, which means that a prerequisite of $FFVP_{it} = 1$ is $\mathrm{d}_{1\mathrm{it}} = 1$.

Hence, the DID interaction term will equal to 1 only when both d_{1it} and $FFVP_{it}$ are both equal to 1. This is the reason why our specification above does not include the treatment dummy variable. To test the robustness of the results, we run the DID regression using both fixed effects and random effects panel estimation and using unmatched and matched samples.

Results

To describe the main results, first we need to compare the estimates of imbalance test from each matching method. These are reported in Table 3. Note that the lower the $\mathcal{L}_1$ statistic, the more similar are the treatment and the control groups on average, which also indicates that the control and treatment samples are better matched. Results depicted in table 3 indicate that if we do not use any matching method, the $\mathcal{L}_1$ statistic is 0.998 and the number of observations in the control group is 62,868. This provides a baseline reference for the unmatched data, which we can use as a point of comparison between matching solutions. As expected, the number of observations in the control group shrinks to 3,045 and the $\mathcal{L}_1$ statistic reduces to 0.992 when using nearest neighbor matching with 3 neighbors. The $\mathcal{L}_1$ statistic and the number of observation continue to decrease if we use nearest neighbor matching with 2 neighbors. When we use Mahalanobis matching without caliper, the $\mathcal{L}_1$ statistic further falls to 0.907 and the number of observations in the control group declines to 754. The number of observations in the treatment group remains at 1,116 under each of these matching strategies.

Once we add a caliper 0.1 to the Mahalanobis matching algorithm, the matching becomes stricter. The $\mathcal{L}_1$ statistic becomes 0.616 and the number of observations in the control group falls to 200 while the number of observations in the treatment group falls to 253. This means that the algorithm could not find matches within the control group for the remaining treated observations. If we reduce the value of the caliper for Mahalanobis matching, the matching becomes even stricter and the $\mathcal{L}_1$ statistic becomes 0.604 for a caliper of 0.075 and 0.533 for a caliper of 0.05 with the number of observation being further reduced. Finally, when we use coarsened exact matching, the $\mathcal{L}_1$ statistic becomes 0.541 and the number of observations in the control and treated groups are 161 and 151, respectively.

While the use of stricter matching routines significantly decreased the number of observations in both the control and treatment groups, the resulting matches still include students widely distributed across different schools. For example, in the CEM sample, the 151 individuals in the treatment group come from 13 schools (out of total of 14 FFVP participating schools) while the 161 individuals in control group come from 76 schools. The same results are found in the matched groups using the Mahalanobis matching technique. Thus, given that the results from the imbalance test suggest that the coarsened exact matching (CEM) and the Mahalanobis matching with caliper 0.05 provide the best matches between the control and treated groups, we rely more on these matching techniques but also report results of the DID matching panel estimates using the other PSM methods for comparative purposes.

Table 3 also presents the estimates of our panel DID models using unmatched and matched samples and BMI z-score as the outcome measure[7]. Using unmatched samples, our results indicate that the DID coefficient is 0.15 for fixed effects and 0.148 for random effects and both are significant at the 0.01 level. This suggests that a student's BMI z-score will increase if she participates in the FFVP program, which is inconsistent with our prior hypothesis. Similar conclusions would be reached from DID estimations using matched samples based on the nearest neighbor algorithms. Again, the DID coefficient is positive and significant at the 0.01 level. This is also true of the DID estimates after Mahalanobis matching with no caliper threshold. As we previously mentioned, the original size of the treatment group is much smaller than the control group and the simple average BMI z-score for the treatment group is higher than that of the control group. Hence, selection bias is likely in the unmatched or weakly matched samples. The estimates of the imbalance test for these four matching methods are quite close, and so it is not surprising that the coefficient of each DID interaction term is similar across these matching strategies.

In contrast, the DID estimates using the Mahalanobis matching with the caliper and from CEM are markedly different. When using the matched samples from the Mahalanobis matching with the 0.1 caliper, the DID coefficient is -0.055 for fixed effects and -0.045 for random effects. These are, however, not significantly different from zero. With a 0.075 or 0.05 caliper, the DID coefficient is still negative and insignificant. However, when using the CEM sample, the coefficients become -0.15 in the fixed effects

[7] The complete set of estimates are available from the corresponding author.

model and -0.139 in the random effects model and both are now statistically significant at the 0.05 level.

Table 4 reports the comparison of descriptive statistics for both control and treatment group after using CEM. As noted above, in the unmatched data there are important differences between the treatment and control groups in mean values for several individual and neighborhood controls. After using CEM, however, the average values of these variables become much closer in the treated and control groups. Since these variables are potentially important determinants of FFVP school participation, reducing the gap in these variables between the treated and control groups can also reduce selection bias issues. Hence, based on the results of the imbalance test and the evidence from the descriptive statistics, the DID estimate using the CEM sample provides the more trustable result that FFVP participation indeed reduces students' BMI z-scores.

To test the robustness of our findings, we also ran our models using BMI percentiles as the outcome measure instead of BMI z-score (also in Table 3). Results are similar to those discussed above. When using unmatched samples and matched samples from nearest neighbor matching and Mahalanobis matching, the DID coefficients of the FFVP effect are always positive and significant at the 0.01 level. However, when using matched samples from Mahalanobis matching with caliper 0.1 and 0.075, the DID coefficient is negative but not statistically significant. For Mahalanobis matching with caliper 0.05, the random effect estimate is negative and significant at the 0.1 level. Remember that the $\mathcal{L}_1$ statistic from Mahalanobis matching with a caliper of 0.05 is very close to that from CEM so this result strengthens our confidence in the estimated FFVP effect obtained from the CEM sample and the conclusion that better matching provides more trustable results. When using the CEM matched sample, the coefficient is negative and significant at the 0.05 level. The magnitude of the FFVP effect is -0.0417 for fixed effects and -0.037 for random effects, which means that those who participate in FFVP have 4.17% (for fixed effects) and 3.7% (for random effects) lower BMI percentiles than those who do not participate in the FFVP.

Conclusion

While a number of studies have examined the effect of FFVP on fruit and vegetable consumption of children, no other known study has examined the FFVP's effect on childhood obesity. To fill this void, we use a relatively unique panel dataset with measured BMI of school children in Arkansas. Arkansas is an interesting case to study since it has one of the highest childhood obesity rates in the US. It is also the first state to mandate measurement of weight and height of school children. We used a panel difference-in-difference estimation procedure to examine the effect of FFVP participation on students' BMI z-score and percentile. Before the panel DID estimation, however, we used several matching methods such as Propensity Score Matching and Coarsened Exact Matching to match FFVP participants to non-participants. We then estimated both fixed effects and random effects DID models using unmatched and matched samples. In addition to being the first to examine the effect of FFVP participation on childhood obesity, another contribution of this paper is the investigation of the sensitivity of the impact estimates to the use of different matching techniques.

Our results show that while the FFVP effects on weight are unexpectedly positive and statistically significant using unmatched samples and matched samples with less balance on the covariates, they are negative when using stricter matching techniques such as the Mahalanobis with calipers and CEM, both of which provided more balance in characteristics between the treated and control groups. Specifically, our panel DID results using matched samples from these two techniques suggest that FFVP participation can reduce BMI percentile by 4 percent, *ceteris paribus*. Considering that the cost for each student in participating schools has been estimated to be only 50-75 dollars per year, the FFVP could have a relatively high benefit-cost ratio.

Our study represents a first attempt at analyzing the effect of FFVP on childhood obesity. Given that the FFVP has only been implemented in Arkansas since 2008, more research is needed to draw more definitive conclusions. For instance, future research should test the robustness of our findings when more data become available (i.e., more years of implementation). It would also be important to examine whether our findings will hold true in other states that have implemented the FFVP program in schools.

References

Anderson, P.M., K.F. Butcher, and P.B. Levine. 2003. "Maternal employment and overweight children." *Journal of Health Economics,* 22 (3): 477-504.

Bartlett S., L. Olsho, J. Klerman, K.L. Patlan, M. Blocklin, P. Connor, K. Webb, L. Ritchie, P. Wakimoto, P. Crawford. 2013. *Evaluation of the Fresh Fruit and Vegetable Program (FFVP).* Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service. Available at: http://www.fns.usda.gov/Ora/menu/Published/CNP/FILES/FFVP.pdf

Becker, S.O., and A. Ichino. 2002. "Estimation of average treatment effects based on propensity scores." *The Stata Journal* 2(4): 358-377.

Blackwell, M., S. M. Iacus, G. King, and G. Porro. 2009. "cem: Coarsened exact matching in Stata." *Stata Journal, 9*(4): 524-546.

Boukhris, T. T. 2007. "A public response to childhood obesity: Evaluating the fresh fruit and vegetable program in Texas schools." PhD dissertation, Texas State University.

Campbell, B.L., R. M. Nayga, J. L. Park, and A. Silva. 2011."Does the National School Lunch Program improve children's dietary outcomes?" *American Journal of Agricultural Economics,* 93(4): 1099-1130.

Casey, P.H., K. Szeto, S. Lensing, M. Bogle, and J. Weber. 2001."Children in food-insufficient, low-income families: prevalence, health, and nutrition status." *Archives of Pediatrics & Adolescent Medicine,* 155(4): 508.

Center for Disease Control and Prevention. 2013. *About BMI for Children and Teens.* Available at: http://www.cdc.gov/healthyweight/assessing/bmi/childrens_bmi/about_childrens_bmi.html

Davis, E.M., K.W. Cullen, K.B. Watson, M. Konarik, and J. Radcliffe. 2009. "A fresh fruit and vegetable program improves high school students' consumption of fresh produce." *J Am Diet Assoc,* 109(7): 1227-31.

Epstein, L.H., C.C. Gordy, H.A. Raynor, M. Beddome, C. K. Kilanowski, and R. Paluch. 2001. "Increasing fruit and vegetable intake and decreasing fat and sugar intake in families at risk for childhood obesity." *Obesity Research,* 9(3): 171-178.

Guo, S. S., and W. C. Chumlea. 1999. "Tracking of body mass index in children in relation to overweight in adulthood." *The American Journal of Clinical Nutrition* , 70(1): 145s-148s.

Haynes-Maslow, L.E. Parsons, S.B. Wheeler, L.A. Leone, 2013. "A qaualitative study of perceived barriers to fruit and vegetable consumption among low-income populations, North Carolina, 2011." *Preventing Chronic Disease* 10.

Heckman, J.J., H. Ichimura, and P.E. Todd. 1997. "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme." *The Review of Economic Studies,* 64(4): 605-654.

Iacus, S.M., G. King, and G. Porro. 2012. "Causal inference without balance checking: Coarsened exact matching." *Political Analysis,* 20(1): 1-24.

——. 2011. "Multivariate matching methods that are monotonic imbalance bounding." *Journal of the American Statistical Association*, 106(493): 345-361.

Islam, A. 2011. "Medium-and long-term participation in microcredit: An evaluation using a new panel dataset from Bangladesh." *American Journal of Agricultural Economics*, 93(3): 847-866.

Jamelske, E., L. A. Bica, D. J. McCarty, and A. Meinen. 2008. "Preliminary findings from an evaluation of the USDA fresh fruit and vegetable program in Wisconsin schools." *Wisconsin Medical Journal (WMJ)*, 107(5): 225.

Justus, M.B., K.W. Ryan, J. Rockenbach, C. Katterapalli, and P. Card-Higginson. 2007. "Lessons learned while implementing a legislated school policy: body mass index assessments among Arkansas's public school students." *Journal of School Health*, 77(10): 706-713.

Nayga, Rodolfo M. 2000. "Schooling, health knowledge and obesity." *Applied Economics*, 32(7): 815-822.

Ogden, C., M.D. Carroll, L.R. Curtin, M.M. Lamb, and K.M. Flegal. 2010. "About childhood obesity." *JAMA*, 303(3): 242-249.

Ohri-Vachaspati, P., L. Turner, and F.J. Chaloupka. 2012. "Fresh fruit and vegetable program participation in elementary schools in the United States and availability of fruits and vegetables in school lunch meals." *Journal of the Academy of Nutrition and Dietetics*, 112(6): 921-926.

Rosenbaum, P.R., and D.B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika,* 70(1): 41-55.

——. 1985. "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score." *The American Statistician,* 39(1): 33-38.

Singh, G.K., M. Siahpush, and M.D. Kogan. 2010. "Rising social inequalities in US childhood obesity, 2003–2007." *Annals of Epidemiology,* 20(1): 40-52.

United States Department of Agriculture, Food and Nutrition Service. 2010. *Handbook for Fresh Fruit and Vegetable Program.* Available at: http://www.fns.usda.gov/cnd/ffvp/handbook.pdf.

Wang, Y. 2001. "Cross-national comparison of childhood obesity: the epidemic and the relationship between obesity and socioeconomic status." *International Journal of Epidemiology,* 30(5): 1129-1136.

Table 1. Description of Variables in the Study

Variables	**Description**
Outcome Variables	
Bmi z-score	Individual's BMI z-score
Bmi percentile	Individual's BMI percentile
Treatment Variables	
D1	Binary indicator (if the period is on 2009−2010 then =1, if 2007-2008 then= 0)
FFVP	Binary indicator (if participating FFVP for two school years then =1, if not participating for both school years then =0)
DID	D1* FFVP (DID interaction term)
Control Variables	
Age*	Age of student in months
White *	Binary indicator (if individual is White then =1, 0 otherwise)
Black *	Binary indicator (if individual is Black then =1, 0 otherwise)
Hispanic	Binary indicator (if individual is Hispanic then =1, 0 otherwise)
Native	Binary indicator (if individual is Native then =1, 0 otherwise)
Asian	Binary indicator (if individual is Asian then =1, 0 otherwise)
Male*	Binary indicator (if individual is male then =1, 0 otherwise)
Spanish	Binary indicator (if individual's language is Spanish then =1, 0 otherwise. English is the base)
Free*	Binary indicator (if individual participated in free lunch then =1, 0 otherwise)
Reduced*	Binary indicator (if individual participated in reduced lunch then =1, 0 otherwise)
Urban*	Binary indicator (if individual lived in urban area=1, 0 otherwise)
Lowaccess*	Binary indicator that describes the accessibility to large grocery stores. It takes the value of one for urban students living more than one-mile from a large grocery store and for rural students living more than 10 miles from a large grocery store.
Hispanic_prp*	Proportion of population which is Hispanic (base is White)
Black_prp*	Proportion of population which is Black (base is White)
Native_prp*	Proportion of population which is Native (base is White)
Asian_prp*	Proportion of population which is Asian (base is White)
Other_prp*	Proportion of population which is other races (base is White)
Publictrans_prp*	Proportion of population that use public transportation to work
Singlemother_prp*	Proportion of families that have children under 18 with female householder with no husband present
Highschool_prp*	Proportion of population with high school degree
Somecollege_prp*	Proportion of population with some college or an associate's degree
Collegeplus_prp*	Proportion of population with college and post-graduate degrees
Asianandpacific_prp*	Proportion of households which speak Asian and Pacific languages
Incomebelowpoverty	Proportion of population below the poverty level
Workingmother_prp	Proportion of families that have children under 18 with mother in the labor force
Novehicle_prp*	Proportion of families with no vehicle availability
Vacant_prp*	Proportion of housing units that are vacant
Medincome*	Medium household income ($'000)
Medhousevalue*	Median value for owner occupied housing units ($'000)
Medyearbulid *	Median year residential structures were built

Note: The control variables highlight as * are used for both matching and DID. Others are used only for matching.

Table 2. Descriptive Statistics for Variables Used in the Study

		Entire Group(N = 63,984)				Treatment Group(N = 1,116)			
Variable	**Unit**	**Mean**	**Std. Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**
Age	Month	119.00	30.02	58	196	112.6	18.74	83	164
White	Binary	0.666	0.471	0	1	0.695	0.460	0	1
Black	Binary	0.208	0.406	0	1	0.148	0.355	0	1
Hispanic	Binary	0.096	0.294	0	1	0.123	0.329	0	1
Native	Binary	0.007	0.081	0	1	0.006	0.078	0	1
Asian	Binary	0.018	0.131	0	1	0.019	0.138	0	1
Male	Binary	0.507	0.499	0	1	0.542	0.498	0	1
Free	Binary	0.449	0.497	0	1	0.510	0.500	0	1
Reduced	Binary	0.099	0.298	0	1	0.118	0.323	0	1
Spanish	Binary	0.077	0.266	0	1	0.093	0.291	0	1
Urban	Binary	0.621	0.485	0	1	0.568	0.495	0	1
Lowaccess	Binary	0.314	0.463	0	1	0.218	0.413	0	1
Hispanic_prp	Proportion	0.070	0.120	0	0.787	0.078	0.122	0	0.445
Black_prp	Proportion	0.162	0.259	0	1	0.136	0.218	0	0.963
Native_prp	Proportion	0.007	0.016	0	0.174	0.011	0.024	0	0.087
Asian_prp	Proportion	0.011	0.027	0	0.312	0.006	0.022	0	0.219
Other_prp	Proportion	0.031	0.080	0	0.609	0.020	0.044	0	0.186
Publictrans_prp	Proportion	0.005	0.020	0	0.306	0.002	0.020	0	0.074
Singlemother_prp	Proportion	0.265	0.238	0	1	0.257	0.205	0	1
Highschool_prp	Proportion	0.350	0.110	0.028	1	0.368	0.093	0.078	0.654
Somecollege_prp	Proportion	0.277	0.089	0	0.748	0.265	0.080	0.062	0.484
Collegeplus_prp	Proportion	0.188	0.141	0	0.830	0.166	0.117	0	0.660
Asianandpacific_prp	Proportion	0.010	0.024	0	0.280	0.005	0.020	0	0.156
Incomebelowpoverty	Proportion	0.177	0.138	0	0.894	0.214	0.141	0	0.725
Workingmother_prp	Proportion	0.255	0.208	0	1	0.248	0.191	0	1
Novehicle_prp	Proportion	0.063	0.076	0	0.634	0.060	0.061	0	0.384
Vacant_prp	Proportion	0.122	0.095	0	0.660	0.111	0.077	0	0.451
Medincome	Dollars(in thousands)	43.6	18.81	6.85	250	37.27	13.60	11.62	81.84
Medhousevalue	Dollars(in thousands)	112.1	62.72	9.99	1000	86.13	35.85	28.7	206.5
Medyearbulid	Year	1980.4	12.4	1939	2005	1979	8.66	1948	2000
Bmi z-score		0.696	1.067	-3.983	3.918	0.761	1.076	-2.887	2.947
Bmi percentile		0.684	0.278	0	1	0.700	0.277	0.001	0.998

Table 3. The Comparison of Results among Different Matching Methods

Matching Method	Balance Test ($\mathcal{L}_1$ statistic)	Number of Observations in Control Group	Number of Observations in Treatment Group	Coefficient of DID (Fixed effects)		Coefficient of DID (Random effects)	
				Bmi z-score	*Bmi percentile*	*Bmi z-score*	*Bmi percentile*
None	0.998	N=62,868	N=1,116	0.150***	0.034***	0.148***	0.033***
Nearest-neighbor matching with 3 neighbors	0.992	N=3,045	N=1,116	0.122***	0.028***	0.125***	0.029***
Nearest-neighbor matching with 2 neighbors	0.990	N=2,068	N=1,116	0.118***	0.024***	0.124***	0.026***
Mahalanobis matching without caliper	0.907	N=754	N=1,116	0.129***	0.027***	0.152***	0.032***
Mahalanobis matching with caliper 0.1	0.616	N=200	N= 253	-0.055	-0.019	-0.045	-0.015
Mahalanobis matching with caliper 0.075	0.604	N=193	N= 244	-0.072	-0.026	-0.057	-0.019
Mahalanobis matching with caliper 0.05	0.533	N=175	N=218	-0.086	-0.031*	-0.073	-0.025
Coarsened exact matching	0.541	N=161	N=151	-0.150**	-0.0417**	-0.139**	-0.037**

Note: *, **, *** denote significance at the 10%, 5%, and 1% levels.

Table 4. Comparison of Descriptive Statistics of Control and Treatment Group after CEM

	Control Group (N=161)				Treatment Group (N=151)			
Variable	**Mean**	**Std. Dev.**	**Min**	**Max**	**Mean**	**Std. Dev.**	**Min**	**Max**
Age	99.79	24.1	58.5	165.4	116.2	20.52	83.7	155.9
White	0.921	0.268	0	1	0.933	0.249	0	1
Black	0.042	0.201	0	1	0.039	0.195	0	1
Hispanic	0.029	0.169	0	1	0.019	0.14	0	1
Native	0	0	0	0	0.006	0.081	0	1
Asian	0.004	0.064	0	1	0	0	0	0
Male	0.541	0.498	0	1	0.543	0.499	0	1
Free	0.323	0.468	0	1	0.344	0.476	0	1
Reduced	0.052	0.223	0	1	0.079	0.271	0	1
Spanish	0.019	0.136	0	1	0.019	0.140	0	1
Urban	0.422	0.494	0	1	0.364	0.482	0	1
Lowaccess	0.213	0.410	0	1	0.198	0.400	0	1
Hispanic_prp	0.055	0.084	0	0.55	0.054	0.082	0	0.439
Black_prp	0.038	0.110	0	0.676	0.040	0.127	0	0.963
Native_prp	0.005	0.013	0	0.113	0.005	0.013	0	0.087
Asian_prp	0.010	0.021	0	0.117	0.008	0.019	0	0.117
Other_prp	0.016	0.051	0	0.529	0.009	0.026	0	0.186
Publictrans_prp	0.001	0.004	0	0.035	0.001	0.007	0	0.07
Singlemother_prp	0.203	0.182	0	0.880	0.190	0.161	0	0.695
Highschool_prp	0.374	0.099	0.153	0.645	0.368	0.098	0.153	0.645
Somecollege_prp	0.270	0.077	0.062	0.499	0.266	0.074	0.062	0.484
Collegeplus_prp	0.180	0.130	0	0.520	0.195	0.127	0.013	0.520
Asianandpacific_prp	0.006	0.017	0	0.125	0.007	0.018	0	0.125
Incomebelowpoverty	0.169	0.12	0	0.568	0.171	0.126	0	0.725
Workingmother_prp	0.201	0.159	0	0.943	0.180	0.138	0	0.613
Novehicle_prp	0.052	0.048	0	0.248	0.053	0.053	0	0.384
Vacant_prp	0.099	0.077	0	0.372	0.100	0.076	0	0.383
Medincome	42.98	16.15	16.29	81.8	42.96	16.27	15.4	81.84
Medhousevalue	100.6	42.00	14.3	223.7	100.5	41.06	31.1	197.7
Medyearbulid	1981.4	9.690	1939	2003	1982.4	8.261	1948	2000
Bmi z-score	0.708	1.082	-3.173	3.331	0.688	1.160	-2.887	2.873
Bmi percentile	0.682	0.282	0.00075	0.999	0.675	0.297	0.0019	0.997