5. Regression Results for Evaluating the School Nutrition Programs

In this section, we present the regression results for evaluating the school nutrition programs. We first present results for the SBP program, and then we consider the sensitivity of our results. Finally, we turn to the implications of our SBP results for evaluating the NLSP program.

5.1. Evaluating SBP Availability

For our main analysis, we implement this difference-in-difference strategy in a regression framework. A regression allows us to take into account observable differences, such as age, gender, race, and income, between our difference-in-difference groups. To the extent that the identification strategy is contaminated, the regression can potentially adjust for remaining observable confounding factors. In addition, to the extent that we can control for other important determinants of the outcomes, the regression framework will improve the precision of our estimates relative to the difference-in-difference results.

We evaluate the impact of SBP availability with the regression

$$Outcome_i = \alpha + sbav_i\beta_1 + inschool_i\beta_2 + sbav_i * inschool_i\beta_3 + X_i\gamma + \varepsilon_i \quad (1)$$

where $sbav_i$ is an indicator variable for school breakfast being available, $inschool_i$ is an indicator variable for school being in session, and X_i is a vector of other important control variables. The coefficient on the interaction $sbav_i$ and $inschool_i$ (i.e., β_3) measures the causal impact of program, the regression analog of the difference-in-difference estimates presented in Table 3. The other control variables include age (indicators for each age), male, race (indicators for Hispanic, non-Hispanic black, and "other race"), income (indicators for \$5,000 increments and for greater than \$50,000), household size, and geography (a complete set of interactions between urban and the four census regions). For simplicity, we use ordinary least squares for all models, regardless of whether the dependent variable is continuous or dichotomous.[21]

For all the results in this section, the regressions account for the complex sample design of the NHANES. Specifically, we use information on the strata, primary sampling units, and weights provided by the NHANES for the regressions.[22] These methods implicitly account for the fact that our sample contains multiple children from some households. We examine the sensitivity of our results to accounting for the complex survey design in the appendix.

[21] Using ordinary least squares with a dichotomous dependent variable (a linear probability model) can lead to difficulties, especially when one is interested in computing predicted probabilities. Because are interests throughout are on marginal effects and for the sake of simplicity, we choose to ignore the dichotomous nature of our dependent variable. We have estimated logit models (results not presented here) and all of our substantive conclusions remained the same.

[22] The NHANES documentation suggests that such methods should be used. We implement these methods by using the "survey commands" in STATA, identifying the underlying selection probabilities, strata, and primary sampling units.

We present the results for the dietary recall measures in Table 4 and the exam measures in Table 5. The regression results largely mirror those presented in Table 3. The availability of SBP generally has positive impacts on nutrition-related outcomes, with statistically significant impacts on improving dietary quality (as measured by the HEI), the percent of total calories from fat, and rates of vitamin C, vitamin E, and folate deficiencies. For example, SBP availability increases the HEI by 3.89, an amount that represents a 6 percent increase of the population mean HEI, and reduces the prevalence of children being low of vitamin C by 7 percentage points, an amount that is twice as large as the population prevalence. Both of these results are statistically significant at the 0.01 level.

Impacts by Income Group

Equation 1 measures the mean impact across all income groups, but there are substantive reasons to expect the impact to vary across the groups. First, the subsidy the children receive varies across groups. Breakfasts are free for children from families with income less than 130 percent of the poverty line and are limited to cost less than $0.30 for children from families with income between 130 and 185 percent of the poverty line. Children from relatively high income families could still benefit because of a small subsidy to their meals. Second, there is a potential additional effect for all children who participate to the extent that school breakfasts are substituting for breakfasts that might have otherwise been consumed at home. Such a substitution effect will vary depending on the quality of the meal that would otherwise be consumed at home, which in turn likely depends on family income.

To examine how the impact of SBP varies by income, we define four different income groups based on the income-to-poverty ratio. These groups correspond to the programmatic rules regarding subsidies. We define the low income group as the children who are in households with an income-to-poverty ratio of less than 130 percent of the poverty line, the medium income group with an income-to-poverty ratio between 130 and 185 percent of the poverty line, and the higher income group with an income-to-poverty ratio greater than 185 percent of the poverty line. We also define an unknown group for those who did not respond to the income question.

Based on the four income groups, we estimate the model,

$$
\begin{aligned}
Outcome = {} & \beta_1 sbav * low + \beta_2 inschool * low + \beta_3 sbav * inschool * low \\
& + \beta_4 sbav * medium + \beta_5 inschool * medium + \beta_6 sbav * inschool * medium \\
& + \beta_7 sbav * high + \beta_8 inschool * high + \beta_9 sbav * inschool * high \\
& + \beta_{10} sbav * unknown + \beta_{11} inschool * unknown + \beta_{12} sbav * inschool * unknown \\
& + \beta_{13} X + u
\end{aligned}
\tag{2}
$$

The impact of SBP availability is measured by the interactions as before (that is, the coefficients β_3, β_6, β_9, and β_{12}). We note two things about this model. First, we include the same specification for the other regressors for this model as we did for equation 1. Second, we include interactions with a complete set of income dummy variables to facilitate comparison of these coefficients with those from equation 1—they are on the same scale.

Tables 6 and 7 present the equation 2 regression results. Recall that the significant overall impacts of the SBP were on dietary quality (measured by HEI score) and the percent of total calories attributed to fat (see Table 4). The results from Table 6 suggest that the impacts on dietary quality and the total calories attributed to fat are most consistently observed for the higher income group, not for the low income group. Similar patterns are apparent in Table 7 in that the overall significant impacts on vitamin C and folate are driven by the high and unknown income groups.

Taken at face value, these results are somewhat surprising. Recall that a school nutrition program is expected to affect the poor groups through an income effect (based on the subsidized meal) that is not supposed to be present for the non-poor. Although all income groups could potentially benefit from the meal substitution effect, we expected to find a larger meal substitution effect among the poorer groups. Quite simply, we would have thought that the poorer individuals would have had a lower quality diet at home.

One potential explanation for our results is that children from higher income families are better able to take advantage of the potential benefits of SBP. For example, it is possible that children across the income distribution are provided a relatively unhealthy breakfast. When SBP is available, the higher income parents may be more likely to have the flexibility to ensure their children are at the program and better able to monitor what the children eat. In addition, it is important to note that the “high income” group still contains households with quite modest income because the cut-off for the group is only 185 percent of the poverty line.

These results also allow us to observe what would happen if we were to use the higher income group as a differencing group, as Bhattacharya and Currie (2001) did. Such a strategy identifies the additional impact of SBP availability to the poorer groups over and above those impacts on higher income groups. In other words, such a strategy ignores any meal substitution impacts on higher income children. However, our results suggest that there exist effects among the higher income groups, and thus such an identification strategy would underestimate the impact on poorer children.

5.2. The Sensitivity of the SBP Availability Results

The results thus far are fairly striking. We have exploited a transparent identification strategy, and we find that the availability of SBP has a significant impact on several outcomes. In this subsection, we examine the sensitivity of our results.

Limiting the Variation in Income

A difference-in-difference strategy is dependent upon a linearity assumption. To the extent that the underlying impacts are linear or to the extent that the underlying impacts are non-linear and the underlying changes are small, our identification strategy will identify the true impact of SBP availability. Because we have not specified an underlying theory that tells us the functional form of the various relationships, we must be cautious whenever large underlying differences exist.

As is apparent in Table 3, there are large differences in income between the schools where SBP is available and where it is not available. Although we are controlling for income fairly flexibly (indicator variables for $5,000 bands up to $50,000, as well as an indicator for income over $50,000 and income unknown), it is possible that these controls are not sufficient to make the underlying groups comparable. We believe that income is an important determinant of dietary outcomes, and indeed our results in the previous section suggest that the impacts of SBP vary by income group. To ensure that these results are not driven by underlying non-linearities, we present some results where we restrict the variation in income to provide evidence about whether such concerns are driving our conclusions.

In Table 8, we repeat the key regressions in Tables 4 through 7, but drop individuals with more than $40,000 in annual household income from the sample.[23] The results look similar to the previous results. There are still significant impacts on several outcomes (HEI score, percent of calories from fat, low vitamin C levels, and low vitamin E levels), and the results are still generally driven by impacts among children from medium and relatively high income families.

Seasonal Variation in the NHANES

One of the key aspects of our identification strategy is that, among the individuals who have SBP available, some receive breakfast and some do not due to school being in session. As we discuss in Section 4, one of the major confounding factors to this strategy is that schools systematically tend to be out of session during the summer months and dietary outcomes could be related to season. For example, fresh fruits or vegetables may be cheaper during the summer (either because of increased availability in stores or because of the opportunity to have gardens) or exercise may be easier. Our strategy aims to overcome seasonal confounders by using those individuals without school breakfast available to identify the true, underlying seasonal variation in diet. However, to the extent that such seasonal variation is large, the differencing strategy may not fully account for the differences.[24]

One aspect of seasonal variation that we can examine directly is food prices. Figure 2 plots the percent change in the Food component of the CPI-U (not seasonally adjusted) between adjacent months for the period January 1993 through December 2002. The heavy black line is the mean change in the index. There is some systematic variation in food prices across the year. For example, there is a large decline in the mean change between January and February, as well as little overlap in the distribution of changes between January and February. However, there does not appear to be too much systematic variation by season. Figure 3 presents the same information for monthly changes in the Fresh Fruit and Vegetables component of the CPI-U. This figure suggests more seasonal variation in prices: three of the four low growth months are during the summer (June, July, and August). However, we do not interpret these differences to be so large as to make the differencing strategy infeasible.

---

[23] The NHANES collects information on income with one question, asking about total family income, that places families into brackets. Thus, we are forced to impose the income restriction in nominal terms across the six survey years rather than real terms.

[24] Importantly, a difference-in-difference strategy is also dependent upon a linearity assumption. To the extent that the underlying impacts are linear or to the extent that the underlying impacts are non-linear and the underlying changes are small, our identification strategy will identify the true effect.

A potentially more problematic source of seasonal variation exists in the NHANES due to its survey design. Specifically, to accomplish its vast undertaking of data collection, the NHANES survey relies on fully equipped medical clinics, a Mobile Examination Center (MEC), that are housed in the back of tractor trailers.[25] An MEC is then transported to each of the data collection sites. Thus, data collection is limited by the number and transportation time of the MECs. Because of this constraint, the NHANES takes far longer to collect a nationally representative sample then most other surveys. For example, the NHANES III was in the field for six years, with the first three years producing a nationally representative sub-sample and the last three years producing a nationally representative sub-sample.

Due to the actual data collection schedule of the NHANES, a further seasonality issue is introduced into the data. Table 9 presents a cross-tabulation of the number of children interviewed by season and census region. The interviewing pattern implies a strong correlation between season and geography. In fact, almost no interviews took place in the South census region (plus Texas) during the summer. To the extent that diets differ across regions, then the NHANES data collection process introduces an additional confounding factor into our analysis.

The potential difficulties of the interviewing schedule can be observed in Table 3. To the extent that the same types of places were visited over the calendar year, then the demographic characteristics should not vary by school being in session. However, individuals are much more likely to be Hispanic when school is in session, regardless of whether or not SBP is available. Although our difference-in-difference identification strategy could net out these underlying differences between the groups as well, such large differences are potentially problematic.

In Table 10, we repeat the regressions from Table 8, except we exclude children born in the South and West. Although the precision of the estimates is much less, presumably because of the smaller sample sizes, many of the point estimates are quite similar. For example, we still find overall impacts on low vitamin C and low vitamin E, and the results are still largely driven by individuals in the higher income groups.

5.3. Evaluating the National School Lunch Program

The difference-in-difference identification strategy that we have used to examine the impact of the SBP relied on two premises: school is not in session year around and the SBP is not available in many places. The latter premise is not true for the NSLP. As can be observed in Table 2, NSLP is available for over ninety percent of the children across all four income groups.

In our proposal to the USDA, we acknowledged a different identification strategy would be necessary because of NSLP's widespread availability and discussed several alternatives. The most promising alternative was to use higher income children as the differencing group to identify the underlying seasonal variation in diet. However, this strategy relies on the assumption that the NSLP program has no effect on the higher income children. Taken at face value, our results thus far invalidate this assumption in that we observe an impact of the SBP on

[25] For more information about the MEC, see the special section on the NHANES website: http://www.cdc.gov/nchs/about/major/nhanes/mectour.htm.

the children of higher income families, and in fact, these are the children for whom we primarily observe an impact. We have little reason to speculate that the NSLP would be any different.

Regardless, it is instructive to examine results for the NSLP. Table 11 presents the basic difference-in-difference results (analogous to Table 3) for the impact of NSLP availability in which we use the identification strategy originally proposed. Specifically, the first two columns compare the outcomes for children who have free NSLP available, distinguishing between those children who are in school and those children who are not in school. This comparison ideally would allow us to hold everything else constant about the children (such as socio-economic status and community characteristics), except some children currently receive NSLP because school is in session and some do not because school is not in session.[26] For example, the results in Table 11 would suggest that the overall dietary quality is better for poor children when they are in school (an HEI score of 63.3) than when they are out of school (an HEI score of 62.7). However, there could also be important seasonal variation in dietary quality that could confound this comparison. If we were to maintain that there were no impacts of NSLP on the children from relatively high income families, then the seasonal variation would suggest the dietary quality is higher when school is out (an HEI score of 64.5) then when school is in session (an HEI score of 63.6). The difference-in-difference estimator then implies that free lunches through the NSLP has an impact of 1.5 [= (63.3 – 62.7) – (63.6 – 64.5)] on the HEI score.

Overall, the results in Table 11 are weak and mixed.[27] The availability of free lunches appears to improve dietary outcomes for some measures (higher HEI score, lower calories, less prevalence of low zinc intake and low vitamin C level), but it appears to harm dietary outcomes for many other measures (increased calories from fat and saturated fat and increased prevalence of low vitamin A, E, folate, anemia, and high cholesterol). However, only one of these impacts is statistically significant (increased prevalence of low vitamin E levels), but this relationship is weak and implausible in that vitamin E is unlikely to change over a short time period because it is not water soluble.

Our conclusion from these results is not that the NSLP has little impact, though. Rather, we conclude that our identification strategy does not provide a good mechanism to identify the causal effects of the NSLP. Again, our identification strategy hinges on the assumption that school nutrition programs do not have an impact on the relatively high income groups. Our previous results suggest that this not the case for the SBP, and we have little reason to expect the NSLP to be different.

---

[26] The regional problem previously identified is readily apparent. For example, 28.0 percent of the children who are attending school report being Hispanic, whereas 10.2 percent of the children not in school report being Hispanic. These differences are what would be expected given data are collected in the south and southwest during the winter and the northeast during the summer.

[27] We have also implemented this difference-in-difference strategy with a regression framework, and all of the conclusions remain the same.