III. Alternative Designs

This chapter outlines four possible designs for obtaining rigorous estimates of the impact of participation in the School Breakfast Program (SBP) on key student outcomes related to learning:

1. The planned experimental evaluation by the U.S. Department of Agriculture (USDA) of the Universal-Free School Breakfast Program pilot projects (USBP design)

2. An experimental study of the effects of classroom implementation of the SBP among breakfast program applicant schools (SBP applicant design)

3. A nonexperimental study of the effects of the SBP based on data from the Early Childhood Longitudinal Study, Kindergarten Cohort (ECLS-K design)

4. A nonexperimental study of the effects of the SBP based on data from the forthcoming National Health and Nutrition Examination Survey (NHANES design)

First, the details of the methodological approach used by each of the four designs are described. Second, the strengths and weaknesses of the designs are assessed and summarized in terms of feasibility and methodological rigor. Third, a single design is selected as the most feasible of the four alternatives.

In Section A, we describe the key elements of the alternative designs. The subsequent four sections (Sections B through E) outline the four alternative design approaches by providing details about the key design elements of each one. Section F summarizes the strengths and weaknesses of each alternative and makes a case for the design we selected as the best of the four. (We describe the selected design in greater detail in the final chapter of this report.)

As described in Section F, we conclude that the ECLS-K design, supplemented with analysis based on the 1988-1994 NHANES III dataset, is the most feasible. We arrived at this conclusion after taking into consideration the fact that the USBP design has been planned and is being implemented. The strengths of the ECLS-K design include its large, nationally representative sample and longitudinal design with carefully developed measures of academic achievement. Although the ECLS-K provides no information on outcomes related to dietary intake or nutritional status, the supplemental analysis conducted with NHANES III data should help address this limitation.

A. Key Elements of the Alternative Designs

As we have discussed in Chapter II, the limitations of the existing studies of the effects of the SBP on learning reduce the usefulness of their results. It is therefore important that the alternative designs considered here address at least some of these limitations. In particular, the designs should do as many of the following as possible:

- Examine multiple outcomes
- Examine the U.S. School Breakfast Program

- Collect and/or use data from a nationally representative sample
- Use an experimental design
- Use large samples of students and schools

There are many tradeoffs both among these desired study features and between the features and study costs. For example, a study using an experimental design would be difficult to implement nationally, thereby making it difficult to use a nationally representative sample. Thus, it would not be possible to collect data on a nationally representative set of USBP participants and nonparticipants, because the USBP is not available nationally. There also are obvious tradeoffs between the size of the samples analyzed, the number of outcomes measured, and the cost of the study. Although it may be feasible to collect data on a limited number of easily measured outcomes from a large sample, it may not be feasible to collect data from a large sample on a wide range of outcomes that include difficult-to-measure ones like dietary intake and short-term cognition.

To define each of the alternative designs fully, we must describe design elements that address such issues as the research objective, the general approach to be used, and the details of how this approach will be implemented. The importance of the following elements will be considered in the discussion of the alternative designs:

1. ***Intervention and counterfactual.*** It is important to clearly define the *intervention* being examined as well as to explicitly state the *counterfactual*. Together, the intervention and counterfactual define the hypothesis that the study will test.

2. ***Basic design approach.*** It is necessary to determine whether the study will use an experimental or nonexperimental design. In an experimental design, the choice of the unit of random assignment is important. In a nonexperimental design, it is important to identify differences between the two groups other than the main outcome in order to limit selection bias. It is also necessary to select the population of interest and a corresponding sampling frame as well as to determine whether the study will use a cross-sectional or longitudinal design. [7]

3. ***Data collection.*** The design should specify whether the study is to be based on primary data or secondary data (or some combination of the two). Study designs involving primary data collection require many more decisions than do those using secondary data.

4. ***Measurement of key characteristics and outcomes.*** To measure SBP participation, the design first must specify whether the intention is to capture SBP participation on a single day or over a longer period. To measure students' learning, the design must specify what type of test will be used, and how the test will be implemented.

[7] The sample design should also specify planned sample sizes and levels of statistical precision. These issues are discussed in Sections B.6, C.6, D.6, and E.6.

5. ***Analysis plans.*** The major analysis issue to be considered by experimental designs is whether substantial *dilution* of the effect of SBP participation on learning is inherent in the design. The major analysis issue that must be confronted by nonexperimental design alternatives is selection bias.

6. ***Statistical power.*** When evaluating a design, it is highly important to determine the degree to which the design has the statistical power to detect effects of a size likely to be produced by the intervention. The statistical power of a design approach typically is specified in terms of its minimum detectable difference (MDD), the smallest *true* intervention effect that would likely lead the evaluation to conclude statistically that the intervention had an effect.

7. ***Design costs***. The cost of carrying out a design is an important factor in assessing its promise. A design's cost can be influenced by each of the six design features described in this list. The cost also may include expenses related to establishing and operating a demonstration program (for designs based on a demonstration).

B. The USBP Design

The William F. Goodling Child Nutrition Act of 1998 (P.L. 105-336) authorized the Secretary of Agriculture, through the Food and Nutrition Service (FNS) of the USDA, to conduct a demonstration and evaluation that will rigorously assess the effects of the USBP on program participation and on a broad range of student outcomes, including academic achievement. Six school food authorities (SFAs) were selected to participate in the demonstration in spring 2000. FNS contracted with Mathematica Policy Research, Inc. (MPR) to develop a comprehensive study design for evaluating the USBP pilot programs. This section summarizes the design plan that MPR produced (Ponza et al. 1999). The plan proposed an experimental approach to randomly assign elementary schools within the six participating SFA's into a treatment group that will use the USBP and a control group that will use the regular SBP.[8]

1. Intervention and counterfactual

The primary intervention in the USBP design is having access to the USBP; the counterfactual is having access only to the regular SBP. In other words, students in the treatment group attend elementary schools over the course of a school year that offer the USBP, whereas control group students attend elementary schools that offer only the regular SBP. In practice, this intervention has two main implications for students. First, all treatment group students can obtain a free school breakfast, but only low-income control group students who have been certified for free meals can do so. Second, because breakfasts are served free to all students in USBP schools, cafeteria workers or other school personnel do not have to distinguish between certified and noncertified students, thus potentially leading to less stigma among SBP participants. Both implications are expected to lead to higher rates of participation in the USBP by treatment group students than in the regular SBP by control group students.

[8]FNS awarded the contract to conduct the USBP evaluation to Abt Associates Inc. The design as implemented may differ somewhat from the design described in this section.

In addition to these implications of attending a USBP school, the implementation of the USBP itself may lead to other changes in the breakfast program available to students at treatment group schools. For example, because the implementation of the USBP is expected to lead to greater program participation, schools may decide to alter the ways in which breakfasts are prepared or served. For instance, schools with USBPs sometimes have opted to serve breakfasts in classrooms as part of the school day, rather than in the cafeteria before the start of classes. A difference in breakfast delivery could influence student outcomes.

It is important to emphasize that the intervention in the USBP design is not participation in the USBP relative to a counterfactual of nonparticipation. Both treatment group students at USBP schools and control group students at regular SBP schools may or may not eat a school breakfast. Thus, the estimated impact arising from differences between treatment group and control group members cannot be used directly to determine the effects of breakfast program participation. Instead, the treatment-control difference simply provides an estimate of the impact of *attending* a school in which the USBP rather than the regular SBP is available. However, attending a USBP school rather than a regular SBP school (that is, being in the treatment group rather than in the control group) is expected to be positively correlated with eating a school breakfast. It is therefore possible to use treatment-control differences to obtain indirect estimates of the effects of participation.

For a direct estimate of the effects of breakfast program participation on learning, a secondary analysis will treat participation as the intervention. Within USBP schools, for example, this secondary intervention will be defined as usually eating a school breakfast, relative to the counterfactual of usually being a breakfast program nonparticipant. Breakfast program nonparticipants may or may not have eaten any breakfast, and these two possibilities will be distinguished. As described in the following section, the basic design approach for examining the effects of the secondary intervention differs from the experimental design used to examine the primary intervention—the effects of USBP availability.

2. Basic design approach

The USBP design uses an experimental approach to examine the effects of USBP availability on student outcomes. The unit of random assignment for the evaluation is the elementary school. Some or all of the elementary schools in each of the six participating SFA's are matched on the basis of their characteristics. One school in each matched pair is then randomly assigned to a treatment group that participates in the USBP or a control group that continues using the regular SBP. After the USBP is implemented in the treatment schools, samples of students are drawn from each school.

a. Random assignment

In this design, random assignment is conducted at the school level rather than the student level, primarily because the USBP typically is implemented at the school level (or at least at the classroom level). Implementing the USBP at the student level would make it possible to randomly assign students, but it would also result in the loss of two of the benefits of the USBP: (1) reduction of administrative burden on schools, and (2) reduction of stigma borne by students. In addition, attempting to randomly assign students into USBP and regular SBP groups would raise practical concerns that would make successful implementation unlikely.

One of the drawbacks of randomly assigning schools rather than students is that random differences between schools and their experiences over time will make statistical detection of program effects more difficult. This loss of precision arises because sample members in a particular group are influenced by shared events or circumstances. For example, if one treatment school has a particularly strong teaching staff, the student outcomes among all students in that school would be positively influenced relative to students in control schools. If students, rather than schools, were randomly assigned, some treatment group and some control group students would be affected by the teaching staff of that school.

To address the loss of precision from using schools as the unit of analysis, the USBP design suggests using as large a sample of schools as the funding constraints of the evaluation will allow. Random circumstances or events in treatment and control group schools will be more likely to cancel each other out if there are more schools in the sample. In addition, the design attempts to minimize random variation across treatment and control schools by matching schools prior to random assignment.

In particular, schools that have similar observable characteristics (for example, size, racial composition, average test scores, or income levels), and that local administrators consider to be similar in quality, will be matched. Each pair of matched schools is then randomly assigned, with one school becoming a treatment school and the other a control school.

The USBP design report discusses two alternative approaches that vary according to the number of schools sampled (Ponza et al., 1999). The preferred approach proposes that 144 schools be randomly assigned, 72 to a treatment group and 72 to a control group. The schools would need to be selected from the six SFAs participating in the demonstration.

b. Student sampling

The USBP design calls for two partially overlapping samples: (1) a cross-sectional sample of students in grades 1 through 6 as of the 2000-2001 school year; and (2) a longitudinal cohort of students in study schools, used to gather data on changes in student outcomes as measured by school records. The students in the cross-sectional sample are to be surveyed once, during spring 2001. That survey also will collect dietary recall data and information about the students' experiences in school and their attitudes toward school; the students also will be given cognition tests. In addition, their parents will be surveyed in spring 2001.

The second sample (the longitudinal cohort of students in study schools) includes students in grades 2 through 6 in school year 2000-2001. Most of these students will have been in grades 1 through 5 during the previous school year (the baseline year) and could be followed through the 3 years of the demonstration period; they will not be followed beyond grade 6, however. Clearly, some of the students (those in grade 1) in the cross-sectional sample will not be included in the longitudinal sample. In addition, for operational reasons, there will be longitudinal sample members who will not be in the cross-sectional sample. Otherwise, the two samples will overlap.

Administrative school records data will be the main form of information collected from students in the longitudinal sample. Ultimately, the aim is to collect school records data for each student from the baseline year and the 3 followup years. Of particular importance will be test scores. Longitudinal data on students' scores provide the best way to measure their achievement in a given school year. Test scores from a single point in time measure their level of achievement at that time, encompassing their progress over an entire lifetime. Longitudinal data enable researchers to construct variables indicating students' *gain* in test scores, which should be influenced primarily by experiences during that school year, including SBP participation, and by their natural aptitudes.

The USBP design also allows for collection of school-level data over the baseline year and the 3 followup years. This effort focuses on collecting data that measure such outcomes as attendance rates and mean test scores. Because the data cover both the preimplementation and the postimplementation periods, analysis can focus on how the change in mean outcomes between the baseline year and a particular followup year differ between treatment (USBP) schools and control (regular SBP) schools.

An important issue in the USBP design is the potential for dilution of the effects of the intervention. Dilution occurs when the intervention (relative to the counterfactual) influences only a small proportion of the target population. In the case of the USBP evaluation, the main expected effect of the availability of the USBP, relative to the availability of the regular SBP, is limited to students who would not have eaten a school breakfast under the regular SBP but who do so under the USBP.

These "new participants" are the only group whose behavior changes as a result of the availability of the USBP.[9] However, all students attending treatment and control schools are to be sampled, not just new participants. Thus, the measured size of the effect of the intervention (that is, the difference in mean outcomes between treatment and control students) depends on the effect of breakfast program participation on new participants and on the percentage of the sample consisting of new participants.

A design option attempts to deal with the possible effect of dilution by maximizing the proportion of the sample made up of new participants. This approach involves conducting a "preimplementation" survey during the summer before the USBP is implemented in the control schools. Students in both treatment schools and control schools would be asked whether they had usually participated in the SBP during the previous school year and whether they would participate in the program if breakfasts were free to all students of the school.

Those who report that they did not participate during the previous year but would do so if breakfasts were free to all students are defined as "likely new participants." A key feature of this design option is that, at the time of the survey, students do not know whether their school has

[9] "New participants" refers to students in both treatment schools and control schools, even though students in control schools do not have the opportunity to participate in the USBP. For that group, being a new participant implies that they do not eat a school breakfast at their regular SBP school but would eat a school breakfast if they were attending a USBP school.

been selected as a treatment (USBP) school; thus, the question about future participation is purely hypothetical for both groups of students. After the likely new participants are identified, they can be oversampled, thereby lessening the problem of dilution.

3. Data collection

The USBP design calls for a wide range of data collection activities, including surveys of students, their parents, and their teachers; the collection of school administrative data, including achievement test data; and the administration of cognitive tests. This variety of data sources will ensure that the evaluation has information on a large number of student outcomes; however, it will require an extensive data collection effort. The key aspects of this effort are described here and are organized according to the source of the data.

a. Data collection from students and their parents

Much of the analysis to be conducted by the USBP evaluation will be based on data collected from students and their parents. These groups will provide data in three forms: (1) 24-hour dietary recalls, (2) student and parent surveys, and (3) cognitive tests.

Dietary intake. The USBP design will use a 24-hour dietary recall instrument to collect dietary-intake information from the students. The 24-hour dietary recall methodology provides information on all the foods eaten during a 24-hour period and will be conducted at the end of the first followup year. The design proposes conducting one 24-hour dietary recall for all sample members (at the end of the first followup period) and a second recall for a second representative subsample to allow estimation of the distribution of usual intake. Ponza et al. (1999) recommended using computerized interviewing methods and using the protocol from the most recent Continuing Survey of Food Intakes by Individuals to elicit complete information on food intake and combat the problem of missing data.

In practice, the design calls for an in-person dietary recall interview to be conducted with the child sample member alone, immediately after the breakfast period at school. The child's intake during the rest of the day would be collected during an interview with the child and his or her parent on the following day. The presence of the parent at the second interview will help the child remember the foods consumed at home during the intake day.

Student and parent surveys. The USBP design suggests conducting in-person interviews with children and their parents to obtain demographic and socioeconomic information and to fully assess breakfast program participation. These interviews would be conducted in conjunction with the dietary recalls, with students interviewed in school on the intake day and parents interviewed after the completion of the 24-hour recall. In the basic USBP design, student and parent interviews are scheduled for the end of the first followup year.

A design option is for the evaluation to include a preimplementation survey of sample members' parents. This survey would be conducted in the summer before the first followup year (and prior to the implementation of the USBP in the treatment schools). The main pieces of information collected would be sample members' participation status during the previous school year and their likelihood of participating in the following school year, if breakfast were free to all. If implemented, the preimplementation survey would last only 10 to 15 minutes and would be administered to a subsample of approximately 150 parents in each school.

Cognitive tests. The design includes administering a short-term test of cognition—the Wechsler Memory Scale. This scale measures students' short-term memory, has performed well in research settings, and is straightforward to administer. It typically takes about 10 to 15 minutes and would be given after breakfast and before lunch on the day that students' dietary intakes are measured.

The USBP design also proposes a second short-term test—the Revised Children's Manifest Anxiety Scale. This test would measure the children's emotional state. It would take 10 to 15 minutes and would be administered at about the same time as the Wechsler Memory Scale.

b. School records data

Schools' administrative records data could provide a wide range of student- and school-level information useful for the USBP evaluation, including rates of breakfast program participation, measures of student behavior, achievement test scores, visits to the school nurse, and attendance and tardiness. The USBP design calls for schools to complete an administrative data form for the school as a whole and for individual students in the sample. This method would ensure that, to the extent possible, data from different districts would be in comparable form.

It is possible, however, that this method would impose a burden on participating schools to gather this information accurately and promptly. As a backup method, schools could be asked to provide electronic data files containing the requested information, which would then be manipulated by the evaluator to bring it into a reasonable form.

The collection of school- and student-level administrative data would begin during the baseline year and would continue throughout the 3-year demonstration period. At the beginning of the evaluation period, the evaluator would seek parental consent to obtain sensitive information on individual sample members and would work with schools to provide both this information and school-level information.

c. Data collection from teachers

The USBP design includes surveys of the sampled students' teachers to collect additional information about the students and about the general school climate. The survey would also collect information about the teachers, their impressions of the school and its students, and their comments about individual sample members. In particular, teachers would be asked about each student's attendance and tardiness, classroom behavior, health, and academic performance. Finally, the teachers would be asked a series of questions as part of the Connors Teacher Rating Scale (CTRS) to assess sample members' classroom behavior and attention level in class.

To collect data about individual students, a student-teacher crosswalk must be developed as part of the initial sampling frame. The teachers would then be sent a general questionnaire and a set of individual questionnaires, one for each sample member in their class. This survey would be conducted at the end of the first followup year.

d. Other forms of data collection

The USBP design also proposes three additional forms of data collection: (1) direct observation of sample members' participation in the SBP/USBP, (2) data collected from SFA and school personnel on the characteristics of demonstration schools and the schools' food service operations, and (3) qualitative data collection based on site visits.

Direct observation of breakfast program participation. Because accurate measurement of sample members' SBP/USBP participation is critical to the USBP design, the design includes an option for observing this outcome directly. Field interviewers or school staff would observe students as they passed through the cafeteria line or ate in the classroom at breakfast, noting when students in the study sample selected a school breakfast. This observation would take place during the 5-day school week in which the 24-hour dietary recall interviews were conducted. School meals for sample students would be observed at the end of the first and second followup years.

Data collection from SFA and school personnel. The USBP design includes surveys of the SFA directors of the six districts and of school administrators and cafeteria managers at the treatment and control schools. These surveys would provide information on SFA operations and school outcomes, such as attendance and SBP/USBP participation rates. They would be conducted at the end of the first followup year.

Qualitative data collection. The USBP design also calls for the collection of qualitative data covering issues related to USBP implementation. In particular, the evaluator's staff would conduct site visits to demonstration schools to collect information on such issues as:

- How schools have implemented the USBP
- How much cross-school variation in program implementation exists, and why this variation arises
- What strategies schools have developed to deliver breakfast program services
- How the USBP and regular SBP differ in their attempts to promote nutrition, learning, and other key outcomes
- How the program costs of the USBP and regular SBP differ
- What the characteristics are of the meals offered by USBP and regular SBP schools

The qualitative data collection effort would include a meals-offered survey; examination of program documents and records; focus groups with school staff and students; and semistructured interviews with SFA administrators, cafeteria managers, and school administrators. This effort would be conducted through a combination of telephone interviews and site visits.

4. Measurement of key characteristics and outcomes

Any design of the relationship between the SBP and learning must be able to accurately measure relevant aspects of students' participation in the breakfast program, outcomes related to their learning, and other relevant characteristics. Because the approach taken toward measurement of these factors influences the details of both data collection and data analysis, the design should carefully consider these measurement issues.

a. School breakfast participation

At a conceptual level, students should be defined as SBP participants if they select a set of foods from the school that qualifies as a USDA-reimbursable breakfast. Because most schools do not keep detailed student-level records on participation based on this definition, the USBP design relies on three approaches to proxy for it.

The first and most direct approach for determining students' participation status is to ask them (and/or their parents) directly whether they ate a school breakfast on a given day or during a given period. The drawback of this approach is that parents may not know, and their children may claim to have eaten a school breakfast to satisfy their parents even if they did not actually do so.

A second approach defines participation based on the foods the students obtained for breakfast from school. For example, the School Nutrition Dietary Assessment (SNDA-1) study defined school breakfast participants as students who obtained at least two food items from their school that contributed to the USDA breakfast pattern requirement. The third approach involves direct observation of breakfast program participation, as described in the previous section.

Participation can be defined either in terms of the target day (that is, the day on which the students' dietary intake information is collected) or over a longer period, such as a week, month, or school year. The USBP design recommends collecting both types of participation measures, because each type is appropriate for different outcomes. For example, one would expect short-term cognitive functioning to be most strongly influenced by breakfast program participation on the day that cognitive functioning is measured, rather than by usual participation. Conversely, academic achievement would more likely be influenced by usual participation, rather than by whether a student ate a school breakfast on any given day.

b. Outcome measures related to student learning

Student learning can be measured directly, through multiple years of achievement test score data, or indirectly, through various outcomes that may in turn influence academic achievement. These direct and indirect outcomes are described here.

Academic achievement. The most direct means of measuring students' learning in school is to measure their academic achievement. In particular, the gain in their level of academic achievement from one year to the next can serve as a measure of the amount that students learned during a school year. The best current measures of students' academic achievement are their scores on standardized tests. If the administrative records of participating schools provide these test scores (or if these tests can be administered by the evaluator), the USBP design recommends using gains in test scores as the primary measure of student learning.

The most appropriate type of test for measuring student learning is a norm-reference test (NRT), which is designed so that scores can be compared with the scores of a reference group of students—typically, students across the country as a whole. This type of norming allows for comparisons of test scores across students taking different standardized tests, provided that the comparisons are made carefully. Examples of NRT's include the Iowa Test of Basic Skills (ITBS) and the Comprehensive Test of Basic Skills (CTBS). Ideally, these tests would be administered in the spring, to correspond with the evaluation's other data collection efforts.
In comparing students' test scores over time (to generate a measure of the gain in scores), across grades, or across schools, the USBP design emphasizes certain types of measures. For example, in addition to using NRT scores, scores can be further standardized by converting percentile scores to normal-curve-equivalent (NCE) scores, which have better mathematical properties. If data from different schools contain information on student scores on different standardized tests, the norming populations for the different tests should be as similar as possible. In addition, the subject matter of the tests should be similar, with reading test scores being compared with reading test scores rather than with math test scores.

Cognitive functioning. There are many aspects of cognition, as well as different measures of any one aspect. The USBP design recommends the Wechsler Memory Scale, based on a review of the literature on the effects of breakfast eating on cognition (Pollitt 1995; Vaisman et al., 1996). This literature indicated that tests of verbal memory are sensitive to breakfast consumption. Although breakfast eating may also may influence other aspects of cognition, such as visual perception, verbal fluency, and time on task, the limited resources of the evaluation and the limited time over which information can be collected from students indicate use of the Wechsler Memory Scale.

Emotional functioning. A number of instruments are available to measure various aspects of children's emotional status. Given time constraints and expectations about the effects of breakfast program participation, the USBP design calls for using the relatively short and easy-to-administer Revised Children's Manifest Anxiety Scale (Reynolds and Richman, 1985).

Attendance and tardiness. Obtaining school-level attendance data should be easy, as schools are required to report some measure of average daily attendance to district and State education authorities. When using school-level attendance data, however, it will be necessary to ensure that comparisons of attendance rates between schools are made only if the measures of attendance are similarly constructed, with similar definitions. Obtaining parental consent is a key issue with respect to collecting student-level administrative data on attendance (and on other outcomes).

Information on attendance and tardiness also would be obtained from the student and parent surveys. These self-reported data may be less reliable than the records data, but it would be possible to collect a wider range of information during these surveys.

Classroom behavior and disciplinary incidents. The USBP design calls for measuring student behavior in three different ways. First, the teacher surveys would include a variety of questions related to sample members' behavior, including the CTRS. In particular, the teacher survey would include the subscale of the CTRS that measures student hyperactivity. Second, the student and parent surveys would include questions on behavior. For example, sample members would be asked (possibly by proxy) whether they have engaged in particular behaviors, such as fighting or talking back to teachers, or whether they have been disciplined (formally or informally) during the past school year. Third, school records may contain some information on students' disciplinary incidents.

5. Analysis plans

The USBP design proposes two major types of impact analysis to examine the relationship between the breakfast program and learning. The first involves estimation of the effects of USBP availability on student outcomes—*the availability analysis*. The second involves estimation of the effects of SBP/USBP participation on student outcomes—*the participation analysis*. The availability analysis is based on an experimental design, whereas the participation analysis uses a nonexperimental design. The primary student outcomes to be examined, as called for by the Child Nutrition Act of 1998, are breakfast program participation, academic achievement, attendance and tardiness, and dietary intake over the course of a day. Finally, either of the two types of analysis can be estimated using student-level data or school-level data.

Two types of analysis, each of which can be estimated using school-level data or student-level data, result in four types of models overall. The four types are described here.

a. School-level USBP availability model

To estimate the effects of USBP availability on student outcomes using school-level data, Ponza et al. (1999) propose the following model:

(1) $Y_j = Z_j \delta_1 + \alpha_1 USBP_j + \varepsilon_{1j}$,

where:

Y_j	=	mean outcome among students at school *j*
Z_j	=	vector of characteristics of school *j*
$USBP_j$	=	binary variable representing USBP status of school *j*.

In the model, an outcome, such as a school's mean test score, is regressed on an indicator of whether the school is a USBP school (treatment school) or a regular SBP school (control school) and a set of school characteristics that potentially influence the mean test score.[10] The estimate of the parameter α_1 represents the estimated effect of USBP availability on academic achievement. The experimental design, if properly implemented, ensures that this estimate is unbiased. The model

[10] In this case, the control variables would include the lagged value of the dependent variable (that is, the mean test score at the school in the previous year). By controlling for the previous mean score, the estimated effect of USBP availability would be its effect on the gain in the school's mean test score, rather than its estimated effect on the level of the mean score.

would be estimated using ordinary least squares (OLS), logit/probit, or tobit estimation techniques, depending on whether the dependent variable is continuous, binary, or truncated.[11]

b. Student-level USBP availability model

The USBP design calls for the estimation of a hierarchical linear model (HLM) of the effects of USBP availability on student outcomes using student-level data. A simplified version of this model is shown here:[12]

(2) $Y_{ij} = X_{ij}\beta_2 + Z_j\delta_2 + \alpha_2 USBP_j + \gamma_{2j} + \varepsilon_{2ij}$,

where:

Y_{ij}	=	value of outcome among student *i* at school *j*
X_{ij}	=	vector of characteristics of student *i*.

The other variables in the model are defined as before. This model explicitly addresses the potential nonindependence of the error term across observations by giving each observation a school-level and a student-level component of the error term. The key estimate from the model is the estimate of the parameter α_2, which represents the average effect of USBP availability on the outcome. Again, random assignment ensures that this estimate is unbiased. If the outcome measure is participation in the breakfast program, α_2 represents the direct effect of USBP availability. For other outcome measures such as dietary intake or test scores, however, α_2 represents primarily an indirect effect. The main way in which USBP availability is expected to influence key student outcomes is by first influencing program participation, which, in turn, influences the outcome of interest.

c. School-level participation model

If the USBP data collection effort yields multiple years of information on school-level participation rates, school characteristics, and outcome measures, the following school-level participation model may be estimated:

(3) $Y_{jt} = Z_{jt}\delta_3 + \alpha_3 P_{jt} + \gamma_{3j} + \varepsilon_{3jt,}$,

where:

Y_{jt}	=	mean outcome among students at school *j* in year *t*
Z_{jt}	=	vector of characteristics of school *j* in year *t*
P_{jt}	=	participation rate among students at school *j* in year *t*
j	=	fixed effect of school *j*.

The key parameter in the model is α_3, which represents the influence of the participation rate on

[11]Furthermore, the sample of schools is unlikely to be a simple random sample. To account for any possible design effects, the standard errors of the coefficients would be estimated using Taylor series approximation methods with a software package such as SUDAAN.

[12]It is possible to increase the complexity of the model by modeling the coefficients on the school-level variables in the model as dependent on other school-level variables and a random error.

the school-level outcomes of interest. The model implicitly assumes that the effect of the participation rate is the same in USBP and regular SBP schools, although this assumption could be relaxed by interacting the participation rate variable with the binary variable indicating USBP availability. For the model to be estimated with a reasonable degree of precision, the participation rate variable must have sufficient exogenous variation over time and across schools. Without sufficient variation, the estimate of the effect of participation will be imprecise, and the analysis will lose statistical power.

The model allows for the direct control of the unobserved-school fixed effect, using binary variables representing each school in the sample. Controlling for the fixed effect in this way would address one possible source of selection bias—unobserved fixed school-level differences that are related to breakfast program participation and that affect the outcomes of interest.

d. Student-level participation model

The USBP design also proposes an HLM model to estimate the effects of breakfast program participation using student-level data. A simplified version of this model is:

(4) $Y_{Ij} = X_{ij}\beta_4 + Z_j\delta_4 + \alpha_4 P_{ij} + \beta_4 P_{ij} * USBP_j + \gamma_4 + \varepsilon_{4ij}$,

where:

P_{ij} = variable indicating the participation status of student *i* at school *j*.

The other variables in the model are defined as before. This model is similar to the student-level availability model, except that the key independent variables are the student's participation status and an interaction between participation status and the school's USBP status. The interaction is included to account for the fact that participation in the USBP may influence student outcomes differently from participation in the regular SBP. In particular, the estimated effect of participation in the regular SBP is represented by α_4, whereas the estimated effect of participation in the USBP is represented by $(\alpha_4 + \beta_4)$.

For the usual estimation techniques (OLS, logit/probit, tobit) to yield unbiased estimates of α_4 and β_4, the assumption that participation status is not correlated with the error terms of the model is necessary. In other words, unobserved factors influencing the outcomes of interest must not be correlated with whether or not students eat school breakfasts. As described previously, the nonrandom nature of the process by which students' participation status is determined leads us to question this assumption. If the assumption does not hold, then selection bias is possible.

The USBP design proposes three approaches to address the possibility of selection bias. The first approach attempts to ensure that selection bias does not arise in the first place. To prevent a correlation between participation status and unobserved determinants of key student outcomes, the model must explicitly control for all such factors. Ponza et al. (1999) identify three categories of factors that must be controlled for in the model: (1) detailed information on the students' socioeconomic status, (2) dietary habits of the sample members, and (3) dietary knowledge and attitudes of the food preparer in the sample member's household.

This approach for dealing with the possibility of selection bias is the best (if it is feasible), because it does not require complex econometric techniques or stringent assumptions. The limitation of this approach is that it is difficult in practice to control in sufficient detail for these three factors or for any other important factors affecting a student's decision to participate.

The second possible approach for dealing with selection bias in the student-level participation model is to estimate a fixed-effects model (that is, to directly control for an individual-specific fixed effect). In the model, this is a term representing unobserved factors that are specific to a given individual but constant over time, and that influence the outcome of interest. This approach requires data from more than a single point in time on the value of the outcome variable and on students' participation status (as well as on selected individual and school characteristics). However, given these data requirements and the fact that this approach does not control for time-varying sources of selection bias, Ponza et al. (1999) do not suggest that the USBP evaluation rely exclusively on fixed-effects methods for dealing with selection bias.

The third suggested approach is the estimation of instrumental variables (IV) and/or selection correction models. These are two-stage approaches to account for selection bias, with a first-stage equation that has participation status as the dependent variable, and a second-stage equation that is a version of the student-level participation model shown. In both models, information drawn from the estimation of the first-stage equation is used in the second stage to prevent a correlation between the participation variable and unobserved determinants of the outcome.

The major challenge to using this approach successfully is that it is necessary to find identifying variables for the model. In practice, identification of either model requires the inclusion in the first-stage equation of variables that influence students' participation status, but that do not directly influence the key outcome of interest in the second-stage equation. Finding appropriate identifying variables generally is difficult. The USBP design suggests the following variables:

- ***Timing considerations.*** These include variables that reflect the time that students must leave for school in the morning, the time that they arrive at school relative to the time school starts, the time that breakfast is served, and so on.
- ***Breakfast price and students' certification status.*** In regular SBP schools, these reflect what students must pay to obtain a school breakfast.
- ***Planted identifying variables.*** The design calls for exploration of an approach in which the evaluation team randomly assigns students to a treatment group that receives some incentive or encouragement to participate in the SBP/USBP and a control group that does not.
- ***Welfare status.*** The welfare status of students' families may reflect their attitudes about any stigma they might associate with receiving public assistance.

6. Statistical power

In the USBP design, Ponza et al. (1999) generated estimates of the MDDs of both USBP availability (based on an experimental design) and SBP/USBP participation. These calculations were based on a specific design in which the sample included 144 schools in the 6 demonstration districts and 30 students from each school, with the schools equally divided by random assignment between treatment and control schools. The authors calculated statistical power for both the experimental availability analysis and nonexperimental participation analysis, although the calculations of statistical power for the nonexperimental participation analysis assumed no selection bias. If this assumption is relaxed, the resulting analysis will have much less statistical power.

For the experimental analysis, two types of statistical power calculations were made. First, power was calculated under the assumption that students would be sampled randomly from each school. The second set of calculations assumed that there would be a preimplementation survey, and that students most likely to be new participants would be oversampled.[13]

Ponza et al. (1999) calculated the statistical power of models with test scores and several other outcomes as dependent variables, but we will focus on the former. The target MDD for this outcome is 10 percent of a standard deviation. A model that has an MDD greater than this target is not sufficiently powerful for our purposes.[14]

With a preimplementation survey, the MDD for the experimental availability analysis with student test scores as the outcome variable is 16 percent of a standard deviation. In other words, the true effect of breakfast program participation on achievement test scores would have to be more than 16 percent of a standard deviation in order for the USBP design to have at least an 80 percent chance of finding the estimated effect of USBP availability to be statistically significant. Thus, the design does not reach the target level of statistical power for detecting impacts on achievement test scores, given the assumptions that were made. If the impact of the USBP on test scores is larger than anticipated, however, the design may have enough power to detect it.

In the nonexperimental analysis of the effects of SBP/USBP participation on test scores, the MDD is just under 9 percent of a standard deviation in the test score measure (regardless of whether the design includes a preimplementation survey). As mentioned, however, this calculation assumes no selection bias. As long as this assumption is credible, the nonexperimental analysis may be sufficiently powerful to determine the influence of breakfast program participation on learning.

7. Design Costs

The costs of the USBP design have been broken down into two components, demonstration costs and evaluation costs. The costs of implementing the USBP demonstration have been calculated assuming a design in which 72 treatment schools would implement a USBP in six districts. The costs are based on the difference between the free meal reimbursements for all participating students in

[13]See Ponza et al. (1999) for an in-depth description of the details of the power calculations conducted without a preimplementation analysis.

[14]However, the target MDD refers to the effect of eating a school breakfast, not the effect of attending a school in which the USBP is available. The target MDD for USBP availability depends on the percentage of new participants in the total sample, as USBP availability will influence the mean outcome in the sample only through its effect on those who become new participants. If 25 percent of the sample consists of new participants, the actual MDD from the models of USBP availability must be multiplied by four before being compared with the target effect size. This calculation has been made in the MDDs shown below.

USBP schools and the combination of free, reduced-price, and paid reimbursements for participating students in the control schools (which offer the regular SBP). In addition, the estimates of demonstration costs are based on the following assumptions:

- Each school includes seven grades (kindergarten through sixth grade), with 70 students per grade.

- The participation rate in control (SBP) schools is assumed to be 30 percent, with 80 percent of breakfasts for students certified for free meals, 6 percent for students certified for reduced-price meals, and 14 percent for students paying the full price.

- The participation rate in treatment (USBP) schools is assumed to be 55 percent.[15]

Under these assumptions, the estimated cost of implementing the USBP demonstration is in the range of **$6.6 to $6.9 million**. The cost estimate for the evaluation based on the USBP design is based on the size of the winning proposal by Abt. This estimate includes the cost of the base contract (which covers the collection and analysis of two years of school records data and a single year of survey data), contract option 1A (which covers the collection and analysis of preimplementation data), and contract option 3 (which covers the collection and analysis of two additional years of data). The winning Abt proposal was approximately **$6.4 million**. Combining the demonstration and evaluation costs leads to overall USBP design costs of **$13.0 to $13.3 million**.

8. Design summary: strengths and weaknesses

The USBP design proposes an experimental approach to estimate the impacts of USBP availability on student outcomes, including measures of learning. Under this approach, schools in the six demonstration districts would be paired and randomized, with treatment schools implementing the USBP and control schools continuing to offer the regular SBP. Both cross-sectional and longitudinal data would be collected from a sample of students at these schools (as well as from the schools), and the outcomes of students in the treatment and the control schools would be compared to determine the impact of USBP availability.

This analysis would include both school-level and student-level analysis. Finally, the design also includes an analysis of the effects of breakfast program participation on student outcomes. This portion of the design would use a nonexperimental approach, with the outcomes of students who chose to participate in the breakfast program compared with those of nonparticipants. Statistical methods would be used to control for the relevant observed and unobserved differences between participants and nonparticipants.

A primary strength of the USBP design is that it uses a rigorous methodology to determine the impacts of USBP availability. This methodology should lead to unbiased estimates of the effect

[15] The actual participation rate in the USBP is likely to depend largely on how the program is implemented. In schools in which free breakfasts are served in a cafeteria, the actual participation rate is likely to be below 55 percent. In schools in which free breakfasts are served in classrooms, the participation rate may well be above 55 percent.

of availability. In addition, within the constraints of the demonstration and evaluation (in particular, the limitation to schools from six districts), funding is available to collect information from a large number of students in a large number of schools. Another strength of this design is that it involves the collection of data on a large number of outcomes. Although the discussion has focused on test score data, the design also calls for collection of data representing dietary outcomes, cognitive function, and attendance in school, as well as other school outcomes.

A final strength of the USBP design is that it already is being implemented. This Congressionally mandated 3-year project, which began at the start of school years 2000 and 2001, is in its second year of implementation. Thus, many of the normal obstacles to implementing an experimental evaluation of a large program have been overcome.

The USBP design has three major weaknesses with respect to informing the debate on the effects of the SBP on learning. First, although the USBP availability analysis is based on an experimental approach, the part of the design that examines the impact of SBP participation is based on a nonexperimental approach. As a result, unobserved factors that are related to students' decisions to eat a school breakfast and that are correlated with the outcomes of interest—or selection bias—may hinder the evaluator's ability to generate unbiased estimates of the effect of SBP participation.

The second weakness of the USBP design results from the dilution effect. The fact that the treatment and control groups will include many students who are not influenced by the intervention hinders the design's ability to produce sufficiently precise estimates of the effects of USBP availability. In other words, some students who eat a school breakfast under the USBP would have eaten a school breakfast under the regular SBP.

Others would not eat a school breakfast under either set of circumstances. Because there is no way to eliminate these students without undermining the beneficial aspects of the experimental methodology, their presence reduces the expected size of the impacts and makes it more difficult to detect these impacts statistically. Consequently, the USBP design does not have sufficient power to detect the expected impacts of USBP availability on one of the key outcomes of the demonstration, academic achievement.

Third, the USBP design is limited in that its data will not be nationally representative. Sample members will come from only the six school districts selected for the USBP demonstration. Furthermore, because of the demonstration requirements, these districts will not necessarily be even qualitatively representative of school districts (and students) nationally.

Despite these weaknesses, the USBP design produces a wealth of useful information on the SBP in general and on the USBP in particular. The data collected as part of the evaluation and the rigorous methodological approach will help minimize the weaknesses and emphasize the strengths of the design.

C. SBP Applicant Design

Although the USBP design uses an experimental approach, it does not entail the random assignment of students into breakfast program participant and nonparticipant statuses. Under the methodologically ideal experimental approach to examining the effects of SBP participation, students within schools would be randomly assigned to a treatment group that participates in the SBP or to a control group that does not.

This experimental approach would be difficult to implement in practice, however, as denying school breakfasts to eligible students would raise ethical questions, and it would be hard to find schools willing to do this. The SBP applicant design uses an experimental design that would be easier to implement, and that would come closer than the USBP design to the methodological ideal of randomly assigning students to SBP participant and nonparticipant statuses.

The SBP applicant design involves schools that apply to participate in the SBP for the first time. Because these schools will not have offered SBP breakfasts, randomly assigning students to a control group that is denied access to the program would not amount to taking away a benefit they already had. However, randomly assigning students to treatment and control groups would be difficult to implement at a practical level. Thus, the SBP applicant design calls for randomly assigning the classrooms of applicant schools to a treatment group that participates in the breakfast program or to a control group that does not. Finally, so treatment group sample members would be most likely to eat a school breakfast, these meals would be free to all students in treatment classrooms.

1. Intervention and counterfactual

The intervention in the SBP applicant design is the offer of a free breakfast; this offer would last only for one school year. During that year, students in the treatment group would attend classrooms in which free school breakfasts were available. The breakfasts could be served in any way the participating schools chose, as long as they were available only to students in treatment classrooms. Two possibilities are that (1) breakfasts are served in the classroom, and (2) cafeteria breakfasts are served only when students in treatment classrooms have access to the cafeteria.

In the SBP applicant design, the counterfactual is not having access to any breakfast program (and thus not participating in one). Control group students would attend classrooms in which, as in the previous year, they do not have the option of eating a school breakfast.[16] However, these students may have eaten a breakfast at home or away from home.

[16]Although no control group students would participate in the breakfast program according to the design, some of these students may “cross over” and obtain school breakfasts in practice. For example, if breakfasts are served in the school cafeteria, control group students may somehow leave their classes and go to the cafeteria for breakfast. Alternatively, treatment group students who already have eaten potentially could obtain a breakfast and then give it to friends who are control group students. An evaluation based on the SBP Applicant design would have to monitor such possibilities closely.

2. Basic design approach

The SBP applicant design uses an experimental approach consisting of five steps: (1) selecting participating schools, (2) randomly assigning classrooms, (3) implementing the breakfast program, (4) selecting a student sample, and (5) collecting and analyzing the data.

a. Selecting participating schools

SBP applicant design schools would be chosen from the pool consisting of schools applying to USDA to join the SBP. The underlying logic is that denying school breakfasts to control group students would be more feasible than it would be in schools already offering the program. In other words, the control group would not lose a benefit it already had; these students would be in the same situation they were in during the previous year. The following schools would be chosen from the pool of SBP applicant schools:

- Elementary schools
- Schools with more than some minimum number of low-income students
- Schools that are willing to implement random assignment and to delay full implementation of the SBP for one year in control classrooms.

Schools applying for the SBP are clearly signaling their desire to implement the SBP in full immediately. Therefore, they would likely need an incentive to agree to delay full implementation for one year and instead implement random assignment. In this design, allowing these schools to implement a USBP in the treatment classrooms would be the incentive. Thus, although control classrooms would not participate in a breakfast program, students in treatment classrooms would receive free breakfasts. Additionally, USDA would reimburse the school at the higher free-breakfast rate for each breakfast served, regardless of students' certification status.

We expect that there would be enough appropriate applicant schools to meet requirements for generating sufficient statistical power. In the most recent year for which data are available, 1,500 schools applied for SBP benefits (Food Research and Action Center, 1999). Many were not elementary schools, and many probably had few low-income students. However, even if only one-fourth the schools that applied were elementary schools with a nontrivial number of low-income students, 325 schools were potentially eligible for a demonstration based on the SBP applicant design. That number of eligible schools would be large enough to support a rigorous study (assuming a substantial number could be persuaded to participate).

Of course, including only SBP applicant schools in the demonstration limits the generalizability of the results from an evaluation based on this design. In effect, the results would not be generalizable beyond the SBP applicant schools meeting the criteria listed here. These schools probably are different in a number of respects from schools that currently participate in the SBP. For example, current SBP schools tend to serve a larger percentage of low-income students than do non-SBP schools.

b. Randomly assigning classrooms

After applicant schools have been chosen for the demonstration, the next step would be to randomly assign classrooms in those schools to treatment and control groups. Random assignment at the classroom level gives the SBP applicant design greater statistical power than would random assignment at the school level.

Such assignment would have to occur prior to the beginning of the school year, to give schools and classrooms time to prepare to implement the program. In each school, classrooms would first be matched by grade level, with the set of classrooms restricted to grades 2 through 5.[17] If the school had any tracking of students across classrooms according to ability level, then classrooms consisting of students with similar ability levels would be matched prior to random assignment. After classrooms were matched in this way, random assignment would be implemented, and one classroom within each matched pair would be assigned to the treatment (breakfast program) group and the other to the control (no breakfast program) group.

c. Implementing the breakfast program

The schools would have to decide where breakfast should be served for each treatment classroom, and at what time. In addition, a key implementation issue would involve how to ensure that control group students do not have access to the breakfast program. To strictly control which classrooms are served breakfast, the breakfast could be served in the classroom. Many USBP schools currently serve breakfasts in the classroom (see Wahlstrom et al., 1997). Another possibility would be to serve breakfast in the school cafeteria, but only after the official start of the school day. Treatment group classrooms would then be brought to the cafeteria as a group and would be allowed either to eat a school breakfast or to engage in some other activity there.

The breakfasts served to treatment group students would be free, for two reasons. First, receiving the free meal reimbursement rate from USDA would serve as an incentive to SBP applicant schools to participate in the demonstration. Second, free breakfasts for all students would encourage student participation in the program, thereby increasing the statistical power of the analysis.

Participating schools would have to determine the activities in which control group classrooms would engage while the treatment group classrooms ate breakfast. If these activities include instruction or study periods, the design could lead to a negative effect of the program on the amount of time devoted to teaching in treatment classrooms. Therefore, teachers in control group classrooms could offer their students some type of noninstructional activity instead.

This design calls for the demonstration to last for one school year. After that, schools would be eligible to participate in the regular SBP. All students (regardless of treatment status) would become eligible for the SBP under the regular rules for free- and reduced-price meal certification, meal pricing, and reimbursement.

---

[17]The classrooms randomly assigned would be limited to those in grades 2 through 5 because baseline data (including test scores, we hope) would be required from the previous school year. This type of data may not be available for first-grade students, who would have been in kindergarten the previous year.

d. Selecting a student sample

Although the SBP applicant design calls for the random assignment of classrooms, data would be collected on individual students from those classrooms. Because collecting survey data from each student in each treatment and control group classroom would likely be prohibitively expensive, the design calls for sampling of students in the classrooms. In particular, the most advantageous sampling plan would involve stratified random sampling, with oversampling of low-income students.

Information on students' income levels would be difficult to obtain prior to sampling, so students certified for free and reduced-price *lunches* would be oversampled, a process that would produce two benefits. First, low-income students would be the most likely ones to participate in the breakfast program. Second, these students probably would be less likely than higher income students to eat a breakfast both at home and at school. Among program participants, the fewer students who have already eaten breakfast at home, the larger the potential impact of the program.

e. Collecting and analyzing the data

The SBP applicant design would use a cross-sectional data collection plan primarily. Sampled students would be interviewed at the end of the school year and asked about their breakfast program participation during that year, as well as about other experiences in school. Administrative data on sample members' experiences and school performance during the year also would be collected. A key purpose of the data collection would be to obtain information on students' test scores.

Ideally, schools would administer their own tests during the spring, and this information would be available in the schools' administrative data systems. Otherwise, the evaluator would have to administer tests during that spring. Finally, administrative data from the previous school year would be collected so that any change in students' outcomes between the predemonstration year and the demonstration year could be measured. This information would be particularly helpful if it included test scores.

3. Data collection

The SBP applicant design data collection effort would be similar to that planned for under the USBP design. The specific data collected would depend on the level of resources to be expended on the evaluation but could include the following components:

- ***Student and parent data collection.*** Surveys of students in the sample and surveys of the students' parents would gather information on students' breakfast-eating habits, participation in the breakfast program (for treatment group students), experiences in school, and demographic/socioeconomic characteristics. If resources allow, dietary intake data could be collected from students, with their parents' assistance. Another option would be to administer short-term cognitive tests to students, provided administration was on a day for which we had information on breakfast program participation status.

- ***School records data.*** The collection of school records data would be an important part of the overall data collection plan. In particular, information on test scores for the spring of the demonstration year and for the previous year would be important, as would information on attendance, tardiness, grades or other measures of classroom performance, disciplinary problems, visits to the school nurse, and other relevant information.

- ***Teacher data collection.*** As in the USBP design, the teachers of sample members would be able to provide useful information on student outcomes and could be surveyed at the end of the demonstration year. They could provide information on students' behavior and performance in class. Scales such as the Connors Teacher Rating Scale (CTRS) or the Pediatric Symptoms Checklist (PSC) should be considered for a teacher survey.

- ***Cafeteria manager data collection.*** Because it would be important to understand how breakfasts were served to treatment group students, a cafeteria manager survey would be useful. This survey would yield information on the time and place that school breakfasts were served, as well as on other characteristics of school breakfasts.

- ***Qualitative data.*** A final way to collect information on the implementation of school breakfasts in treatment classrooms and on any treatment-control differences not apparent from other forms of data would be to collect qualitative data. This form of data would be collected through evaluation staff's site visits to treatment and control classrooms of participating schools during the demonstration year.

4. Measurement of key characteristics and outcomes

Since school classrooms will be randomly assigned to the treatment and control groups, a key focus of the data collection will be classroom characteristics, which will serve both as control variables in the models measuring the impacts of breakfast program availability and as outcome measures. Such characteristics as the number of students in a classroom and the teacher's educational attainment and experience would be important control variables in the model.

Other variables, such as teachers' assessments of disciplinary incidents and the level of student performance, could serve as outcomes; the effect of school breakfasts being available in a classroom on student behavior and performance would be measured. The teacher survey would be used to collect the data on classroom characteristics and outcomes. Any additional classroom-level data available through school records also would be collected.

Other measurement issues and the appropriate way to address them are similar in the SBP applicant design and the USBP design. Accurate measurement of SBP participation is critical, as is measurement of student achievement as an indicator of learning. For a full description, see Section B of this chapter.

5. Analysis plans

Unlike the USBP design, which has several types of analysis and models to estimate, the SBP applicant design calls for a single type of analysis and model that should be sufficient to determine the impact of the breakfast program on student learning.

This type of analysis involves estimating the effect of breakfast program availability on learning at the student level. Because each participating school will have both treatment and control group classrooms, there is no school-level analysis.[18]

Although it would be feasible to conduct a separate analysis of the effects of breakfast program participation (analogous to the student-level USBP participation analysis), we hope that this step would not be necessary. If the availability analysis yields a sufficiently precise estimate of the impact of being in a classroom that offers school breakfasts, the estimate of this impact, along with the participation rate in treatment group classrooms, can be used to generate an estimate of the impact of participating in the breakfast program.

In particular, because the overall impact of breakfast program availability would occur entirely through students who actually eat a school breakfast, dividing the estimate of the impact of availability by the participation rate will yield an estimate of the impact of participation. For example, if attending a treatment group classroom that offers school breakfasts leads to an increase in test scores of 0.06 of a standard deviation, and if half the students in treatment group classrooms eat school breakfasts, then the impact of participation would be 0.12.[19]

Given the experimental approach of the SBP applicant design, a simple comparison of mean outcomes among treatment and control group students would yield an estimate of the impact of breakfast program availability. However, controlling for other relevant factors in a regression model would yield a more precise estimate of this effect. In particular, the SBP applicant design analysis plan calls for estimation of the following model:

(5) $Y_{ic} = X_{ic}\beta + W_c\gamma + \alpha\, SBP_c + w_c + \varepsilon_{ic},$

where:

Y_{ic}	=	value of outcome among student *i* in classroom *c*
X_{ic}	=	vector of characteristics of student *i* in classroom *c*
W_c	=	vector of characteristics of classroom *c*
SBP_c	=	treatment status of classroom *c*.

In this model, the error structure includes a random classroom-specific term and an independent and identically distributed error term. The model's estimate of the coefficient α is the estimate of the overall breakfast program impact. As mentioned, dividing this impact estimate by the participation rate in treatment group classrooms yields an estimate of the impact of participation.

The potential complications in the estimation of this model are similar to those in the estimation of the student-level availability model under the USBP design. Because the availability of the

---

[18]Conceivably, there could be classroom-level analysis, with classroom-level measures of student performance in treatment group classrooms compared with those in control group classrooms. However, we do not recommend pursuing this type of analysis, as it is unlikely that classroom-level data will be available.

[19]Making an adjustment of this type requires the assumption that nothing in the process of random assignment itself affects the outcome variable.

breakfast program influences only students who participate, this effect is "diluted" by the presence of nonparticipants in treatment group classrooms. If this dilution effect is large enough, the design may lack sufficient power to detect impacts of the size likely to be generated by the demonstration.

In this case, we could turn to nonexperimental methods to examine the effects of breakfast program participation on student outcomes. In particular, mean outcomes in participants in treatment group classrooms would be compared with mean outcomes in nonparticipants, after controlling for relevant characteristics in a regression framework. Because design is nonexperimental, we would be concerned about the presence of selection bias.

6. Statistical power

We must make several assumptions in order to determine the statistical power of the SBP applicant design. Following the USBP design, the key outcome for which we will examine minimum detectable differences (MDDs) will be test scores. The target level for the MDD will also be based on the same estimate, from Meyers et al. (1989), that the impact of SBP participation on test scores is 0.10 of a standard deviation. However, some students in the treatment group do not participate in the breakfast program, so this MDD target is not the relevant one for the SBP applicant design.

In particular, if the true effect of participation is 0.10 of a standard deviation, then the true effect of breakfast program availability will be (0.10) × (participation rate) standard deviation. We would oversample certified students, so we assume that the resulting participation rate will be 40 percent.[20] Thus, the target MDD for this design would be 0.040 of a standard deviation. In other words, the experimental-based analysis of breakfast program availability should be able to detect an effect of availability at least as low as 0.040 of a standard deviation.

Additional assumptions related to the power calculations include the following:

- Estimates based on the design are intended to generalize to the schools participating in the demonstration, but not beyond those schools to a wider population. This assumption is analogous to the assumption in the USBP design that the power calculations are not intended to generalize beyond the six demonstration districts. It is based on the fact that the schools will not be randomly chosen for the demonstration.

- Up to 100 schools will be included in the demonstration and evaluation.

- Each demonstration school includes grades 2 through 5 (the grades included in the demonstration) and has four classrooms per grade that can be evenly divided between treatment and control classrooms. Each classroom has at least 15 students.

[20]Oversampling certified students will create a design effect. Although we do not know the exact magnitude of this design effect, we do know that the greater the extent of oversampling, the greater it will be. For simplicity, we ignore the design effect in calculating the MDDs in this section. However, if a full SBP applicant design plan is to be developed, this design effect should be taken into consideration.

- Overall, 10 percent of the total variation in test scores is explained by cross-school variation, and another 5 percent is explained by cross-classroom (but within-school) variation.

Table III.1 shows the MDDs under alternative assumptions about the number of schools and number of students per classroom in the demonstration. The table suggests that increasing the sample size by increasing the number of schools in the demonstration has the largest effect on the MDDs. For example, moving from a sample of 5 students in each of 16 classrooms in 50 schools (for a total sample size of 4,000 students) to a sample of 5 students in 16 classrooms in 100 schools (8,000 students) leads to a decrease in the MDD from 0.069 to 0.049.

By contrast, increasing the total sample to 8,000 students by increasing the number of students per classroom from 5 to 10 leads to a decrease in the MDD from 0.069 to only 0.054. After the maximum number of schools has been reached, however, the decrease in the MDD caused by an increase in the number of students per classroom from 5 to 10 is greater than the decrease caused by an increase from 10 to 15.

Table III.1—SBP applicant design minimum detectable differences

SCHOOLS	**Classrooms per school**	**Students per classroom**	MDD
Number			
50	16	5	0.069
50	16	10	0.054
50	16	15	0.048
75	16	5	0.056
75	16	10	0.044
75	16	15	0.039
100	16	5	0.049
100	16	10	0.038
100	16	15	0.034

The analysis indicates that, to achieve sufficient statistical power, the SBP applicant design requires 16 classrooms per school (including 8 treatment and 8 control classrooms) and a sample of either 100 schools and 10 students per classroom (for a total sample size of 16,000) or 75 schools and 15 students per classroom (for a total sample size of 18,000).[21]

[21]Actually, with 10 students per classroom and 16 classrooms per school, a sample of 94 schools (for a total sample size of 15,040) would be sufficient to yield an MDD of 0.039.

This result is dependent on the assumptions shown above, which should be more rigorously assessed if this design goes forward. However, this preliminary power analysis suggests that the design would require a large effort to recruit participant schools and to implement the demonstration program in treatment classrooms, as well as to collect data on the sample and to conduct the analysis.

7. Design costs

The cost estimates for the SBP applicant design include three components: (1) demonstration costs, (2) data collection costs, and (3) analysis and reporting costs. The key assumptions on which each component is based are described here.

The SBP applicant design calls for free breakfasts to be served in half the classrooms of demonstration schools and for no breakfasts to be served in the other half. If this demonstration program is not implemented, we assume that all these schools would offer students the regular SBP.

Thus, demonstration costs for the SBP applicant design have been calculated as the difference between the cost of providing free breakfasts to participating students in half the participating schools' classrooms (but with no additional costs, as the remaining classrooms have no breakfast program) and the cost of providing a combination of free, reduced-price, and full-price breakfasts to participating students in all classrooms.

In other words, the implementation of the demonstration includes not only the cost of free meals served to students in treatment classrooms, but also the savings from not serving any breakfasts to students in control classrooms.

Additional assumptions on which the calculation of demonstration costs is based include the following:

- The demonstration involves 100 schools with 8 classrooms (of 25 students each) assigned to the treatment group and 8 classrooms (of 25 students each) assigned to the control group.
- The participation rate in treatment classrooms is assumed to be 75 percent.[22]
- The participation rate under the regular SBP is assumed to be 30 percent, with 80 percent of breakfasts served to students certified for free meals, 6 percent to students certified for reduced-price meals, and 14 percent to students who pay the full price for breakfast.

Given these assumptions, the estimated cost of implementing the SBP applicant demonstration is $0.9 to $1.0 million.

As described, the evaluation costs include both data collection costs and analysis and reporting costs. In addition to the assumption that 100 schools (spread across 80 school districts) would

[22]This assumed participation rate is higher than that assumed for the USBP demonstration; we feel that it would be more likely in this demonstration for breakfasts to be served in the classroom, rather than in the cafeteria.

participate, with 16 classrooms in each school, we assume that a sample of 10 students in each of the 1,600 classrooms would be selected for the evaluation sample, leading to a total sample size of 16,000.

The data collection effort would include two years of school records data collection and the one-time administration of a single set of surveys to school principals, cafeteria managers, classroom teachers, students, and their parents. Data to be collected from students would include dietary intake data and the administration of short-term cognitive tests. With these data collection activities, the estimated total data collection costs for the SBP applicant design are estimated to be $10.0 to $10.5 million.

The estimated analysis and reporting costs assume a two-year evaluation period with one major report. These estimated costs fall in the range of $0.7 to $0.8 million. Together with data collection costs, total evaluation costs for the SBP applicant design are $10.7 to $11.3 million. Total costs for fully implementing the SBP applicant design and evaluation are $11.6 to $12.3 million.

8. Design summary: strengths and weaknesses

The main strength of the SBP applicant design is its experimental design, which yields unbiased estimates of the effect on student achievement and other outcomes of being in a classroom that offers free breakfasts (relative to having no breakfast program). Furthermore, treatment status is much more closely correlated with SBP participation status in the SBP applicant design than it is in the USBP design, the other experimental design.

Although the SBP applicant design does not reach the methodological ideal of randomly assigning students to participant and nonparticipant statuses, it comes as close as is feasible. In addition, by randomly assigning classrooms, as opposed to schools, the design generates more statistical power than does the USBP design. Finally, although this design does not yield nationally representative estimates, the large number of participating schools should give the results broad geographic representation.

One of the primary weaknesses of the SBP design is its difficulty of implementation. While it easy to stipulate in the design that classrooms will be divided into control and treatment groups, at the school level, this type of differentiation can cause problems, such as resentment amongst the teachers and families associated with the classrooms chosen for the control group. Furthermore, not all schools would be willing to serve breakfast in the classroom, as this creates extra work for teachers and custodial staff. Teachers may also feel that meal service takes away valuable instruction time. Finally, it could potentially be very difficult to match four classrooms at each grade level.

Another weakness of the SBP design is the possibility that the resulting estimates would not have sufficient statistical power. The power of the design relies largely on the ability of the evaluator to recruit a sufficient number of schools for the demonstration. In particular, the design requires roughly 100 elementary schools that are applying for participation in the SBP; have a nontrivial number of students certified for free or reduced-price meals; and are willing and able to delay full SBP implementation for one year, conduct random assignment of classrooms into breakfast program and no breakfast program groups; and implement a free breakfast program in treatment classrooms only. Analysis based on substantially fewer than 100 schools is likely to result in estimates of the impact of the SBP that are statistically insignificant, whether or not the true impact of the program is positive.

The SBP applicant design also has several other weaknesses. If breakfasts are delivered in treatment group classrooms during the school day, it is likely that treatment group classrooms will have less time than control group classrooms for instruction. In addition, free breakfasts delivered to classrooms may induce some students who already have eaten breakfast to participate in the program, an outcome that the evaluation should track.

This participation, if it occurs, would decrease the likelihood that program participation will have a detectable positive influence on learning outcomes. Finally, because the schools eligible for the demonstration are limited to those applying to participate in the SBP, the sample will not be representative of students in current SBP participant schools, which tend to serve large percentages of low-income students.

An additional consideration in assessing the SBP applicant design within the context of the four alternative designs presented in this report is that it is fairly similar to the USBP design, which currently is being implemented. Thus, the degree to which it can provide insights in addition to what is learned from the USBP evaluation should be carefully considered. However, although the USBP design examines the impact of free school breakfasts relative to the regular SBP, the SBP applicant design examines the impact of free school breakfasts relative to no breakfast program. Thus, despite the similarity of the two designs, the SBP applicant design has the potential to answer a unique question.

D. ECLS-K-Based Design

In this section, we describe a design for using the ECLS-K, Kindergarten Cohort, to relate information on students' SBP participation to cognitive performance. The ECLS-K consists of a nationally representative sample of 16,906 students who were in kindergarten programs in 866 schools as of fall 1998.[23] The study is scheduled to include multiple assessments of children's cognitive, physical, social, and emotional development over time; it will also include surveys of families and schools for children in kindergarten, grade 1, grade 3, and grade 5. With information in the ECLS-K, perhaps supplemented by an expanded questionnaire, it is possible to estimate, in a nonexperimental context, the consequences of SBP participation on both children's cognitive performance and related outcomes from kindergarten through grade 5.

1. Intervention and counterfactual

The primary intervention that this design examines is SBP participation, or eating a school breakfast in a school breakfast program funded (presumably) by USDA. In particular, the intervention is defined as usually eating a school breakfast over the course of a school year while attending a school in which the school administrator reports that students are certified to receive

[23]The study originally targeted approximately 21,000 students in 995 schools, but only 16,906 survey respondents have both child and parent information, and only 866 schools have positive sample weights.

free breakfasts.[24] Furthermore, the intervention refers to participation in the SBP *as it is currently administered in schools throughout the country*.

Although policymakers may be interested in examining versions of the SBP that deliver breakfasts in new and/or innovative ways, the ECLS-K is not suitable for this purpose. In addition, the definition of the intervention requires more specifics as to what "usually eats a school breakfast" means.

The counterfactual in the ECLS-K design is "not eating a school breakfast," or "not participating in the SBP." This design can examine the impact of a school simply offering the SBP to students, as well as the impact of actual SBP participation. In particular, the design can address the following four questions:

- What is the impact of SBP participation on current participants?

- What would be the impact of SBP participation on current nonparticipants in schools offering the SBP, schools not offering the SBP, or both?

- What is the impact of offering the SBP in schools where it currently is offered?

- What would be the impact of offering the SBP in schools where it currently is not offered?

To address the first two questions, the impact of SBP participation would be inferred by using statistical methods to compare outcomes in a group of current SBP participants and a comparable group of nonparticipants. To address the third and fourth questions, the impact of offering the SBP would be inferred by using statistical methods to compare outcomes in a group of students attending SBP schools and a comparable group of students attending non-SBP schools.

Because the counterfactual includes both students who eat no breakfast and those who eat a nonschool breakfast, the main impact being estimated does not tell us the effect of eating breakfast versus no breakfast. Instead, it tells us the effect of eating a school breakfast versus eating no school breakfast. This could occur both because the school breakfast program makes children more likely to eat breakfast and/or because the program influences the foods consumed by those who do eat breakfast. This issue can be examined by comparing outcomes among SBP participants, nonparticipants who eat breakfast, and children who eat no breakfast.

2. Basic design approach

The ECLS-K design uses a nonexperimental, comparison group design. This approach involves comparing key outcomes in a group of students who participate in the SBP with outcomes in a group of students who do not eat a school breakfast. As suggested in Section 1, these comparison group students could attend SBP schools, but choose not to eat a school breakfast, or could attend

[24] Some schools may operate school breakfast programs of their own, without USDA funding. However, we assume that, in such instances, the school administrator would be less likely to report that students are certified as eligible for free breakfasts at the school.

non-SBP schools and therefore not have the option of eating a school breakfast. The analysis would be conducted primarily at the student level, but there also would be some school-level analysis.

Because this design is nonexperimental, students are not randomly assigned to a treatment group that receives breakfast and a control group that does not receive breakfast. Therefore, SBP participants and nonparticipants may differ in ways that influence outcomes related to learning. The nonexperimental design must account for the extent to which differences in key outcomes are the result of these differences in student characteristics, rather than the result of eating a school breakfast. Statistical methods, such as instrumental variables models, may be used for this purpose.[25] After the differences in participants' and nonparticipants' characteristics have been accounted for, any remaining differences in key outcomes can be attributed to the influence of SBP participation.

The ECLS-K design is based primarily on cross-sectional analysis. This analysis would involve measuring students' cognitive functioning at a point in time and comparing this outcome in students who report (at the same point in time) that they are SBP participants versus those who report that they are not participants. The analysis is cross-sectional because the cognitive functioning outcomes and the SBP participation variable each are reported at a single point in time for each student.

The primary strength of the ECLS-K data for studying the impact of SBP participation on learning is that it contains detailed information on various aspects of children's cognitive development. One of the primary motivations for the ECLS-K data collection was to track various aspects of young children's development, so great efforts were made to collect a broad range of high-quality data on cognitive outcomes. The resulting data will provide the most accurate picture available on the cognitive functioning of a large, nationally representative sample of elementary students.

In addition to having high-quality measures of students' learning, the ECLS-K data include several questions related to SBP participation. This information is available at both the school level and the student level. When combined with the detailed information available on student outcomes, family income and demographics, and school characteristics, the breakfast information in the ECLS-K enables researchers to estimate the effects of SBP participation on learning, accounting for a variety of characteristics that distinguish SBP participants from nonparticipants.[26]

Two other aspects of the ECLS-K make it particularly appealing for a study of the relationship between SBP participation and learning. First, it is based on a very large, nationally representative sample. The sample of more than 16,000 students is large enough to support analysis of the relationship between SBP participation and learning both in the overall sample and among key

---

[25]These methods are described in greater detail in Chapter IV.

[26]See Chapter IV for further discussion of ECLS-K data and measurement issues.

subgroups. Second, the longitudinal nature of the data not only provides the analytic advantages described here, but also allows the participation-learning relationship to be estimated at various points in time. Thus, the analysis can reveal whether the effects of eating a school breakfast are different for children of different ages.

3. Data collection

The ECLS-K collects some breakfast-related information in its surveys of parents, teachers, and school administrators. However, we recommend collecting supplemental data to better evaluate the impact of SBP participation on learning at both the student and the school levels. A major goal of this additional data collection is to address the issue of selection bias by identifying factors that contribute to school breakfast participation but not directly to learning outcomes. Supplemental data also would help researchers measure school breakfast and school lunch participation more accurately on the student and school levels.

Additional information collected from parents would contribute to the analysis of variation in student outcomes *within* schools, whereas information collected from school administrators would contribute to the analysis of the variation in outcomes *between* schools. Accounting for more of the variation in student outcomes will produce more precise estimates of the SBP participation.

Additional data can be collected most conveniently by adding targeted questions to existing surveys of parents and school administrators. In Chapter IV, we discuss these surveys as well as our reasons for not recommending the collection of dietary intake data in the context of this evaluation design.

4. Measurement of key characteristics and outcomes

The ECLS-K measures various characteristics of children, their families, and their schools that are relevant to a study of how school breakfasts influence learning. This section briefly discusses those variables, indicators of SBP participation, and student and school characteristics that can be used as control variables. A more detailed description of the variables is presented in Chapter IV.

a. Outcome variables

The ECLS-K includes a rich set of variables that measure three types of outcomes: (1) a student's cognitive development; (2) events and processes associated with learning, such as school attendance and tardiness; and (3) other aspects of a child's growth, including emotional, social, and physical growth. Analysis of ECLS-K data should therefore enable researchers to estimate differences in a wide range of outcomes by SBP participation status.

b. SBP participation

Participation in the SBP is a key variable for the analysis of the link between school breakfast and learning. The ECLS-K collects information on SBP participation at the school and child levels. On the school level, principals are asked how many students are eligible for, and how many students participate in, the *free* breakfast program at the school. Principals are *not* asked how many other students receive school breakfasts, at either the reduced or the regular rate. This information could be gathered through the supplemental survey.

At the child level, the ECLS-K collects data from parents on whether their child's school serves breakfast, whether the child usually eats a school breakfast, and the number of times the child ate the breakfast during the previous 5 school days. The SBP participation variable can be defined differently, depending on the nature of the analysis. For an analysis of within-school differences in individual student outcomes, participation might be measured using the number of school breakfasts students ate during the previous 5 school days.

In a linear model, that definition would imply that the effect of SBP participation on learning is proportional to the number of days the child ate a school breakfast during the previous week, with the implicit assumption that the number of breakfasts consumed during the previous week is a good proxy for the number of breakfasts consumed in the typical week. A less restrictive specification would have separate indicators for each level of weekly participation—1, 2, 3, 4, or 5 days, to allow the effect of the number of breakfasts consumed per week to influence learning nonlinearly.[27]

c. Background characteristics of students and schools

Because the ECLS-K collects information from both schools and parents, it includes a rich set of background characteristics on students and schools. Access to rich data is particularly important because it makes it possible to control for factors important to SBP participation as well as to learning. Previous research on the determinants of SBP participation (based on data from the School Nutrition Dietary Assessment study for the 1991–1992 school year), indicate that a wide variety of factors are related to whether or not children eat school breakfasts.

Among children attending schools that offer the SBP, for example, participation has been found to be more likely among children who are younger, male, black or Hispanic, living in rural or suburban areas, and who are certified for free or reduced-price meals (Gleason, 1996). In addition, participation is correlated with the region in which the child lives and with the full price charged for breakfast in the school.

5. Analysis plans

This ECLS-K-based approach to analyzing the impact of school breakfasts on learning is nonexperimental in nature. The design must therefore rely on multivariate statistical methods to infer the difference that SBP participation makes on the educational outcomes of students.

Let us assume that we observe a particular learning-related outcome, Y_{is}, for student *i* in school *s*. We assume that the outcome is a function, $g(.)$, of the availability of school breakfasts at a child's school, A_s; of other school characteristics, Z_s ; of student background characteristics, X_{is}; and of school breakfast participation for the student, P_{is}:

(6) $Y_{is} = g(A_s, Z_s, X_{is}, P_{is})$

Examples of Y_{is} include the growth in student test scores from one year to the next and attendance patterns in a given year. Examples of characteristics in Z_s include teachers'

[27] Of course, any school breakfast participation variable will be affected by the degree of parental accuracy in reporting on participation, which is unknown at this point.

characteristics and factors related to a school's decision about whether and how to offer the SBP to students, such as the proportion of students at the school certified as eligible for free or reduced-price school *lunches*.

Examples of characteristics in X_{is} include family income, parental education, household food security, and a student's prior test scores.[28] (We also assume that X_{is} includes a constant term.) The participation variable, P_{is}, may be defined in a variety of ways, but for simplicity, we assume that it is a single indicator for "usual" SBP participation, as indicated through the ECLS-K parent survey.[29] Note that when the school does not offer school breakfasts ($A_s = 0$), students do not receive any school breakfasts ($P_{is} = 0$).

To estimate the relationship between SBP participation and learning outcomes, it is necessary to make assumptions about the relationship between the variables contained in equation (6). When using linear regression methods, we assume that equation (6) has the following form:

(7) $Y_{is} = \alpha A_s + \beta Z_s + \gamma X_{is} + \delta P_{is} + u_s + e_{is}$,

where u_s is a school-specific error term, and e_{is} is an individual-specific error term. In this equation, the coefficient α represents the contribution of the availability of school breakfasts to learning for nonparticipants; the coefficient matrix β represents the contribution of other school characteristics to learning; the coefficient matrix γ represents the contribution of personal and family characteristics to learning; and the coefficient δ represents the contribution of school breakfast participation to learning.

For simplicity of presentation, equation (6) specifies that a given level of SBP participation has a uniform effect on learning, regardless of the characteristics of the school or individual students. In practice, researchers should try estimating separate equations for different subgroups of students.[30]

Because we are likely to have multiple observations of students from each school, we can subtract the school-level means of *Y, X,* and *P* from equation (7) and estimate the following:

(8) $(Y_{is} - Y_s) = \gamma (X_{is} - X_s) + \delta (P_{is} - P_s) + (e_{is} - e_s)$.

---

[28]The ECLS-K includes an 18-item food security module that can be used to categorize households as "food secure," "food insecure without hunger," or "food insecure with hunger." In addition, ERS is currently developing a children's hunger scale from these items to assess whether there is hunger among children in the household.

[29]In the basic model, we will not use the variable measuring the actual number of days of participation during the previous five days at school, because—as we note in Chapter IV—this variable likely measures with some error the level of SBP participation contributing to educational outcomes. In practice, most students who "usually" participate in the SBP have received breakfasts during the preceding 5 days at school.

[30]In this chapter, we discuss the estimation of separate impacts of attending an SBP school for students with different propensities of SBP participation.

Note that the estimates obtained from equation (8) are the same as those that would be obtained by using a dummy variable for each school in equation (7) to capture the combined effect of $(\alpha A_s + \beta Z_s + \gamma X_s + \delta P_s + u_s)$. This "fixed school effects" estimator allows us to account for *all* time-invariant school-specific characteristics affecting outcomes without having to specify what those characteristics actually are.

If we believe that the existing school-level variables contain enough school-level information to predict school-level learning outcomes (that is, that there are no unobserved school-level effects correlated with unobserved individual-level effects), we can improve on the efficiency of the fixed-effects estimator through a method known as "random-effects." Assuming that $e_s = 0$, we can estimate the parameters of using the variation in average outcomes between schools, as expressed by the equation:

(9) $Y_s = \alpha A_s + \beta Z_s + \gamma X_s + \delta P_s + u_s.$

Equation (8) therefore captures the within-school variation in learning outcomes, whereas equation (9) captures the between-school variation in learning outcomes. Under the assumption that e_{is} and u_s are uncorrelated (that is, that there are no missing variables in equation (9) correlated with the unexplained portion of individual outcomes), it is possible to obtain efficient estimates of γ and δ as an optimally weighted average of the corresponding estimates from equations (8) and (9).[31] If this assumption does not hold, then the random-effects estimator is biased, and the fixed-effects estimator is preferable.[32]

Using a linear, fixed-, or random-effects model, we can include in the analysis schools that do not offer the SBP and can assume that the impact of SBP participation in these schools would be the same as in schools offering the SBP. Because this assumption may be unrealistic, researchers should estimate equations (8) and (9) both with and without schools that do not have an SBP. In this way, the researchers can determine whether including nonparticipating schools in the sample has any substantial effect on the parameter estimates of interest.

In general, including non-SBP schools will improve the efficiency of estimates of the effects of SBP participation on learning and will enable researchers to estimate the impact of attending a school offering the SBP. Various alternative models of the effect of SBP participation on learning can be estimated using ECLS-K data. Two of these alternatives are instrumental variables and propensity score models, each of which is designed to address different limitations of the basic model described above. Details of the alternative models, along with their strengths and weaknesses, are presented in Chapter IV.

---

[31] The assumption that e_{is} and u_s are uncorrelated is a particularly strong one. If the assumption does not hold, then the random-effects model will lead to biased estimates. However, it is possible to test the validity of this assumption using a specification test developed by Hausman (1983). For additional information on the formation of the random effects estimators, see Greene (1997).

[32] In addition to estimating fixed- and random-effects models, investigators may want to consider estimating hierarchical linear models (HLMs) that take into account the clustering of students in observation schools and that also allow the effects of the characteristics of classroom units on the outcomes of interest. See Bryk and Raudenbush (1992) for more details.

6. Statistical power

MDDs are related to the "power" of an analysis. The power of an analysis refers to the likelihood that it will detect, at a given level of statistical significance, a certain magnitude of difference in outcomes between program participants and nonparticipants. For given levels of power and statistical significance, the minimum difference between outcomes that can be detected by a particular analysis is known as the MDD. We calculate MDDs that can be detected with 80 percent power using a 95 percent confidence interval and a one-tailed test (details are discussed in Chapter IV). To do this in the context of the nonexperimental evaluation of the SBP, we rely on formulas developed by Gleason (2000).

The logic underlying these MDD calculations (both with and without selection bias) is that they are very similar to the calculations based on an experimental design except that treatment status (participation) is not independent of other components of the regression model—the control variables and, in the case of selection bias, the error term.

Thus, there is less "useful," or exogenous, variation in participation status on which to base the estimates. This problem is essentially one of the multicollinearity of the treatment variable: the greater the multicollinearity, the larger the MDD for any given level of power (or, conversely, the less the power of the analysis for any given MDD).

Given the large number of schools and students in the ECLS-K dataset, the models we propose for estimating the impact of SBP participation on learning have high levels of statistical power under a given set of assumptions. One of these assumptions is that the model does not suffer from selection bias and can be estimated without using an instrumental variables (IV) model. However, if this assumption does not hold, and an IV model must be estimated, the MDDs are much higher and the statistical power much lower. Additional details on the statistical power of the ECLS-K design can be found in Chapter IV.

7. Design costs

Because the ECLS-K-based design is based on analysis of secondary data, the estimated costs of the design do not include demonstration costs or data collection costs (except for the costs associated with supplemental data collection). The main design costs are associated with analysis and reporting. These costs fall in the range of $0.4 to $0.5 million.

Although the design does not include primary data collection, we suggest supplemental data collection activities; namely, adding questions to the ECLS-K parent and school administrator questionnaires (as described in Chapter IV). Although we do not have specific cost estimates for these supplemental data collection activities, we have estimated the implications of the suggested additions to the surveys in terms of the length of the questionnaires. In particular:

- Supplemental questions to the ECLS-K parent survey in the spring of 2002 (as well as in the spring of 2004) will increase the length of the existing instrument by approximately 2 pages, from 146 to 148 pages (or 1.4 percent). This instrument is administered to approximately 20,000 parents.

- Supplemental questions to the ECLS-K school administrator survey in the spring of 2002 (as well as in the spring of 2004) will also increase the length of the existing instrument by approximately 2 pages, from 36 to 38 pages (or 5.6 percent). This instrument is administered to approximately 1,000 school administrators.

8. Design summary: strengths and weaknesses

An ECLS-K-based design for evaluating the impact of the SBP on learning has strengths and weaknesses. The major strengths are that it relies on a large, nationally representative sample of elementary school students and schools and on a database that includes a rich variety of outcome measures and student background characteristics. The weaknesses are related to selection bias and the resulting difficulties of obtaining reliable impact estimates for certain subgroups of students.

Although IV methods may help correct for selection bias, and propensity score methods may help construct comparable subgroups of SBP participants and nonparticipants, each method requires the researcher to make what may be unreasonable assumptions about the determinants of SBP participation. That is, the researcher must assume either that certain variables affect participation but not outcomes or that the unobserved factors influencing outcomes are uncorrelated with the propensity to participate in the SBP. Supplemental data collection from parents and schools on additional factors influencing SBP participation (but not learning) can reduce the need to make unreasonably restrictive assumptions when using these methods.

Even if the assumptions required for unbiased estimation were true, however, the small number of observed SBP participants among higher income students makes it difficult to use ECLS-K-based impact estimates to generalize about the likely impact of a universal SBP. To understand the likely consequences of universal-free school breakfasts, an experimental USBP evaluation would appear to be necessary. At the same time, an experimental demonstration of the USBP would not enable researchers to estimate the impact of the SBP on the *current* participants, because this group presumably would continue to receive school breakfasts under the USBP.

In an experimental context, testing the impact of the SBP on current participants would require denying SBP benefits to them. This type of experiment would be both politically infeasible and ethically questionable. Consequently, using the ECLS-K to construct a comparison group of SBP nonparticipants may be the most feasible way to determine the impact of school breakfasts on current participants' learning.

E. NHANES Design

In this section, we describe a design for using NHANES to study factors associated with SBP participation and learning. The NHANES design is similar to the ECLS-K design in that it is nonexperimental and based on a national survey of school-aged children that contains information on SBP participation and family background.

Unlike the ECLS-K, however, NHANES includes comprehensive information on dietary intake, nutritional status, and health status but contains little information on learning outcomes, school attendance, social and emotional development, or school characteristics. Supplemental data collection to capture information on academic achievement, cognitive function, and school performance is recommended if the NHANES design is to be used to full advantage.

1. Intervention and counterfactual

In the NHANES design, the intervention is participating in the SBP (that is, usually eating a school breakfast). The counterfactual is nonparticipation (not usually eating a school breakfast). To study learning outcomes, we are most interested in an intervention that covers a substantial period of time—"usual" participation over the school year.

This information is collected in the NHANES as the number of times per week that the child selects a school breakfast. For the intervention group, we are interested in defining participation as selecting a school breakfast on most school days. In terms of the population targeted by the intervention, children in elementary school are of greatest interest; however, we are also interested in older children (especially girls), who are less likely to report eating breakfast.

The main counterfactual condition is attending a school in which the SBP is offered but not "usually eating" a school breakfast during the year. However, we may also want to examine the counterfactual condition of attending a school in which the SBP is not offered. Finally, it will be useful to examine differences in outcomes among SBP participants, nonparticipants who eat breakfast outside of school, and children who do not usually eat breakfast.

The last would require the addition of a survey question on children's usual breakfast habits. This analysis would involve examining whether the relationship between SBP participation and the outcomes of interest arises simply because SBP participants are eating any breakfast or because of the composition of the breakfasts they are eating.

2. Basic design approach

The NHANES design uses a nonexperimental approach to examine the effects of SBP participation on learning relative to the counterfactual of not eating a school breakfast. We would observe students' participation status and would then compare mean outcomes in participants and nonparticipants. The key to the design lies in the strengths and weaknesses of NHANES, the secondary data source to be used by the design.

The NHANES is conducted by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), U.S. Public Health Service, to collect information about the health, nutritional status, and diet of people in the United States (National Center for Health Statistics 1994). The NHANES is unique in that it combines a home interview with health and nutrition assessments conducted in a mobile examination center.

The NHANES covers a representative cross-sectional sample of the U.S. civilian, noninstitutionalized population. The sample design is a multistage, complex, stratified survey design of individuals living in households. Previous NHANES surveys were conducted on a periodic basis, but, beginning in 1999, the sample design was transformed into an annual nationally representative sample (National Center for Health Statistics NHANES web site, 2001).

Continual data collection increases the availability of timely data for health and nutrition policymaking and provides flexibility in changing the sample design and the survey content. These design features offer an opportunity for future funding to increase the sample size of school-aged children or to add content to the NHANES to study the relationship between SBP and learning.

The current survey design includes participants of all ages and racial/ethnic groups. The first three years of the survey (1999–2001) oversample of adolescents, blacks, and Mexican Americans. In 2000, oversampling of low-income white people was added to the design requirements in order to produce reliable estimates for the total population with incomes at or below 130 percent of the poverty line. Oversampling of the low-income population will provide a larger sample of children who are eligible for or participate in the SBP.

Data are collected on approximately 5,000 people from 15 primary sampling units (PSUs) per year. Although each year provides a national sample, data can also be aggregated across survey years to provide reliable national estimates for more detailed age, gender, and racial/ethnic groups. Table III.2 shows the total sample sizes expected for school-aged children by race/ethnicity for three survey years: (1) 2000, (2) 2001, and (3) 2002. The design calls for equal numbers of boys and girls in each age group.

Table III.2—Expected sample sizes for NHANES in 2000, 2001, and 2002

RACE/ETHNICITY	6–11 years	12–15 years	16–19 years	6–19 years
Non-Hispanic White/other	176	176	176	528
Non-Hispanic black	176	183	183	542
Mexican-American	176	190	190	556
Total for one year	528	549	549	1,626
Total for three years	1,584	1,647	1,647	4,878

Future decisions about the sample design requirements and annual sample sizes will depend on federal data needs for national data for specific populations groups. The potential integration of NHANES and the Continuing Survey of Food Intakes by Individuals (CSFII) may change the annual sample size of people for whom sociodemographic and dietary intake data are available.

However, the exact nature of the sample design and of the oversampling by age, socioeconomic, and race/ethnicity for survey years beyond 2005 is unknown at this time. Given uncertainty about the post-2005 period, our discussion of the NHANES design is based on currently known sample sizes and sample design.

The current survey sample includes about 1,626 school-aged children each year. Because the data are nationally representative, the information they provide on the relationship between SBP participation and learning is generalizable to all school-aged children nationally. The cross-sectional nature of the data means that key outcomes are measured at a given point in time in the school year. Thus, the mean values of key outcomes are collected at the same time and cover the same time period as the information on SBP participation.

There is possible selection bias related to the availability of the SBP in schools, parents' knowledge about the availability of SBP, and parents' decisions about their children's participation in the program. Accounting for SBP participation decisions in the analysis is important, because there is a risk that nonrandom selection into the sample of SBP participants will bias estimates of the impact of SBP participation on learning.

3. Data collection

NHANES measures much of the same family and school information that other national surveys, such as the ECLS-K, measure; however, it does not measure learning or other school-related measures. Nevertheless, the NHANES is unique in that it assesses the intermediate outcomes that may affect learning.

Nutrition and health status are potentially affected by SBP participation, but they in turn may also affect learning outcomes. Section 3.a describes the current NHANES data collection plans. Section 3.b describes possible supplemental data collection that might be conducted to expand the usefulness of the information for studying learning outcomes.

a. Current data collection plans

This section summarizes the current data collection plans for NHANES 1999-2002, by general topic area. Table III.3 contains a more detailed list of variables currently collected.

- ***Family characteristics.*** Information on the family's income, food assistance program participation, and use of emergency feeding assistance in the past year is available to characterize the food security and socioeconomic status of the child's family. Information on parents' education, health insurance coverage, and sources of medical care is also collected.

- ***School meals and behavior.*** Information on whether the child's school offers the SBP or the National School Lunch Program (NSLP), the child's frequency of participation in the SBP and NSLP, and whether the child receives free or reduced-price school meals is available from the parent interview. Additional information on the child's grade level, attendance, suspensions, expulsions, and skipped grades is also collected in the parent interview.

- ***Dietary intake and behavior.*** Dietary intake is assessed using 24-hour dietary recall methodology and additional interview questions about dietary habits. At least one 24-hour recall is collected per person, with a second day collected on a subsample. The 24-hour dietary recall provides information on whether breakfast was consumed, the time and source of breakfast(s), and the foods and amounts consumed at breakfast, as well as the total day's intake. Total nutrient intake is estimated using information collected about dietary supplement use, discretionary salt use, and water intake. Dietary habits, such as the frequency of eating away from home, are also collected.

Table III.3—Current variables in NHANES pertinent to school breakfast and learning

Family Characteristics

Head of household
- Education level
- Country of birth
- Employment
- Occupation

Health insurance coverage
Source of medical care
Number of families in household
Family members' relationships
Housing characteristics
Income, past 12 months
Food security instrument (18-item instrument used in Current Population Survey), past 12 months
Emergency Feeding, past 12 Months

Food Program Participation (Family-Level)

WIC: Number of months in past 12 months and current month
Food Stamp Program: Number of months in past 12 months and current month

Individual Characteristics

Age
Race/ethnicity
Country of birth

School Meals and Behavior

National School Lunch Program: Does school offer? Number of times per week has free/reduced-price/full-price lunch
School Breakfast Program: Does school offer? Number of times per week has free/reduced Price/full-price breakfast
Grades at school

Suspensions and expulsions
Grade level and skipped grades
Attendance (missed days due to illness)

Dietary Intake and Behavior

One in-person 24-hour dietary recall (second day on subsample)
Use of dietary supplements
Additional food security questions
Number of times the child eats away from home per week
Salt usage

Blood Determinations Related to Nutrition or Health Status

Iron status (hemoglobin, hematocrit, serum iron, tibc, serum ferritin, transferrin saturation)
Serum and RBC folate
Serum vitamin E
Serum vitamin A and retinyl esters
Serum carotenoids
Plasma homocysteine
Methyl malonic acid
Serum vitamin B_{12}
Cotinine (passive smoke or cigarette exposure)
Lead

Health-Related Behaviors

Number of times per week child plays or exercises enough to sweat or breathe hard
Number of hours of television/video watching yesterday
Number of hours of computer use yesterday
Smoking
Alcohol and drug use

Health Interview Data

Reported medical conditions
Whether mother smoked during pregnancy
Birthweight and whether full-term
Vision problems; need for corrective lenses

Health Examination Data

Height, weight, anthropometric measures (for assessing growth, overweight)
Bioelectrical impedance analysis, body composition (ages 8 and older)
Dental exam for caries, periodontal disease
Blood pressure (ages 8 and older)

- ***Blood determinations related to nutrition or health status.*** These variables provide information on intermediate outcomes relating to short-term and long-term nutritional status and health. Iron, B-vitamins, and lead are related to brain development while the nutritional biochemistries have the potential to serve as biomarkers of dietary intake and status. Increased dietary quality or dietary status would be expected to result in improved nutritional status.

- ***Health-related behaviors.*** Health behaviors assessed for younger and older children include physical activity levels and time spent on sedentary activities, such as watching television, playing video games, and using personal computers. Risk behaviors include smoking for children aged 8 years and older and alcohol and drug use for those aged 12 years and older. Risk behavior information is collected during private, self-administered computer-assisted personal interviewing (CAPI) in the mobile examination center.

- ***Nutritional status.*** The NHANES provides the most comprehensive picture of nutritional status available on a national sample of school children. Precise anthropometric measurements, such as height and weight, are used to assess growth and overweight in relation to the revised CDC growth charts (Kuczmarski et al., 2000). Blood and urinary measurements provide an assessment of vitamin and mineral status for a wide variety of nutrients, such as B vitamins and iron. Iron status is of particular interest, because iron deficiency is related to developmental and behavioral disturbances that may affect mental performance and learning in young children (Centers for Disease Control and Prevention, 1998).

- ***Health status.*** General health status measures, such as physical fitness, blood pressure, and respiratory disease, provide an overall picture of the child's health and readiness to learn. Other health components included in the NHANES related to a study of breakfast and learning include visual acuity and hearing problems, which may affect classroom learning, and environmental exposures, such as to lead. Frequent health problems and illnesses may lead to more missed days of school and fewer opportunities to learn. Elevated levels of lead in the blood can be associated with iron deficiency anemia and are higher among low-income children (Centers for Disease Control and Prevention 1998). Variables that relate to the child's prenatal environment, such as low birth weight or exposure to smoke, have been shown to relate to growth and development and are collected in the parent interview.

b. Supplemental data collection

The primary weakness of the NHANES for a study of SBP and learning is that there are no current plans to collect cognitive measures or achievement data for school-aged children. Another weakness is that the current measure of SBP participation may be somewhat imprecise. We recommend the following supplemental data collection to compensate for these weaknesses in the NHANES design:

- Additional interview questions about school breakfast participation and usual breakfast patterns
- Cognitive and behavioral testing
- Achievement tests
- Administrative school records

Because parents may be unaware of their child's school breakfast consumption or the availability of SBP at their child's school, we recommend that additional information be collected to verify school breakfast participation. During private interviews, children could be asked additional questions about their frequency of SBP participation during the school year, and about their breakfast eating patterns at home and at school.

Cognitive and behavioral testing to evaluate short-term cognition and academic performance (achievement) could be administered in the mobile examination center during private interviews with children. Subtests of the Wechsler Intelligence Scale for Children, Revised (WISC-R) and the reading and arithmetic sections of the Wide Range Achievement Test, Revised (WRAT-R) were successfully administered in the third NHANES (1988–1994) to 6- to 16-year-old children (National Center for Health Statistics, 1994; and Kramer et al., 1995).
Two subtests of the WISC-R, the Block Design and the Digit Span, were administered to serve as indicators of cognitive functioning. The WRAT-R was used to assess academic performance in reading and mathematics. Other behavioral and cognitive tests relating to school performance and learning could also be considered for supplemental data collection. Decisions about adding survey content will depend on the availability of survey time and the usefulness of linking new components with other interview and health examination components in the survey.

Current NHANES plans do not include the collection of information directly from schools. Because SBP participation and test scores are important outcomes in this study, we recommend the collection of school-level information for these two variables. Schools may keep student records of SBP participation over the course of the year. They also have records of students' test scores that we would have to have to measure the effect of SBP on learning.

We would have to assess the students' level of academic achievement at the point in time of their general data collection, and we also would need test scores for a previous year to assess the students' gain in achievement. It would be necessary to obtain parents' consent to obtain the school records of surveyed students. School records also could provide information on the students' attendance, tardiness, nurse visits, disciplinary events, and grades. School records data are an independent source of relevant information on school behaviors that would supplement the NHANES design.

We considered using the longitudinal followup of a cohort of NHANES children but judged this option too costly to recommend. In a longitudinal design, information on children's SBP participation and related variables, such as school attendance, dietary intake, nutritional status, and health, would be assessed at baseline in the annual NHANES.

With parental consent, a sample of children could be interviewed and examined one to several years later to assess SBP participation, cognitive performance, and academic performance. School records could also be collected to obtain information on academic performance and achievement tests. Learning outcomes, such as academic performance and achievement tests, could be compared between SBP participants and nonparticipants at baseline and at followup.

4. Measurement of key characteristics and outcomes

Participation in the SBP is a key variable for the study of learning and school breakfast. The NHANES includes "usual" participation in the SBP during the school year (that is, the number of times per week the child receives school breakfast, and whether it is free or at a reduced price), based on the parent interview. However, because participation information could be subject to reporting error by the parent, we recommend directly observing students' SBP participation.

Another crosscheck to the parent's report of the child's SBP participation is the 24-hour dietary recall, which captures where breakfast was consumed. During this session, most children aged 6 to 11 years report their dietary intake with the assistance of a parent or guardian. Children 12 and older report their dietary intake alone. Information collected in the dietary recall could be used to identify whether breakfast was consumed in school for children interviewed from Tuesday through Saturday (to reflect Monday through Friday intakes).

One weakness of this approach is that this information would not be available for all children; some children would be interviewed on Sunday and Monday (about Saturday and Sunday intakes), and some would be interviewed during the summer, when school is out of session. Therefore, school day dietary intakes would be unavailable for about 25 to 30 percent of the total sample of school-aged children. However, for children who are not given 24-hour dietary recall interviews, we could compare SBP participation on a sample day with the frequency of weekly SBP participation reported in the parent interview.

The data collected and the desired analysis create various alternative methods of measuring SBP participation. As described earlier, one option would be to define "usual participants" as students who eat a school breakfast on at least 3 of 5 school days. Alternatively, participants and nonparticipants could be distinguished on the basis of whether or not they usually eat a school breakfast five times per week.

Dietary intake is assessed through 24-hour dietary recalls. The current NHANES design calls for one 24-hour recall per child and a second 24-hour recall on a subsample. The second day's intake on a subsample provides information to adjust nutrient intake distributions using statistical software that takes into consideration the day of the week and within- and between-person variability.

Adjusted distributions of nutrient intake could be used to estimate the proportion of SBP participants and nonparticipants who meet dietary recommendations and dietary adequacy. This approach provides information for comparing group dietary data but does not provide a better measure of individual students' usual dietary intake for use in regression analysis.

The 24-hour dietary recall provides information on current dietary intake. Longer-term dietary status is reflected in nutritional biochemical assessments of blood and urine, hematologic determinations, and anthropometric measurements. Nutrition and health outcomes, such as growth, overweight, and iron deficiency anemia, may relate to children's readiness to learn in school.

Other lifestyle and risk behaviors, such as alcohol and drug use, provide information that may relate to poor school performance, missed days of school, suspensions, and expulsions. This type of behavioral information collected in NHANES can be used as control variables when comparing learning outcomes in SBP participants and nonparticipants.

5. Analysis plans

The primary analysis in the NHANES design compares mean differences in dietary, nutrition, health, and (possibly) learning outcomes among SBP participants, nonparticipants, and children who do not eat breakfast. Descriptive analysis would include, but would not be limited to, comparisons of the following outcomes between SBP participants and nonparticipants or nonbreakfast eaters:

- Mean nutrient and food group intakes for breakfast and for the day
- Mean dietary quality and variety score, assessed by the Healthy Eating Index[33]
- The proportion meeting current dietary recommendations or dietary requirements for nutrient adequacy, as defined by the Recommended Dietary Allowances, the Dietary Reference Intakes, and the *Dietary Guidelines for Americans* (U.S. Department of Agriculture and U.S. Department of Health and Human Services, 2000)[34]
- The proportion defined as underweight, at a healthy weight, or overweight based on height and weight measurements and the revised CDC growth charts (Kuczmarski et al., 2000)
- The proportion with iron deficiency anemia or low levels of specific vitamins, based on biochemical test results
- Mean academic test scores and grades (conditional on data availability)

[33]The Healthy Eating Index is a summary measure of diet quality, developed by Kennedy et al. (1995). It measures individuals' food group intake, compliance with dietary guidelines, and assesses the variety of individuals' diets. To the extent that the SBP influences any of these areas of children's diets, then it would also influence the Healthy Eating Index.

[34]Software from Iowa State University can be used to adjust nutrient intake distributions. The program considers the within- and between-person variability in intake and the skewness of nutrient intake distributions (Nusser et al. 1996). However, as described above, regression-adjusted comparisons of the proportion of the groups that meet dietary recommendations cannot be made, because the procedure does not generate estimates of individuals' usual dietary intakes.

- Mean number of missed days of school, tardiness, suspensions, expulsions, and skipped grades (conditional on data availability)

- Mean cognitive test scores (such as from the WISC-R and WRAT-R) and mean scores on composite measures of behavior (conditional on data availability).

To estimate the effects of SBP participation, we would use regression analysis to compare the outcomes of participants and nonparticipants, after controlling for relevant factors that can be measured. Important factors to account for in comparing academic and learning outcomes include food insecurity and poor nutritional status, and factors relating to prenatal nutrition, such as low birth weight and exposure to cigarette smoke. Environmental exposures, such as to lead, revealed by elevated blood lead levels, should also be considered in interpreting the results of cognitive tests and academic performance.

In addition, we would have to consider the important issue of selection bias. We could use several approaches, including IV models or switching regression models to help account for this. These approaches are described in detail in Chapter IV, in the discussion of the ECLS-K design, but also are applicable to the NHANES design.

6. Statistical power

The calculations of the statistical power of the NHANES design focus on the comparison of SBP participation and nonparticipation. We expect about 75 percent of the NHANES sample to attend SBP schools, and the expected total sample size of school-aged children is about 1,600 per year; thus, we expect a sample of about 1,200 children attending SBP schools per year. Using this estimate, we calculated the minimum detectable differences (MDDs) on achievement based on the following assumptions:

- 30 percent SBP participation rate
- No selection bias

- Available prior-year test scores

Average design effect of 1.3 based on NHANES III information or estimated design effect of 2.5 for NHANES 1999-2001[35].

We estimated that a one percentile change in an achievement test score is equivalent to 0.025 of a standard deviation. Therefore, collecting data for one year for 1,200 children (assuming a design effect of 1.3) provides the power to detect a difference of 0.155 standard deviations, or six percentiles, in the achievement test score (see Table III.4). Three years of data collection would

[35]We selected a design effect of 1.3 as a lower bound based on the average design effect in NHANES III. However, the design effect in the current NHANES is likely to be higher. We selected a design effect of 2.5 as a more realistic estimate of the average design effect.

detect a difference of 0.090 standard deviations, or 3.6 percentiles, and six years would detect 0.063 standard deviations, or 2 percentiles.

Table III.4—Minimum detectable differences on achievement test scores

Years of data collection	Sample size	Test score change (SD) for DEFF = 1.3	Test score percentile change for DEFF = 1.3	Test score change (SD) for DEFF = 2.5	Test score percentile change for DEFF = 2.5
1	1,200	0.156	6.2	0.215	8.6
2	2,400	0.11	4.4	0.153	6.1
3	3,600	0.09	3.6	0.125	5.0
4	4,800	0.078	3.1	0.108	4.3
5	6,000	0.07	2.8	0.097	3.9
6	7,200	0.064	2.5	0.088	3.5

MDDs would be higher with a design effect of 2.5. In particular, the MDD would be 0.215 standard deviations with one year of data, 0.125 with three years, and 0.088 with six years.

Thus, with somewhat optimistic assumptions, including the assumption of no selection bias, achieving the target MDD of 0.10 standard deviations would require a minimum of three years of data. With less optimistic assumptions about NHANES design effects, five years of data would be required. Some degree of selection bias likely exists, so it may be necessary to aggregate more years of data collection, or to expand the currently planned annual sample sizes to achieve sufficient statistical power.

7. Design costs

The two cost components of the NHANES-based design are supplemental data collection costs and analysis and reporting costs. Unlike the ECLS-K-based design, the supplemental data collection activities in this design are a critical part of the research effort, so we have generated a cost estimate for the supplemental data collection costs. In generating this estimate, we have made the following assumptions:

- Three years of supplemental data collection would be conducted for NHANES sample members between 6 and 16 years old

- One additional minute of interviewing time would be added to the NHANES parent interview to collect information on usual breakfast patterns and children's school breakfast participation patterns

- Twenty minutes of examination time would be added for trained interviewers to conduct cognitive tests and academic achievement tests, such as the WISC-R and WRAT-R

- Individual school records, including information on academic achievement test scores, would be collected from the sample members' school.

We estimate that the cost of these supplemental data collection activities would be in the range of $2.9 to $3.0 million for a three year period.

The remaining evaluation costs would cover analysis and reporting of the results. These costs would be $1.0 to $1.1 million. Thus, total evaluation costs for the NHANES-based design would fall in the range of $3.9 to $4.1 million.

8. Design summary: strengths and weaknesses

The analytic approach in using data from the NHANES design is to describe and compare mean differences in dietary, nutritional, health, and learning outcomes among SBP participants, SBP nonparticipants, and students who do not eat breakfast. To estimate the effects of SBP participation, we would use regression analysis to compare the outcomes of participants and nonparticipants after controlling for all measurable relevant factors (for example, prenatal exposure to smoke, low birthweight, iron deficiency anemia, and elevated blood lead levels)

One weakness is possible selection bias due to parents' selection of SBP participation. The approach to account for potential selection bias in the NHANES design is similar to that described in the ECLS-K design—using IV models or switching regression models.

The NHANES design offers the advantage of an existing national survey of school-aged children with comprehensive information on family background, SBP participation, dietary intake, nutritional status, and health. The design's primary strength is that it captures many of the important domains needed to link SBP participation and learning (for example, dietary intake and short-term and long-term nutritional status and health.)

The survey also provides a framework for supplemental data collection on short-term cognitive function, school behavior, and academic performance. These domains are needed to fully study the relationship between SBP participation and learning.

Another advantage of the NHANES design is that low-income children are oversampled, which increases the potential sample size of SBP participants available for study. A national sample of about 1,600 school-aged children is drawn from 15 PSUs each year. However, fairly large design effects may occur with few PSUs per year, and the annual sample size is probably too small to support detecting differences in test scores of SBP participants and nonparticipants.

We estimate that, under optimistic assumptions, it would require three years of academic test score data for sufficient power to detect a change in test score of four percentiles (the equivalent of 0.10 standard deviation). Under less optimistic assumptions, more years of data would be required.

Although the NHANES provides a rich database on SBP participation, diet, nutrition, and health variables, the absence of plans to collect information on cognitive functioning and academic performance is a major design weakness. Supplemental data collection would therefore be required.

F. Assessing the Alternative Designs

This chapter presents four alternative designs for estimating the impact of the SBP on learning. Because they use different methodologies and are based on different types of data, each has particular strengths and weaknesses. Given the infeasibility of conducting all four designs, this section identifies the most feasible one. The recommendation has been made with the knowledge that the USBP design already has been funded and is currently being implemented. Thus, the chapter considers whether implementing an alternative design (in addition to the USBP design) would be worthwhile and, if so, which should be implemented. The chosen design—a modified version of the ECLS-K design—is developed more fully in Chapter IV.

1. Summary of alternative designs

The four alternative designs described in this report are summarized in Tables III.5 and III.6. The first table describes the key features of each design. The second one presents the strengths and weaknesses of each design, as well as each one's estimated costs. A key distinction among the four alternatives is the design type; two designs use an experimental approach, and two use a nonexperimental approach.

a. Experimental designs

The USBP design and the SBP applicant design use an experimental approach to examine the relationship between the SBP and learning. In particular, the USBP design involves the random assignment of schools into a treatment group that serves free USBP breakfasts and a control group that serves regular SBP breakfasts. Thus, this design estimates the effect of being in a school that offers the USBP as opposed to being in a school that offers the regular SBP.

The SBP applicant design calls for the random assignment of classrooms in participating schools (that are applying to become SBP schools for the first time) into a treatment group offering USBP breakfasts and a control group offering no breakfast program. This design estimates the effect of being in a classroom that offers free breakfasts as opposed to being in a classroom with no breakfast program.

Table III.5—Summary of evaluation design options

	USBP design	SBP applicant design	ECLS-K design	NHANES design
Design type	Experimental	Experimental	Nonexperimental	Nonexperimental
Intervention and counterfactual	Availability of USBP versus regular SBP	Availability of USBP versus no breakfast program	Participation versus nonparticipation in the regular SBP	Participation versus nonparticipation in the regular SBP
Basic design approach	Experimental, randomly assign schools	Experimental, randomly assign classrooms	Nonexperimental, compare SBP participants and nonparticipants	Nonexperimental, compare SBP participants and nonparticipants
Coverage	Six school districts, not nationally representative	Up to 100 schools, not nationally representative	Nationally representative	Nationally representative
Data collection	Primary	Primary	Secondary	Secondary
Sample	Elementary students in grades 1 through 5 in SY 2000-2001	Elementary students in grades 2 through 5 in SY 2001-2002	Students in kindergarten programs in fall 1998	School-aged children in years 2000, 2001, and 2002
Sample size	144 schools 4,320 students	100 schools 1,600 classroom 16,000 students	16,906 students[a] 866 schools	1,626 students per year 3 years
Key outcomes	USBP and SBP participation student achievement cognitive functioning attendance / tardiness teachers' evaluations dietary intake health status	USBP participation student achievement cognitive functioning attendance / tardiness teachers' evaluations health status	student achievement cognitive functioning social and emotional development attendance / tardiness	student achievement[b] dietary intake nutritional status health status
Statistical power (minimum detectable differences, measured as a percentage of a standard deviation)	16.0 percent	9.5 percent	5.3 percent[c]	9.0 percent[d]

NOTES:
[a] The number of students who continue in the sample to grade 5 may be lower due to sample attrition.
[b] Supplementary data collection is needed to acquire information on student achievement.
[c] This calculation assumes no selection bias. The minimum detectable effect with selection bias (assuming good instrumental variables) is 16 percent of a standard deviation.
[d] This calculation assumes three years of NHANES data, a design effect of 1.3, two years of test score data for each sample member, and no selection bias.

Table III.6—Strengths and weaknesses of evaluation design options

	USBP design	SBP applicant design	ECLS-K design	NHANES design
Strengths				
Design	Experimental design permits rigorous estimation of effect of USBP availability relative to regular SBP availability; 3-year study with longitudinal data	Experimental design permits rigorous estimation of effect of USBP availability relative to no breakfast program		
Implementation	Already funded and being implemented		Basic version of the design is straightforward to implement	
Sample			Nationally representative	Nationally representative
Measurement of SBP participation	Primary data collection methodology permits careful measurement of participation	Primary data collection methodology permits careful measurement of participation		
Outcomes	Full range of outcome measures to be examined	Large number of outcome measures (except for dietary intake) can be examined	Large number of outcome measures relating to academic achievement and cognitive functioning	Good measures of dietary intake, nutritional status, and health status
Weaknesses				
Design			Comparison group design potentially subject to selection bias	Comparison group design potentially subject to selection bias
Implementation		Getting schools to agree to implement random assignment of classrooms likely to be difficult		Supplemental data collection time consuming and potentially difficult
Measurement of SBP participation			Existing ECLS-K participation measures reported by proxy (usually parents) and limited to a single week during year	Existing NHANES participation measures potentially subject to error
Statistical power	Insufficient power to detect impacts on student achievement under experimental design; fallback position is to estimate effects on achievement using nonexperimental comparison group design	If a sufficient number of schools fail to agree to participate in the demonstration, design will have insufficient power to detect impacts on student achievement under experimental design; fall position is to estimate effects using nonexperimental comparison group design	If there is selection bias, design has insufficient power to detect effects of participation	If there is selection bias, design has insufficient power to detect effects of participation
Estimated design cost				
Demonstration	\$6.6 to \$6.9 million	\$0.9 to \$1.0 million	\$0	\$0
Evaluation	\$6.4 million	\$10.7 to \$11.3 million	\$0.4 to \$0.5 million	\$3.9 to \$4.1 million
Total	\$13.0 to \$13.3 million	\$11.6 to \$12.3 million	\$0.4 to \$0.5 million (plus supplemental data collection costs)	\$3.9 to \$4.1 million

[a] Supplementary data collection is needed to acquire information on student achievement.
[b] This calculation assumes no selection bias. The minimum detectable effect with selection bias (assuming good instrumental variables) is 16 percent of a standard deviation.
[c] This calculation assumes three years of NHANES data, a design effect of 1.3, two years of test score data for each sample member, and no selection bias.

The key strength of the two experimental designs is that they produce unbiased estimates of the impacts of the intervention versus the counterfactual. For example, because random assignment of schools in the USBP design ensures that students in treatment schools and students in control schools differ systematically only with respect to their access to the USBP, resulting systematic differences between the groups in outcomes related to learning likely arise from access to the USBP. Given sufficiently large samples, experimental designs have the potential to yield the most rigorous estimates possible of the effects of interventions.

Another advantage of the USBP and SBP applicant designs is that they include primary data collection. They can therefore include a rich set of control and outcome measures specifically tailored for a study of breakfast and learning. The USBP design in particular offers a broad range of outcome measures ranging from academic achievement and cognitive functioning to health and dietary intake. The SBP applicant design also has the potential to include a broad range of outcome measures, with the exception only of dietary intake data.

A unique advantage of the USBP design is that it has been funded and is currently being implemented. Thus, the demonstration and evaluation designs have been fully developed and adequate resources should be available to carry them out. Results from the USBP evaluation will come within a relatively short period of time, given its size.[36]

The two experimental designs do have weaknesses. Although they are experimental, they are not specifically set up to estimate the effects of SBP participation on student outcomes by randomly assigning individuals to participant and nonparticipant categories. Instead, they measure the effects of program *availability.* If they have sufficient statistical power, estimates from the experimental designs of the effect of program availability could be used to generate *indirect* estimates of the effects of participation. Otherwise, they must rely on a nonexperimental approach to estimating these effects.

Unfortunately, the USBP design is unlikely to generate sufficient statistical power to detect the likely effects of USBP availability on test scores.[37] Even with a successful preimplementation survey, the minimum detectable impact on test scores is about 16 percent of a standard deviation, whereas the expected size of the effect is more likely to be closer to 10 percent. Thus, the USBP design will likely have to rely on nonexperimental methods to estimate the effect of breakfast program participation on student achievement in school.

In theory, the SBP applicant design can attain sufficient statistical power to detect effect sizes of participation of 10 percent of a standard deviation. However, this design is likely to be difficult to implement—a major weakness—implying that the target sample sizes may be hard to obtain. In particular, it will be challenging to identify enough SBP applicant schools that will agree to delay full SBP

[36]According to the schedule described in the USBP Evaluation Request for Proposals released by FNS in February 2000, the final draft of the first USBP report covering year 1 of the implementation of the program will be submitted to FNS in June 2002.

[37]The design is likely to have sufficient statistical power to detect estimates of the effects of USBP availability on other outcomes, such as program participation rates and dietary intake (Ponza et al., 1999).

implementation and face the ethical questions involved in randomly assigning classrooms to breakfast program and nonbreakfast program groups.

Finally, implementing either of these experimental designs would be quite costly. When the costs of implementing the demonstration and the costs of conducting the evaluation are summed, each has a cost of roughly $12 to $13 million. Given that the USBP study has been funded and the demonstration and evaluation are under way, the high cost of the SBP applicant design (to address a similar set of questions) becomes an even greater drawback.

b. Nonexperimental designs

The ECLS-K and NHANES designs use a nonexperimental approach to estimating the impact of participation in the SBP (versus the counterfactual of nonparticipation) on learning. Each is based on analysis of a secondary data source that includes information on whether or not students eat school breakfasts and on student outcomes.

One of their strengths is that they are based on nationally representative datasets. In contrast to most previous research and to the proposed experimental designs, results from these analyses would be generalizable to students nationally. In addition, the ECLS-K dataset is very large, with information on about 17,000 students in just under 900 schools.

This sample size implies that the ECLS-K design has a high level of statistical power (given the assumption of no selection bias). The NHANES dataset is smaller; it is expected to provide information every year on more than 1,600 students aged 6 through 18 and will be conducted annually.

The wealth of information available in the ECLS-K and NHANES datasets is another strength of these nonexperimental designs. However, each dataset is limited in an important respect. The ECLS-K dataset includes a rich variety of outcome measures on students' academic performance and background characteristics, but it does not provide information on their dietary intakes or nutritional status. The NHANES dataset has information on dietary intake and a wealth of information on nutritional and health status but lacks information on students' achievement or performance.

Addressing the data limitations of the ECLS-K and NHANES datasets through supplementary data collection would be possible, although it would add substantially to the expense of these designs. As described in Chapter IV, given the expense and the fact that dietary intake information is not central to the objective of the study, we do not recommend supplementing the ECLS-K data with dietary intake information from ECLS-K sample members.

However, we do recommend supplemental ECLS-K data collection to obtain improved measures of SBP participation, as well as relevant student characteristics. For the NHANES design, we recommend supplemental data collection of student achievement and/or cognitive performance measures, including the collection of school records data.

The nonexperimental nature of the ECLS-K and NHANES designs is an important limitation of these approaches to studying the effects of the SBP. Because they are based on observations of students' SBP participation statuses rather than on random assignment of students to participant and nonparticipant statuses, selection bias is possible in estimates of the effect of participation.

In particular, students (or their families) who decide that they should eat a school breakfast may differ in unmeasured ways from nonparticipants. If these unmeasured differences are related to the student outcomes of interest, selection bias will result when the standard regression methods are used to estimate the effects of participation on these outcomes. Although statistical methods, such as IVs, can address selection bias, they also have limitations. Thus, results from a nonexperimental study typically do not carry the same weight as results from a well-designed experimental study.

The costs of the two nonexperimental designs differ dramatically, primarily because of the extensive supplemental data collection necessary in the NHANES-based design. The ECLS-K-based design should cost $400,000 to $500,000, plus the supplemental data collection costs necessary to add questions to the ECLS-K survey instruments. The ECLS-K-based design is the lowest-cost option. Given that the NHANES-based design has many of the same strengths and weaknesses as the ECLS-K design, its cost of roughly $3 million decreases its attractiveness as an option.

2. Choosing the most feasible design

A key factor in the choice of the most feasible design is that the USDA is now implementing the USBP design. This suggests that its results and the results of another recommended design could be used together to give additional perspective on the relationship between the SBP and learning. Although the USBP design is strong, it does have limitations, and we believe pursuing an alternative design that addresses some of these limitations would yield more persuasive evidence on the SBP-learning relationship than would relying solely on the USBP design.

We rule out recommending the SBP applicant design for two main reasons. It is similar to the USBP design in that both use an experimental approach, and both examine the effects of an intervention defined as the availability of free school breakfasts. In addition, both designs call for similar types of data collection activities, variable measurement, and analytic approaches.

Furthermore, neither design is nationally representative, and there are questions about their ability of each to achieve sufficient statistical power. The SBP applicant design does not share the USBP design's strength of being a three-year longitudinal study, and the evaluation costs of the SBP applicant design are greater than the evaluation costs of the USBP design. Thus, the contributions of the SBP applicant design do not justify its costs.

In addition, the SBP applicant design is somewhat risky given its implementation challenges. The success of the approach relies on there being enough SBP applicant schools that serve elementary students and enroll a substantial number of low-income students. Its success also relies on convincing enough of these schools to delay full SBP implementation for a year while classrooms are randomly assigned to a treatment group and a control group.

Schools may not want to deny breakfasts to any students, and they may not want to face the challenges of serving free breakfasts in a limited number of classrooms one year and then implementing the regular SBP the next year.

If the SBP applicant design is not to be chosen as the most feasible, this leaves the two nonexperimental designs to consider. For the reasons we have discussed, we do not consider the NHANES design the most feasible of the two nonexperimental designs. This dataset has information on relatively small samples of school-aged children, so that a large number of years of data would have to be assembled before any analysis that would yield sufficiently precise estimates could be conducted.

More important, the current NHANES data collection plans do not include the collection of information on children's levels of achievement and cognitive performance. For the NHANES design to address the relationship of interest, relatively expensive supplemental data collection activities would have to be conducted.

We consider a modified version of the ECLS-K design to be the most feasible of the four alternatives. The ECLS-K design addresses at least two of the limitations of the USBP design: (1) it is nationally representative; and (2) its large sample size is likely to give its estimates sufficient statistical power, assuming that selection bias is a relatively minor problem.

Furthermore, its rich information on students' achievement and cognitive functioning addresses the key outcomes of interest. Although the ECLS-K design is nonexperimental, the fact that the USBP design will examine the relationship between the SBP and learning, and thus will provide a second methodological perspective on this issue, reduces the impact of any weaknesses.

We recognize that the ECLS-K design would not provide information on students' dietary intake and nutritional status. These outcomes are not necessary to estimate the impact of the SBP on learning, but they may be intermediate outcomes that mediate this impact. In this case, for SBP participation to influence students' learning, it must first influence what they eat and their overall nutritional status. Although the program could influence learning via other routes, this one appears to be at least as important as any other.

To address this limitation, we recommend a modified ECLS-K design in which the analysis of ECLS-K data (described in Chapter IV) is supplemented with analysis of data from the 1988-1994 NHANES III dataset. In addition to information on children's SBP participation, dietary intake, and health and nutrition status (as well as on most of the other variables listed in Table III.3), that dataset also includes information on their academic achievement and cognitive functioning.

In particular, two standardized tests were used to assess intellectual functioning and academic performance: (1) the WISC-R, and (2) the reading and arithmetic sections of the WRAT-R. The entire Wechsler scale can be used to assess children's IQs. Two subtests of the WISC-R, the Block Design and the Digit Span, were administered to serve as indicators of cognitive functioning. They were selected because they are the least culturally sensitive.[38]

[38]Findings for the NHANES III test data have been reported by Kramer et al. (1995). They found that lower income, minority status, and lower educational attainment were independently associated with poorer performance on some subtests. General health status, birth complications, and gender also were predictors of performance for some subtests. This analysis was conducted with data for 1988 through 1991 but did not include SBP participation or dietary or nutritional variables.

To more fully investigate the relationship between the SBP and learning, we recommend an expanded analysis of the full NHANES III dataset for 1988–1994, which would produce a sample size of more than 5,000 school-aged children. The analysis would focus on estimating the relationship between SBP participation and the outcomes of dietary intake and nutritional status. It would then estimate the relationships between dietary intake/nutritional status and achievement/cognitive functioning (as well as the relationship between SBP participation and achievement/cognition). Combined with the original ECLS-K design, this supplemental analysis should provide a fuller understanding of the relationship between the SBP and learning and the pathways through which this relationship arises.

Thus, the most-preferred design includes the following features:

- Use of a nonexperimental basic design approach
- Use of nationally representative data
- Analysis of ECLS-K data to estimate the relationship between SBP participation and student achievement/cognitive functioning
- Possible supplemental ECLS-K data collection to obtain improved information on SBP participation and potential identifying variables for IVs or propensity score models
- Analysis of NHANES III data to estimate the relationships among (1) SBP participation, (2) dietary intake/health and nutrition status, and (3) academic achievement/cognitive functioning.

This design is outlined in detail in the following chapter.