Supporting Statement – Part B for

OMB Control Number 0584-XXXX

Erroneous Payments in Child Care Centers Study (EPICCS)

January, 2017

Chanchalat Chanhatasilpa, Ph.D.

Social Science Research Analyst

Office of Policy Support

Food and Nutrition Service

United States Department of Agriculture

3101 Park Center Drive

Alexandria, Virginia 22302

Phone: 703-457-6791

Email: Chanchalat.Chanhatasilpa@fns.usda.gov

This page is intentionally blank

TABLE OF CONTENTS

TABLES AND FIGURES

LIST OF APPENDICES TO EPICCS OMB SUPPORTING STATEMENT

APPENDIX M: EPICCS BURDEN TABLE

1. Respondent Universe and Sampling Methods

Describe (including a numerical estimate) the potential respondent universe and any sampling or other respondent selection method to be used. Data on the number of entities (e.g., establishments, State and local government units, households, or persons) in the universe covered by the collection and in the corresponding sample are to be provided in tabular form for the universe as a whole and for each of the strata in the proposed sample. Indicate expected response rates for the collection as a whole. If the collection had been conducted previously, include the actual response rate achieved during the last collection.

Respondent Universe

The respondent universe for this Erroneous Payments in Child Care Study (EPICCS)[1] includes sponsored and independent child care centers participating in the Child and Adult Care Food Program (CACFP), as well as the institutions that sponsor the child care centers (where applicable), and the households representing the children served in the centers. During Fiscal Year 2015 CACFP served nearly 3.3 million children daily through about 21,000 sponsoring organizations[2] enrolled at over 64,000 child care centers.[3] Sponsored child care centers include Head Start centers and non-Head Start centers.

---

[1] The household survey component of the study is referred to as National Assessment of Meal Eligibility and Services (NAMES).

[2] Child care providers participate in the CACFP under the umbrella of a sponsoring organization that assumes fiscal responsibility and provides training and monitoring to ensure that its providers comply with all of the CACFP regulations. However, independent child care centers may act as their own sponsor (i.e., self-sponsor) for the CACFP. As such they are included in counts of both sponsors and providers.

[3] All program data are taken from the FNS National Data Bank Data (NDB), last accessed on March 17, 2016 at: http://www.fns.usda.gov/sites/default/files/datastatistics/keydata-december-2015.pdf.pdf

Table B1-1 Respondent Universe, Samples, and Expected Response Rates presents a summary of the universe, samples, and expected response rate for each respondent type, and overall. The numbers of households representing children served in the child care centers are estimated based on the average number of children served. A description of the efforts that have and will be implemented to ensure a high response rate is described in response to Question 3 of this Supporting Statement Part B.

Sampling Overview

A multistage stratified probability sampling design will be used to select a nationally-representative sample of CACFP centers (and their sponsors) and participating households. At the first stage of sampling, a probability sample of 25 States was selected to develop the required State-wide universe lists (sampling frames) of CACFP child care centers. To facilitate subsequent in-person data collection within the sampled States, large geographical clusters referred to as primary sampling units (PSUs) will be defined and selected from the State lists. At the third stage of sampling, a universe list of child care centers including (non-Head Start) sponsored child care centers (SCCCs), independent child care centers (ICCCs), and Head Start centers (HSCs)[4] will be compiled for each of the 50 PSUs, and from these lists stratified samples of centers, and their sponsors, will be selected. Finally, at

[4] Head Start centers will be excluded from analysis of certification errors. However, Head Start centers will be included in analysis of aggregation and meal claiming errors.

the fourth and final stage of sampling, eligible household applications will be randomly selected from each participating center for the household survey.

Table B1-1. Respondent Universe, Samples, and Expected Response Rates by Respondent Category[5]

Respondent	Universe	Initial Sample	Expected Response Rate	Targeted Completed Cases
Child Care Centers:	56,753	474	95%	450
Sponsored Non-Head Start Centers	37,087	158	95%	150
Sponsored Head Start Centers	11,463	158	95%	150
Independent Child Care Centers	8,203	158	95%	150
Sponsors*:	19,186	474	95%	450
Non-Head Start Center Sponsors	6,965	158	95%	150
Head Start Center Sponsors	6,408	158	95%	150
Independent Centers	5,813	158	95%	150
State Agency	49	25	100%	25
Households Representing Child Served:	2,945,758	6,750	80%	5,400
Non-Head Start Centers	2,012,585	2,250	80%	1,800
Head Start Centers	500,333	2,250	80%	1,800
Independent Centers	432,840	2,250	80%	1,800
TOTAL	3,021,746	7,723	82%	6,325

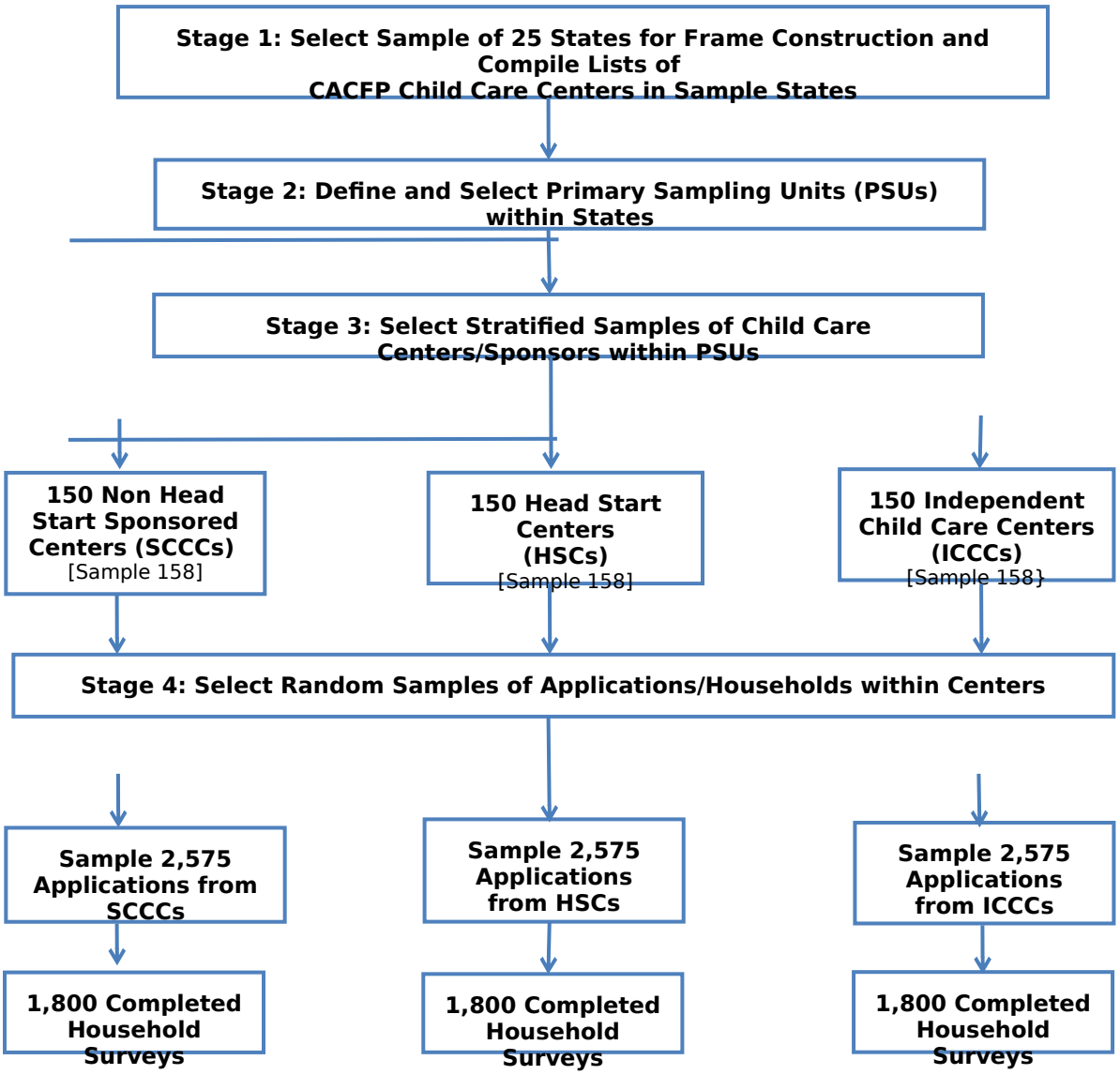
***The study design begins with sampling centers, followed by engaging sponsors for these centers. Thus, the estimates for sponsors are the same as for the centers. However, the number of sponsors will be less because in some cases more than one of a sponsor's centers is sampled. For independent child care centers, the number of sponsors will always be the same as the number of centers.

Figure B1-1 summarizes the general sampling approach for the study. Note that these sample sizes in the table refer to the required numbers of "respondents" (i.e., centers and households that provide useable data for analysis), as well as the number that will be sampled to meet this requirement. Slightly larger numbers will be sampled to offset losses due to nonresponse and ineligibility. We assume that 95 percent of the sampled

[5] All counts are based upon a data file delivered by the FNS COR on April 14, 2015. The file was extracted from the FNS National Data Bank (NDB) on CACFP participation through March 2015. Counts for Sponsored non-Head Start and Independent Child Care Centers are imputed based on proportions of those types of centers.

centers will agree to participate in the study. Although participation of the centers is technically mandatory, we anticipate that some of the sampled centers will be ineligible (e.g., closed) or otherwise unable to participate in the study during the planned data collection period (nonresponse). Thus, we will sample 158 for each type of center to ensure 150 per type of center (for a total of 474 centers).

Figure B1-1. Sampling approach: Selection of child care centers and households



For the household survey we assume an 80% response rate among the eligible sampled applications. Thus, we will select 2,250 eligible applications to obtain 1,800 completed interviews for each type of center across all three data collection periods. In addition, 325 eligible applications (270 for the initial data collection and a total of 55 for the two additional data collections) for each type of center will be selected as a reserve sample to be set aside. This reserve sample will be used if the study is unable to achieve the 80% response rate to ensure the targeted number of completes. Non-response bias analysis will be conducted to address statistical considerations regarding the response rate.

2. Procedures for the Collection of Information

Describe the procedures for the collection of information including:

- **Statistical methodology for stratification and sample selection,**
- **Estimation procedure,**
- **Degree of accuracy needed for the purpose described in the justification,**
- **Unusual problems requiring specialized sampling procedures, and**
- **Any use of periodic (less frequent than annual) data collection cycles to reduce burden.**

Statistical Methodology for Stratification and Sample Selection

A multistage stratified probability sampling design will be used to select a nationally-representative sample of CACFP centers, and their sponsors, and participating households to fulfill the data collection requirements.

Stage 1: Selection of States. A sample of 25 States was selected from the 48 contiguous States and District of Columbia with probabilities

proportionate to an appropriate measure of size[6]. The required sampling frame of States was derived from quarterly administrative reporting to FNS's National Databank (NDB) as available on April 10, 2015. Specifically, reported average daily attendance counts were used as the basis for constructing a State-level aggregate sampling measure of size (MOS). Thus, each State's selection probability was proportional to the State's estimated total daily attendance.

Table B2-1 Sampling Frame for Selecting State Sample illustrates the variation of the measure of size by State based on NDB data for March 2015. Also shown are the nominal probabilities of selection for a 25-State sample. It can be seen that some States are so large that their nominal probabilities are close to or exceed 1.0. These very large States are included in the sample with certainty. Specifically, any State that would otherwise have a probability of selection of 2/3 or greater under probability-proportionate-to-size (PPS) sampling[7] was designated as a certainty State. Application of this rule lead to the designation of the nine certainty States indicated in the table which together account for 59 percent of the total measure of size of all States in the nation. Westat submitted a memorandum to FNS on April 29,

---

[6] Alaska and Hawaii were not included for four reasons: 1) they were not included in the prior APEC studies, 2) both States combined account for less than 0.7 percent of the overall NSLP payments in 2015, 3) being geographically outside the continental United States, their school meal application and delivery systems may be influenced by local factors that are not present in the contiguous United States, and 4) the cost of travel for multiple data collection visits is prohibitive.

[7] Hansen, M., Hurwitz, W., and Madow, W. (1953). Sample Survey Methods and Theory, Vol. I., New York: J. Wiley & Sons.

2015 that provide a summary of the State sampling procedures and results (Appendix J).

Table B2-1. Sampling frame for selecting State sample

State	Measure of size (MOS)*	Nominal Prob. Sel. Prior to Cert. Selection	State Cert. Flag	Prob. of Sel. Non-Cert. State	Exp. No. PSUs in Cert. States	Exp. No. PSUs in Non-Cert. States If Sampled
Alabama	38,442	0.3263		0.5089		1.28
Arizona	30,029	0.2549		0.3975		1.28
Arkansas	70,391	0.5974		0.9318		1.28
California	380,494	3.2292	1		6.46	
Colorado	24,008	0.2038		0.3178		1.28
Connecticut	15,388	0.1306		0.2037		1.28
Delaware	14,347	0.1218		0.1899		1.28
District of Columbia	5,453	0.0463		0.0722		1.28
Florida	231,571	1.9653	1		3.93	
Georgia	124,185	1.0539	1		2.11	1.28
Idaho	9,003	0.0764		0.1192		1.28
Illinois	106,179	0.9011	1		1.80	1.28
Indiana	47,303	0.4014		0.6262		1.28
Iowa	27,779	0.2358		0.3677		1.28
Kansas	23,816	0.2021		0.3153		1.28
Kentucky	48,828	0.4144		0.6463		1.28
Louisiana	53,648	0.4553		0.7101		1.28
Maine	4,413	0.0375		0.0584		1.28
Maryland	46,973	0.3986		0.6218		1.28
Massachusetts	38,175	0.3240		0.5053		1.28
Michigan	57,146	0.4850		0.7564		1.28
Minnesota	27,915	0.2369		0.3695		
Mississippi	38,098	0.3233		0.5043		1.28
Missouri	53,571	0.4546		0.7091		1.28
Montana	8,318	0.0706		0.1101		1.28
Nebraska	26,843	0.2278		0.3553		1.28
Nevada	12,420	0.1054		0.1644		1.28
New Hampshire	7,879	0.0669		0.1043		1.28
New Jersey	61,184	0.5193		0.8099		1.28
New Mexico	23,137	0.1964		0.3063		1.28
New York	224,922	1.9089	1		3.82	
North Carolina	108,665	0.9222	1		1.84	
North Dakota	7,976	0.0677		0.1056		1.28
Ohio	98,742	0.8380	1		1.68	
Oklahoma	42,727	0.3626		0.5656		1.28
Oregon	31,856	0.2704		0.4217		1.28
Pennsylvania	121,002	1.0269	1		2.05	
Rhode Island	6,918	0.0587		0.0916		1.28
South Carolina	27,825	0.2361		0.3683		1.28
South Dakota	9,247	0.0785		0.1224		1.28
Tennessee	57,193	0.4854		0.7571		1.28
Texas	341,283	2.8964	1		5.79	
Utah	18,453	0.1566		0.2443		1.28
Vermont	6,367	0.0540		0.0843		1.28
Virginia	54,845	0.4655		0.7260		1.28
Washington	57,083	0.4844		0.7556		1.28
West Virginia	19,864	0.1686		0.2629		1.28
Wisconsin	47,986	0.4072		0.6352		1.28
Wyoming	5,876	0.0499		0.0778		1.28
Total	**2,945,758**	**25**	**9**	**16**	**29.5**	**20.5****

* Average daily attendance in CACFP centers as of March 2015.

** Total number of PSUs to be sampled from the 16 non-certainty States.

An additional 16 States were selected with probabilities proportionate to the measure of size

from the remaining 41 non-certainty States. The corresponding selection probabilities of the non-certainty States were determined after excluding the certainty States from the State sampling frame, and are shown in the fifth column of the table. Note that the noncertainty States were selected from a sorted list of States (e.g., sorted by Census region and measure of size within region) to induce limited implicit stratification. The last two columns of Table B2-1 show the expected numbers of PSUs to be selected within the State if sampled (see following section).

The general approach for constructing the sampling frames for the selected States will be to

acquire universe lists of CACFP centers, and their sponsors, from State offices, and then to format and integrate the lists into a single electronic dataset created by the Westat sampling team for subsequent sampling purposes.

The type of information available for CACFP child care centers varies from State to State, but

generally all States can provide the core information needed to construct the sample frames such as: (1) type of center, i.e., HSCs, SCCCs, and ICCCs, and (2) average daily attendance. Except for ICCCs, centers can be linked to their sponsoring agency. In addition, we will incorporate sponsor level characteristics in the construction of the frame to ensure representation of

the diversity among sponsors. For this purpose, we will request or derive sponsor characteristics such as (1) number of meals claimed in March 2015; (2) number of centers sponsored; (3) type of centers sponsored; (4) location of sponsor, i.e. rural, urban, or suburban; and (5) proximity to center (including sponsors located in a different county or State).

Stage 2: Sampling PSUs within Selected States. After the selection of States, the process moved to acquisition of universe lists of CACFP centers from the appropriate State offices. The sample of child care centers was restricted to 50 geographical areas referred to as PSUs.[8] PSUs were defined to be either Core Based Statistical Areas[9] (CBSA) or groups of rural counties and will cover the entire State. The PSUs were created by geocoding the

[8] As the name implies, the term primary sampling unit (PSU) is usually reserved for the primary- or first-stage sampling units in a multistage probability sample design. For EPICCS, the States (in particular, the noncertainty States) are the true first-stage sampling units. Since the type of geographical units we propose as sampling units are commonly referred to as PSUs, we will continue to use this term with the understanding that they are actually the second-stage sampling units within the noncertainty States.

[9] The United States Office of Management and Budget defines a core based statistical area as one or more adjacent counties or county-equivalents having at least one urban cluster of at least 10,000 population, plus adjacent territory that has a high degree of social and economic integration with the core as measured by commuting ties. The core based statistical areas comprise the metropolitan statistical areas and the metropolitan statistical areas.

addresses of the child care centers in the State-wide sampling frame to the appropriate CBSA or county within the State. Once the PSUs are formed, a sample of PSUs were selected from each State with probabilities proportionate to an appropriate function of total enrollment counts that reflects the use of the differential sampling rates required to select the various types of centers. For example, since the goal is to sample a fixed number of centers (i.e., 150 centers) for each type of center, the overall (national) sampling rate will be lowest for the predominant type of centers (SCCCs) and much higher for the less common types of centers (ICCCs and HSCs). Since the samples will be weighted appropriately, the resulting samples will be nationally representative of each type of center. While the current approach will not use sponsors as the sampling unit, the centers will be stratified by key sponsor characteristics prior to sampling centers.

A total of 50 PSUs were sampled. On average, about two PSUs will be selected from

each State, but at least one PSU will be selected from every sampled State, and more than two PSUs will be selected from the largest (i.e., certainty) States. As can be seen in the last two columns of Table B2-1, the expected numbers of PSUs in the certainty States was expected to range from one to six, while the average number of PSUs to be sampled in the non-certainty States was expected to be about 1.3 (i.e., either one or two per PSU). The types of geographically-based PSUs proposed for the study are large enough to ensure that a relatively diverse population of centers is available for

sampling at the subsequent stage. Note that the proposed sampling procedures are not designed to represent individual States, but rather when taken as a whole, the entire 50-PSU sample (appropriately weighted) will produce nationally-representative estimates that are subject to measurable sampling errors.

Stage 3: Sampling Centers within PSUs. At the third stage of sampling, centers within the selected PSUs were assigned to strata defined by the three types of centers: (non-Head Start) SCCCs, HSCs, and ICCCs. To ensure a total of 150 responding centers per stratum after anticipated sample losses due to non-cooperation or ineligibility, we sampled 158 centers for each type. The goal will be to obtain an average of nine cooperating child care centers per PSU (three for each type of center). However, the actual numbers could vary depending on the distribution of the different types of centers within the PSU.

Westat submitted two memorandums to FNS, on February 22, 2016 and March 14, 2016 respectively, with a summary of the EPICCS Center Selection (Appendices K and L).

Stage 4: Sampling Households within Centers. The fourth and final stage of sampling is to randomly select households among those who submitted an income eligibility application for meal benefits at each sampled center. For the household survey an 80% response rate is assumed among the eligible sampled applications. Thus, the minimum initial sample size for each type of center is 2,250 eligible applications to obtain 1,800 completed

interviews[10]. Given the uncertainty about response rates, the samples will be randomly assigned to a “primary sample” for initial contact and recruitment, and a smaller “reserve sample” will also be selected to be used if needed. Data collectors will be trained to conduct sampling on site using a programmed algorithm that will be accessed from their computer. Two subsequent rounds of household survey data collection will also be done, at four month intervals, to allow for sampling of applications submitted after September 2016. Thus, an additional 1,125 households will be sampled from two groups of households: those submitting applications November 2016 to February 2017 and another group between March and June 2017 to obtain an additional 900 interviews for this comparison.

To maintain roughly equal workloads across the sampled centers, a fixed number of children will be sampled from each center whenever possible. However, if the actual enrollment of the center is more than twice the enrollment (measure of size) reported on the sampling frame, the consequences of selecting the specified fixed number of households will be to increase unequal weighting design effects. To minimize the impact of large sampling weights, the sampling program will flag instances in which the actual enrollment of the center exceeds the sampling measure of size by a factor of two or more. In these cases, it will be desirable to attenuate the impact on sample precision by selecting more than the specified fixed

[10] A reserve sample of 325 application will be drawn and set aside for use if needed.

number of children. Such cases will be flagged for review by the statistician, and a decision will be made as to the number of households to be sampled.

Estimation Procedures

Weighting. Descriptive statistics derived from the study will be weighted to reflect the relevant probabilities of selection specified under the complex multistage sample design and to compensate for differential rates of nonresponse.

For analyses that are at the application level, weighting the samples of applications (i.e., children and associated households) will involve the following steps. First, a "base" weight, w_{ki}^{base}, equal to the reciprocal of the overall probability of selection will be assigned to each sampled application. This probability is the product of four components: (1) the probability of selecting the State; (2) the probability of selecting the primary sampling unit (PSU) within the State; (3) the probability of selecting the center within the PSU; and (4) the probability of selecting the application/household within the center.

Next, the base weights will be adjusted for nonresponse by applying appropriate nonresponse adjustment factors to the application-level base weights. The required nonresponse adjustment factors will be computed within subsets referred to as "weighting cells." The weighting cells will be defined by type of center and other characteristics found to be correlated with survey response propensity. Logistic regression or Chi Square Automatic Interaction Detector (CHAID) will be used to develop the weighting cells,

using information available for the sampled center as well as characteristics of the sampled application as potential predictors of nonresponse. These may include variables such as geography, level of urbanization, type and size of provider, age and sex of children in household, certification status, and other relevant household characteristics. Thus, the final weight for an application for which a household interview was completed in a specified weighting cell k will be computed as $w_{ki}^{final} = R_k w_{ki}^{base}$, where R_k is the inverse of the weighted response rate of sampled applications in weighting cell k.

For analyses that are at the meal level, the data will be aggregated to the center level. Hence, the number of reimbursable or non-reimbursable meals for a given center must be determined. However, it may not always be possible to independently observe all meals at a given center[11]. In such cases, a random sample of tables (or child trays for cafeteria style) will be selected for observation. For example, suppose for simplicity that there are N serving tables in a particular center, and that 2 are randomly selected for observation. In this case, the reported meal claiming data for the center will be weighted by a factor of N/2, to account for the fact that data were reported for only 2 of the N locations. The sampling weight for the meal claiming data will provide unbiased estimates of the total numbers of meals and reimbursements at each center subject to sampling error. As with the application-level data, post stratification of the sample-based estimates can

[11] Section C1.2.3 of Appendix C1 provides a summary of the two components of the requirements for reimbursable meals and how the meal observations will collect the necessary data.

be employed to align the sample-weighted counts to known numbers of meals (or reimbursements) if such national data are available.

Variance Estimation. In addition to the full sample weights described above, a series of replicate weights will be created and attached to each data record for variance estimation. Replication methods provide a relatively simple and robust approach to estimating sampling variances for complex survey data.[12] Under the replication approach, balanced repeated replicates (BRR) will be formed by deleting selected cases from the full sample and adjusting the base weights of the retained cases accordingly. The entire weighting process developed for the full sample will then be applied separately to each replicate resulting in a series of replicate weights. The replicate weights can be imported into variance estimation software (e.g., SAS, SUDAAN, WesVar, STATA) to calculate standard errors of the survey-based estimates and conduct significance tests. In addition to the replicate weights, stratum and unit codes will also be provided in the data files to permit calculation of standard errors using Taylor series approximations if desired. Note that while replication and Taylor series methods often produce similar results, replication has some advantages in reflecting statistical adjustments used in weighting such as nonresponse and post stratification.

Degree of Accuracy (and Levels of Precision) Needed for the Purpose Described in this Justification

[12] Rust, K.F., and Rao, J. N. K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5: 283-310.

The precision requirements specified for the study are that (a) overall national estimates have a ±0.05 level of precision with 95 percent confidence, and (b) subgroup estimates have a ±0.05 level of precision with 90 percent confidence. The latter requirement means that the sample for each subgroup must be of similar size and sufficiently large to meet the specified precision requirements. Thus, the total study sample size is driven primarily by the subgroup precision goals. The three key subgroups of interest are the three types of centers: SCCCs, HSCs, and ICCCs. Based on certain assumptions about design effects and the underlying variability of payment errors, we have estimated that a subgroup sample size of 150 centers and 12 households per center will be adequate to ensure that the specified precision goals can be met for both subgroup and national estimates when the underlying error rates are sufficiently low (less than 7.5 percent).

Two types of statistical measures related to improper payments will be derived from the study results: (1) a dollar-weighted error rate defined to be the ratio of the total dollar amount in error to total payments (reimbursements), and (2) the percent of cases (e.g., household-level records) with errors of a specified type. In order to derive the sample sizes required to meet the stated precision goals, we made a number of informed assumptions about the expected percent of cases with errors of a particular type, the corresponding dollar-weighted error rate, and the expected underlying relative standard deviation (RSD) of the amounts in error per

case. The latter parameter reflects the variability of the error amounts among those cases with an incorrect payment. For sample planning purposes, we assumed RSD = 1, which implies that the standard deviation of the error amounts is equal to the average error amount among those cases with an error. This is a relatively large standard deviation which is used in the calculations to guard against understating the sample size needed to meet the specified precision targets. We also made the following assumptions about the design effects associated with the stratified multistage probability sample: an unequal weighting design effect of 1.2 for subgroup estimates and 1.5 for national estimates, a between-center within PSU intraclass correlation of τ_1 = 0.02, and a within-center intraclass correlation of τ_2 = 0.10 (see equation 2.23 of Skinner, Holt, and Smith, 1989). The values of the intraclass correlations used in the precision calculations are speculative, but believed to be reasonable.

Note that under the specified sample design, an additional clustering effect is introduced by the first-stage selection of States. However, as indicated in Table B2-1 Sampling Frame for Selecting State Sample, the majority of PSUs will come from the certainty States for which there is no clustering effect since the States are essentially sampling strata rather than sampled clusters. For the sampled noncertainty States, an average of a =1.5 PSUs will be selected per State. The clustering effect for these States is given approximately by the formula $D_1 = 1+(\boldsymbol{a}-1)\delta_1$, where δ_1 = the intraclass correlation between PSUs within the noncertainty States. Since PSUs within

States generally vary widely with respect to many demographic and socio-economic characteristics, we expect δ_1 to be fairly small. For example, for PSUs defined in terms of counties, Hansen, Hurwitz, and Madow (1953)[13] state that "it would not be unusual for the value of δ_1 to be as small as 0.01, or even considerably smaller." Thus, if δ_1 = 0.01, the clustering effect due to sampling the non-certainty States would only be 1.005, and the overall clustering effect for the *entire* sample of States would be considerably smaller (i.e., closer to 1.0 and therefore negligible).[14] Thus, a term reflecting the clustering effect due to sampling States is not explicitly included in calculations of the expected levels of summarized below, but instead can be assumed to be incorporated with the between-center within PSU intraclass correlation τ_1.

Estimates of the expected levels of precision under the specified sample design for centers
serving breakfast or lunch are summarized in Table B2-2 Expected Levels of Precision for Estimates of Errors for Breakfast or Lunch for total error rates

[13] Hansen, M., Hurwitz, W., and Madow, W. (1953). Sample Survey Methods and Theory, Vol. I., New York: J. Wiley & Sons.

[14] While Hansen, Hurwitz, and Madow refer to counties, the PSUs we propose are CBSAs in metropolitan areas and groups of counties in rural areas. Thus, a PSU will often consist of two or more different counties, and as such will be more internally heterogeneous than individual counties. In other words, the magnitude of intraclass correlations cited in Hansen, Hurwitz, and Madow will likely overstate the intraclass correlations for the types of PSUs we are planning for the study, and thus will have an even smaller impact on clustering design effects.

(encompassing all meal types) ranging from 0.05 to 0.15. There are currently no available data on improper payments to child care centers. However, APEC II (which examined errors in the NSLP and SBP programs) can be used as a point of reference. Findings from APEC II suggest that total error rates may range from 0.05 to 0.10. Given that the NSLP and SBP programs are much more complex meal programs in which there is more opportunity for error, we postulate that the error rates in child care centers will be lower, or at most equal to, the APEC II error rates. The current estimates include error rates at a low (0.05), medium (0.10), and high estimate (0.15), with the high estimate being very unlikely and mainly shown for illustration. The sample sizes shown in the table reflect the actual number of centers that agreed to participate, and is based on the assumption that 100% of the centers serve breakfast, lunch or both. Shown in the table are 95- and 90-percent confidence bounds for both estimated error rates and the percent (prevalence) of cases with an error. As can be seen in the last two columns of the table, the levels of precision for prevalence of errors are expected to be appreciably better (much lower) than the ±0.05 (±5 percentage points) requirement. The corresponding confidence bounds for estimates of dollar-weighted error rates are larger, but generally meet or exceed the precision goals specified for national estimates, and either meet or are close to meeting the precision goals for the three type-of-center subgroups. Even with an error rate as high at 10%, the ±2.5% requirement on 90% confidence

bounds (as specified in Appendix C to Circular No. A-123 dated October, 2014) will be achieved.

Table B2-2. Expected levels of precision for estimates of errors for breakfast or lunch

Assumed error rate		Subgroup	Number of centers (resp.) *	Sample size (children)	Total design effect [2]	Precision of estimated error rate [1]		Precision of estimated error prevalence [5]	
Prop. cases with error	Error rate[1]					90% conf.	95% conf.	90% conf.	95% conf.
0.05	0.050	SCCC	155	1,550	2.78	±2.2%	±2.6%	±1.5%	±1.8%
(low estimate)		Head start	162	1,620	2.82	±2.1%	±2.6%	±1.5%	±1.8%
		ICCC	160	1,600	2.81	±2.2%	±2.6%	±1.5%	±1.8%
		Total	**477**	**4,770**	**5.41**	**±1.7%**	**±2.1%**	**±1.2%**	**±1.4%**
0.10	0.100	SCCC	155	1,550	2.78	±3.0%	±3.6%	±2.1%	±2.5%
(expected rate)		Head start	162	1,620	2.82	±3.0%	±3.6%	±2.1%	±2.5%
		ICCC	160	1,600	2.81	±3.0%	±3.6%	±2.1%	±2.5%
		Total	**477**	**4,770**	**5.41**	**±2.4%**	**±2.9%**	**±1.7%**	**±2.0%**
0.15	0.150	SCCC	155	1,550	2.78	±3.7%	±4.4%	±2.5%	±3.0%
(high estimate)		Head start	162	1,620	2.82	±3.6%	±4.3%	±2.4%	±2.9%
		ICCC	160	1,600	2.81	±3.6%	±4.3%	±2.5%	±2.9%
		Total	**477**	**4,770**	**5.41**	**±2.9%**	**±3.5%**	**±2.0%**	**±2.4%**

[1] Error rate defined to be ratio of dollar amount of errors to total payments.

[2] Assumes unequal weighting design effect of 1.2 for subgroups and 1.5 for the overall (total) sample, an intraclass correlation between centers within PSUs of 0.02, and an intraclass correlation between students within centers of 0.10.

[3] Applies to subgroups.

[4] Applies to overall national estimates.

[5] Percent of cases in error.

Nonresponse Bias Analysis. Although efforts will be made to achieve as high a response rate as practicable with the available resources, nontrivial nonresponse losses are likely to occur. OMB requires that a nonresponse bias analysis (NRBA) be conducted if the overall response rate falls below 80

percent.[15] In this case, a nonresponse bias analysis will be conducted to assess the impact of nonresponse on the survey estimates and the effectiveness of the weight adjustments to dampen potential nonresponse biases. The types of analyses to be conducted to evaluate nonresponse will include:

- Comparing characteristics of non-respondents (or the total sample) to those of respondents using information available for both non-respondents and respondents;
- Modeling response propensity using multivariate analyses;
- Evaluating differences found in comparisons between survey respondents and comparable data from extant outside sources;
- Evaluating if significant differences exist between households submitting applications in the initial four month time period and those submitting in the remaining eight months of the study year;
- Comparing cases completed at different levels of data collection effort (e.g., cases completed with limited follow-up compared to those requiring considerable follow-up);
- Comparing weighted estimates of characteristics available for both respondents and non-respondents using unadjusted (base) weights versus nonresponse-adjusted weights; and
- Comparing weighted survey estimates using unadjusted (base) weights versus nonresponse-adjusted weights.

[15] http://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf

Unusual Problems Requiring Specialized Sampling Procedures

There are no unusual problems that require specialized sampling procedures for the sampling of States, PSUs, centers, and/or households. However, it is important to note that the largest portion of the household survey sample will be drawn from applications submitted between July and September of the study year. As noted, two more rounds of smaller household sample selection will also be done, at four-month intervals, to ensure that applications submitted after September are included in sampling.

Any use of periodic (less frequent than annual) data collection cycles to reduce burden

The data collection effort, with multiple visits to sponsors/centers, is planned to be done one time only during the 2016-2017 program year.

Pre-Data Collection Procedures

First, as part of the administrative component of participation in CACFP, the appropriate State agencies will be asked to provide administrative data to develop the sample frame. Subsequently, sampled sponsors and centers will be notified about the study and their selection and their participation confirmed.

FNS regional offices will notify the appropriate State Child Nutrition offices that their State has been selected for participation in the study. Following this FNS contact, the study team conducted follow-up contacts (emails and calls) with the 25 sampled States requesting that they submit

the administrative data to develop the sampling frame. The request included instructions on the submission of the administrative data file (Appendices B1 – B4). Once the sample of centers and sponsors is selected, the States will be asked to inform the sponsors (Appendices B5 – B7). Next, research team will contact sponsors and centers to confirm participation (Appendices B8 – B15).

Recruitment Procedures

Recruitment of parents/guardians for the household survey will be conducted by the data collection team. Recruitment activities will be conducted by mail, email, and telephone. Recruitment materials will include letters, brochures, and call scripts (Appendices B16 – B21).

Data Collection Procedures

EPICCS data collection will include the following:

1) Pre-visit Interviews with sponsoring organizations and child care centers (see Appendix C2) by phone
2) On-site data collection visits to the sponsoring organization or child care center for abstraction of income eligibility applications and sampling for the household sample (see Appendix C4)
3) In-person NAMES survey (see Appendix C25)
4) On-site data collection visits to child care centers to obtain/scan center records for later data entry:
    a) Administrative records including center enrollment and attendance information (see Appendices C3 and C8)

b) Meal count and claiming data (see Appendices C9- C10)

5) Observation of meal preparation and service at the child care centers (see Appendices C5 – C7)
6) Sponsor survey (mail survey) to collect details on operational functioning (see
   Appendices C13 –C15)
7) Request for meal count and claiming data records that sponsors submitted to the State (See Appendices C11 – C12) for data abstraction
8) Request for additional administrative records from centers several months after onsite data collection (see Appendices C16 and C17)
9) Request for meal count and claiming data from State-level agency (see Appendix C18)

Sponsoring organizations and child care centers will be contacted by phone to complete the pre-visit interviews to obtain information to prepare for the data collection visits, and schedule the data collection visits. Sponsoring organizations will also be asked to complete the survey as a mail survey (see Appendices C14 and C15), which will provide relevant sponsor characteristics for later analyses and comparisons.

During the first round of on-site data collection visits, the data collector’s first priority is to access application records to conduct sampling for the household survey. Sampled households will be mailed a recruitment package, followed by a recruitment call. Once a parent or guardian agrees to

participate and has a scheduled appointment, a reminder letter, as well as the income worksheet (see Appendix C21 and C23), will be sent via mail, or email if provided. At the appointed time the data collectors will travel to sampled households to conduct the in-person household survey as a computer assisted personal interview (CAPI) with each participant. Once the parent/guardian has completed the survey, they will receive the incentive and incentives received form (Appendix C29).

Also during the first round of data collection, the data collector will be able to move forward with conducting data collection visits at the child care centers when it is practical and efficient. All data will be entered into a computer-based system and securely transmitted to Westat on a daily basis. Westat is also prepared to request meal count and claiming data from the sponsor organization separately when data collectors are unable to obtain it during the on-site visit. All State-level data will be requested directly from the State, for electronic submission directly to Westat's home office team. All data are received and stored in a secure, password protected computer-based system at Westat.

3. Methods to Maximize the Response Rates and to Deal With Nonresponse

Describe methods to maximize response rates and to deal with issues of non-response. The accuracy and reliability of information collected must be shown to be adequate for intended uses. For collections based on sampling, a special justification must be provided for any collection that will not yield "reliable" data that can be generalized to the universe studied.

As reported in Table B1-1 Respondent Universe, Samples, and Expected Response Rates, a 95 percent response rate is anticipated from centers and sponsors, and 80 percent response rate from parents/guardians for the household survey. We expect a 95% response rate from sponsors and centers for five primary reasons: 1) participation in this study is part of Healthy, Hunger-Free Kids Act (HHFKA) requirements (Appendix A8); 2) the sponsors will be encouraged to participate by their State agency, who have already agreed to participate; 3) the centers will be encouraged to participate by their sponsors; 4) the Study has a letter of endorsement from the National CACFP Forum, Childcare Food Program Roundtable, and the National CACFP Sponsors Association (Appendix B8), which was shared with the sponsors and centers; and 5) we have worked with corporate sponsors and secured their support and agreement to participate. If response rates fall below 80 percent, the reserve sample will be used to achieve the targeted number of completes. Further, nonresponse bias analysis will be conducted as described in the response to question B2. Several steps will be taken to maximize response rates including:

a) Use of endorsement letters from CACFP associations and organizations.
b) Use of study materials that represent the study effectively and clearly.
c) The recruitment packet sent to sampled households will include information about the importance of the study, privacy protections, incentive information, and the fact that meal benefits will not be affected by study participation.

d) Use of experienced senior-level staff to recruit reluctant sponsors, centers, and/or parents/guardians.

e) Use of qualified, trained and experienced individuals to conduct recruitment and household interviews.

f) Use of data collection methods for the NAMES survey that shifts the burden to the data collector (i.e. computer assisted person interview), minimizing data entry and writing for the NAMES survey respondents.

g) Use of methods for sponsor and center data collection that shifts the burden of data entry to the data collector (data abstraction from records, minimizing the level of effort for the sponsor and center staff to simply providing the data collector access to records).

h) Maximizing the use of existing data that sponsors and centers are required to have in their records as per CACFP requirements.

i) Use of techniques for gaining cooperation, practicing sensitivity, and providing specialized support to sponsors/centers as needed to minimize refusals

j) Use of a modest incentive payment to parents/guardians who complete the household survey.

4. Test of Procedures or Methods to be Undertaken

Describe any tests of procedures or methods to be undertaken. Testing is encouraged as an effective means of refining collections of information to minimize burden and improve utility. Tests must be approved if they call for answers to identical questions from 10 or more respondents. A proposed test or set of tests may be submitted for approval separately or in combination with the main collection of information.

The household survey instrument is comparable to the survey instrument that was approved by OMB for the APEC-II study in 2012 (Approval # 0584–0530 NSLP/SBP Access, Participation, Eligibility, and Certification Study, Expiration 08/31/2015). A cognitive pretest of the survey instrument was also conducted for this Study in April 2015 with 8 respondents. The pretest was conducted as a formal 60 minute in-depth semi-structured cognitive interview. Cognitive test respondents were recruited from a pool of parents with a child enrolled at local CACFP participating child care center. A second methods test was conducted in July 2015 to test strategies to improve response rates for income documentation among households who complete the household survey. These strategies included informing parents in advance of the income documentation requirements, providing a list of acceptable forms of documentation, providing respondents with a short income worksheet before the survey to help them prepare for the income questions, and providing a base incentive for completing the survey, as well as an additional incentive for taking the time to gather and provide the income documentation.

Data collection instruments and procedures for records abstraction and meal observations were all developed based on previously approved APEC-II procedures, with minor modifications. These were developed in consultation with professionals with experience in CACFP participating child care centers. Our research staff also consulted with local sponsor and center directors (2 sponsor directors and 2 center directors) to get their input on the data

collection approach and forms. The data collection forms and procedures were modified based on their feedback. It is important to note that only trained data collectors will use these data collection forms, the staff from the centers will not be asked to complete these forms.

5. Individuals Consulted on Statistical Aspects of the Design & Individuals Collecting and/or Analyzing Data

Provide the name and telephone number of individuals consulted on statistical aspects of the design and the name of the agency unit, contractor(s), grantee(s), or other person(s) who will actually collect and/or analyze the information for the agency.

Table B5-1 Individuals Consulted on Data Collection or Analysis presents a summary of staff consulted on statistical aspects of the design. These staff will be responsible for the collection and analysis of the study's data.

Table B5.1. Individuals Consulted on Data Collection or Analysis

Westat Staff (contractor)	Title	Phone Number
Adam Chu, Ph.D.	Senior Statistician	301-251-4326
Mustafa Karakus, Ph.D.	Senior Economist	301-294-2874
Rene Gonin, Ph.D.	Senior Statistician	301-517-8084
Laurie May, Ph.D.	Vice President, Associate Director	301-517-4076
Roline Milfort, Ph.D., PMP	Senior Study Director	301-251-8229
Subcontractor/ Consultants		
Ted Macaluso, Ph.D.	President, Ted Macaluso, LLC	571-214-9658
Richard Mantovani, Ph.D.	Independent Statistical Consultant	204-506-4112
Frederic Glantz, Ph.D.	President, Kokopelli Associates	505-983-0785
FNS Staff		
Chanchalat Chanhatasilpa, PhD	Social Science Research Analyst	703-305-2115
NASS Staff		
Lori Harper	Mathematical Statistician	202-690-0694