Information Collection Request

New

The Childcare Survey of Activity and Wellness (C-SAW) Pilot Study

Supporting Statement: Part B

Program Official/Contact

Carrie Dooyema, MPH, MSN, RN

Division of Nutrition, Physical Activity and Obesity

Atlanta, GA

Telephone: (770) 488-5039

E-mail: igb7@cdc.gov

Fax: (770)-488-5369

July 1, 2019

Table of Contents

PART B:COLLECTION OF INFORMATION EMPLOYING STATISTICAL METHODS

B.1 Respondent Universe and Sampling Methods

Respondent Universe: The Childcare Survey of Activity and Wellness (C-SAW) pilot study aims to obtain reliable estimates and a representative sample of early care and education centers at the state level. Thus, it requires the construction of a sampling frame that covers the target population of legally operating childcare centers serving children aged 0 to 5 years old, but not yet in kindergarten. This includes pre-kindergarten, but excludes programs specifically for children with special needs that are part of a public school. Hereafter, we call such centers simply centers, unless otherwise specified.

The survey universe is restricted to all centers operating in pilot states selected purposely for the study. Pilot states have been selected to provide a variety of different licensing structures and environments for operation of early childhood childcare centers so that we can learn and address various issues that may be encountered if the surveillance system were to be scaled up in size over time. Based on this consideration, we decided to use four important operational characteristics by which states are categorized: 1) States participating in CDC's cooperative agreements such as the Pediatric Obesity Mini COIIN collaborative (Collective Improvement and Innovation Network); 2) States with high level of license exemption; 3) States that have a high number of tribal nations; and 4) Large-sized states with more than one million early age children (age 0-5). States suitably categorized by these characteristics are shown in Table B-1.

State health departments and other partner organizations that have been working closely with CDC and others to support improved obesity prevention policies and practices through partnerships and quality improvement activities within their state are good candidates for participation in this pilot. We selected Ohio from the COIIN states; Ohio also represents the Midwest region. Furthermore, Ohio does not allow license exemption at all, and thus, preparing the sampling frame will be easier than doing it for states with a high level of license exemption. As the backup, we chose Wisconsin, which comes from the same region. Wisconsin also has a low level of license exemption.

From the high-level license exemption states, we selected Florida, with Texas as the backup. Both of these are in the South region. We believe that it will be more difficult to build the sampling frame for a state with a high level of license exemption because they may not have a centralized list of license-exempt centers.

States that have a high number of tribal nations may not have their centers included in centralized state childcare licensing lists because tribal reservations may have oversight for childcare centers in their jurisdiction. We selected New Mexico from this group with Oklahoma as the backup.

There are four large states with more than one million children age 0-5 based on the 2016 population estimates. We selected New York from the Northeast region. As the backup, we selected California, which belongs to the West region.

Table B-1. States in four sampling categories

COIIN States	Selection[1]	License Exemption States	Selection[1]	Tribal Nation States	Selection[1]	Large States	Selection[1]
Arkansas		Alabama		Alaska		New York	X
California		Florida	X	Arizona		California	*
Indiana		Illinois		Montana		Texas	
Iowa		New Jersey		New Mexico	X	Florida	
Kentucky		Oregon		North Dakota			
Louisiana		Texas	*	Oklahoma	*		
Missouri				South Dakota			
North Dakota							

Ohio	X						
Oklahoma							
Oregon							
Pennsylvania							
Wisconsin	*						

[1] Note: "X" indicates primary selection, and "*" indicates backup selection.

The final proposed pilot states will include the 4 states listed above (Ohio, Florida, New Mexico, and New York). CDC will reach out to their early childhood contacts in each of these states and ask them for their cooperation and endorsement of the survey being conducted in their state. If for some reason the state declines the invitation to participate, backup states will be selected to replace those noted above. The survey universe will then be defined by all legally-operating early childhood childcare centers in each of the final set of four collaborating pilot states.

Sampling Methods: For Ohio, Florida, New Mexico and New York, we expect that a comprehensive list frame will be available by request. In Ohio and Florida a specific state agency maintains a centralized list for all child care centers within the state. In New Mexico the University of New Mexico Division of Continuing Education maintains a list for all child care centers within New Mexico in cooperation with the state the Office of Child Development, Early Childhood Division of the located within the New Mexico Children, Youth and Families Department. We will use these lists to develop the sampling frame for each state.

For New York State, there is not such a comprehensive list of all eligible centers. All licensed centers, including centers in the tribal areas, are listed on available databases, but for license-exempt centers, there is no such database available. A mixed frame approach that uses both list frame and area frame will be used to address this issue. Those license-exempt centers not in the list frame will be covered by an area frame that consists of counties, from which a random sample of counties will be selected with probabilities proportional to the number of centers that appear on the list for each county. Then working with the Child Care Resource and Referral Agency, who typically maintain a comprehensive up to date list of all centers in their county, a list of centers

for each sampled county will be compiled, including those that may be located in school buildings. From this list, the centers that already appear in the sampling frame will be removed to avoid over coverage by the area frame.

Each sampling frame will be stratified by the following center characteristics depending on their availability:

- Whether it is operated by a public or religious or other organization (Variable A);
- Whether it is located in an urban or rural area (Variable B); and
- Whether or not it is license-exempt (Variable C).[1]

For the sampling frames for Ohio and New York, the stratification variable (C) is irrelevant because the frames contain only licensed centers. However, for Ohio an alternative variable, that is, participation in the quality rating system called Step Up To Quality (SUTQ) will be used, and among those who do participate they can be further stratified by their rating in SUTQ.

For the pilot study, there is no particular group to be oversampled. Therefore, an equal probability sampling method will be used as much as possible. For sampling frames, after allocating the sample size proportionally to strata defined above an equal probability systematic sample will be selected using some site characteristics available in the frame as sort variables that are not used for stratification. We do not know yet what characteristics will be available in the frame but will make an effort to use them to enhance the sampling efficiency.

As mentioned above for New York, a sample of counties will be selected from which a list of non-licensed centers will be created. Given the large differences in number of children in different counties in New York State, it is important to select counties with probabilities proportional to size (PPS). The measure of size will be the number of licensed centers for the county in the list stratum. Each of these non-licensed centers will have a weight (equal to the inverse of the chance of selection for that county)

[1] License exempt status is different from legally operating status. In some states, centers run by religious organizations are license-exempt but not legally operated.

associated with it, defining how many unlicensed centers it represents. The overall New York sample will be allocated to the list and area frame strata according to the number of centers represented in each stratum.

From the list stratum a simple random selection of centers will be drawn, but from the area stratum centers with larger weights will have a higher likelihood of being included in the sample. This will provide for a more precise estimator than a non-PPS method.

The sampling procedure is summarized in the following table:

Table B-2. Sample designs for the pilot states

State	Sampling Frame	Stratification	Sampling Method
Ohio	List	Variables A, B, and SUTQ	Equal probability systematic
Florida	List	Variables A, B, and C	Equal probability systematic
New Mexico	List	Variables A, B, and C	Equal probability systematic
New York	List	Variables A and B	Equal probability systematic
	Area (counties)	N/A	PPS of counties and centers

Our goal for precision is to achieve a standard error of 5 percent to estimate a population proportion of 50 percent.

To determine the sample size to meet the precision goal, we assume a design effect (DEFF) of 1.5. The DEFF comes from two sources, the first is the unequal probabilities of selection in New York for unlicensed centers. The main source of the DEFF, however, is from variation in the final weights used for estimation due to nonresponse weighting adjustments that cause some loss of sampling efficiency.

We also assumed a response rate of 55 percent. This rate is based on a review of previous childcare center survey response rates. Recent rates have been below 50 percent for state nutrition and physical activity surveys of licensed childcare providers, where the survey was not limited to providers participating in federal programs such as USDA's Child and Adult Care Food Program (CACFP) or Head Start[2]. The response rate is also based on the assumption that license-exempt childcare centers may be less likely to respond to government agency sponsored surveys and thus would lower the overall expected response rate to the C-SAW study. However, we believe that Westat, our contractor for this pilot, working closely with state childcare organizations in each state, will be able to achieve a higher response rate than previous state surveys referenced in the footnotes. Westat has decades of history executing surveys in various settings including childcare centers, schools, and other facilities. This includes surveys of childcare centers participating in CACFP, and studies of childcare sponsor and provider characteristics for the USDA's Food and Nutrition Service. Additionally, Westat also carried out the National Assessment of Educational Progress for National Center for Education Statistics.

Given that the assumed response rate does not meet OMB's standard of 80 percent, it will be necessary to do a nonresponse bias analysis by comparing responding and nonresponding centers based on information available from the frame of legally operating centers.

Incorporating the finite population correction (FPC),[3] the sample size determination formula is given by the following equation:

---

[2] Two statewide healthy eating and physical activity surveillance efforts most similar to the planned C-SAW were conducted in Washington State and Minnesota and Wisconsin in 2013 and 2010, respectively. The *Washington State Survey of Nutrition and Physical Activity in Child Care* was conducted with a statewide census of licensed childcare centers and family childcare homes, endorsed by state childcare agencies, and administered primarily as an online survey with hard copy surveys mailed to centers without email addresses. This survey achieved a 46% response rate among licensed childcare centers. (http://depts.washington.edu/uwcphn/work/ece/waccsurvey.shtml). Similarly, *Supporting Healthy Food and Activity Environments in Child Care Settings* was conducted with a random stratified sample of licensed childcare centers and family childcare homes in both Minnesota and Wisconsin. It was conducted by universities and endorsed by state childcare agencies, and allowed respondents the choice of online, mail or telephone response. This two-state survey achieved a 48% response rate for all childcare centers and a 30% response rate for tribally administered childcare centers. (https://www.healthdisparities.umn.edu/research-studies/supporting-early-child-care-food-and-activity-environments)

$$n = \frac{DP(1-P)}{\sigma^2 + P(1-P)/N},$$

where n is the target respondent sample size, D is the design effect, P is the population proportion for which the precision requirement is set, σ is the precision level in terms of the standard error, and N is the population size. In our case, $P = 0.5$, $D = 1.5$, and $\sigma = 0.05$ for States except New York; for New York we assume $D = 2.0$ because the PPS method for the non-licensed frame will increase the design effect.

It appears that the available sampling frames for the pilot states have information about the type of childcare centers so that screening of eligible centers for the study will not be necessary. Nevertheless, it would be a good idea to include some screening questions to filter out ineligible respondents (e.g., centers without study-eligible children). Furthermore, any sampling frame contains some out of business centers, and therefore, it is necessary to take this into account in the sample size calculation. An ineligibility rate of 10 percent is assumed to counter this issue.

To use the formula, we need N, which is not available and was estimated.[4] Table B-3 shows the estimated universe size and the target respondent sample size and field sample size (with FPC incorporated) for the primary and backup pilot states. Note that the sample size is not very sensitive to the state universe size (N). Therefore, it is not required to use a very accurate estimate for N.

---

[3] The finite population correction (FPC) is a correction to the variance of a statistic obtained from a sample selected from a finite population rather than an infinite population. If the sample is selected by SRS, the correction factor is $1-f$, where f is the sampling fraction. When the population size is very large compared to the sample size, f is small and the FPC is near one.

[4] State's population size of eligible centers was estimated using information garnered from the NSECE study reporting that there were 129,000 center-based ECE programs in the country in 2012. Using 2016 child population data to estimate the state-level number of centers, the total national number of 129,000 is prorated to each state's population size of children of ages 0-5. This methodology is based on the assumption that the number of eligible centers is highly correlated with the population size of children of age 0-5 across the states.

Table B-3. Estimated number of eligible ECE centers and sample sizes for pilot states

Sample Class	Pilot State	Main/ Backup	Number of Children of Ages 0-5	Estimated Number of ECE Childcare Centers (*N*)	Target Sample Size (*n*)	Field Sample Size
Tribal nation	New Mexico	Main	156,168	841	134	271
	Oklahoma	Backup	320,424	1,725	142	286
License-exempt	Florida	Main	1,353,098	7,285	148	299
	Texas	Backup	2,424,168	13,052	149	301
COIIN	Ohio	Main	836,763	4,505	147	296
	Wisconsin	Backup	405,912	2,185	143	290
Large States	New York	Main	1,395,060	7,511	198	400
	California	Backup	2,996,726	16,135	149	301
Total		Main	3,741,089	20,142	627	1,266
		Backup	6,147,230	33,097	583	1,178

B.2 Procedures for the Collection of Information

Procedures for the collection of information addressed below include:

- Statistical method for stratification and sample selection
- Estimation procedure
- Unusual problems requiring specialized sampling procedures
- Any use of periodic (less frequent than annual) data/information collection cycles to reduce burden
- Data collection procedures

As described in Section B.1, the basic sample design is the stratified design with simple random sampling of centers for each pilot state, but in New York unlicensed centers are sampled using probabilities proportional to size. To produce unbiased estimates, we will weigh the data for each state first by the base weight, which is the inverse of the sampling probability. The base weight will be then adjusted for nonresponse – explained in detail later. Survey estimates will be produced using the nonresponse-adjusted weights. A jackknife variance estimator will be developed to estimate the precision of weighted estimates.

This is a one-time data collection effort with no unusual problems that require specialized sampling procedures. The contractor, Westat, will conduct sample frame development activities, sampling, data collection, data cleaning, weighting, and analysis.

Data Collection: A sample of approximately 1,200 ECE centers across four states will participate in this one-time data collection effort.

Each center director will receive a recruitment letter (Attachment 4) introducing the survey, explaining its objectives and the importance of their participation. It also identifies state organizations endorsing the study, provides instructions for completing the survey, including a URL and personalized identification number (PIN) for Internet access; gives confidentiality assurances, identifies the incentive, and lists information on how to seek assistance.

The C-SAW questionnaire (Attachment 3) will collect information on the centers practices and policies across seven topic areas including:

- Nutrition information including whether meals/snacks are served at the center, who provides the meals and snacks, where meals and snacks are prepared and who prepares them, who develops the menus, information on the frequency of certain foods and beverages such as fruits and vegetables, fried, sweet and salty foods and beverages such as juice and milk. The final questions in the nutrition section cover miscellaneous topics such as the food

security, if parents are able to bring in food from home for special occasions, farm to ECE activities, and if there is a space for mothers to breastfeed.

- Physical Activity (PA) information such as the amount of time each day that is provided for physical activity, opportunities for outdoor playtime and adult-led physical activity, tummy time for infants, and policies around time infants are placed in swings or seats.

- Screen time information such as the amount of daily screen time for children in the ECE center.

- Training for staff around nutrition, physical activity, child development and stress management

- Activities undertaken by the ECE center to improve nutrition and PA offerings

- Wellness topics such as activities to enhance child development and child behavior

- Role of the person (administrative, teaching, or both) who completed the questionnaire. At the end of the survey, ECE directors will also be given the option to upload last week's center menu. The menu will be used as a quality control check for the nutrition information responses.

We anticipate that most responses will be submitted through the web. The survey will take approximately 30 minutes to complete, and respondents may complete it over multiple sessions. The web survey will be hosted on a secure Westat server. A URL to the website will be included in the recruitment letter. Each survey will start on a screen that requires respondents to enter their assigned PIN code. PIN entry will be required each time a respondent accesses their survey online, and partially completed surveys will resume on the last screen completed.

If an ECE center director prefers, he or she may call the toll-free number for the study and request a paper version of the questionnaire.

Approximately two weeks after the initial recruitment letter is mailed, all sampled providers will receive a postcard (Attachment 5a) reminding them to complete the survey, if they have not already done so.

Approximately two weeks after the postcard reminder, nonrespondents will be sent another letter (Attachment 5b) along with a hardcopy of the questionnaire.

B.3 Methods to Maximize Response Rates and Deal with No Response

Based on similar state surveys, we estimate a 55 percent response rate. To reduce the potential for nonresponse bias, we will implement a wide array of strategies, as described below.

The instrument (Attachment 3) has been designed in a user-friendly manner that minimizes complicated skip patterns and encourages participation and survey completion. The instrument was pretested with 8 ECE directors and revised based on comments from the pretest participants. Respondents will have the option of completing the survey via the web or on paper. Providing different modes of data collection allows the respondent to select the approach with which they are most comfortable, thus increasing the likelihood they will participate. The contractor is also offering a toll-free help line and dedicated email account, providing an opportunity to immediately reach out for assistance, when desired.

Sampled early childhood education centers will receive a recruitment letter, addressed to “Childcare Director”, on study letterhead, (Attachment 4). The letter introduces the study, explains the study objectives and the importance of their participation, identifies state organizations endorsing the study, provides instructions for completing the survey, gives confidentiality assurances, identifies the incentive, and lists information on how to seek assistance.

Approximately two weeks after the initial mailing, a reminder postcard will be sent to all sampled providers (Attachment 5a). The postcard reminds them to complete the survey, if they have not already done so. It stresses the importance of participating. It also includes the toll-free help line number as well as the dedicated email account information, should they need assistance. For those providers with an email address on the sample file, we will also send a reminder notice with a link to the URL for their survey.

Approximately two weeks after the postcard mailing, nonrespondents will be sent another letter (Attachment 5b) along with a hardcopy of the questionnaire and a postage-paid business return envelope.

A non-cash incentive valued at $20 will be offered as a token of appreciation for the provider's participation.

In spite of the use of extensive refusal avoidance procedures, participant refusal is unavoidable. We expect a complex nonresponse pattern. When the response rate is low it is important to perform good nonresponse adjustments. We will consider using the propensity score method that first calculates the response propensity through logistic regression or running a nonparametric tree-building algorithm with available auxiliary variables in the frame (1, 2, 3). We do not expect many auxiliary variables in the sampling frame, which will be compiled from state's databases. However, we can use Census-tract community-level demographic variables from the 5-year American Community Survey (ACS) data through geocoding of the center address.[5] Since there are a considerable number of auxiliary variables available from the frame and the ACS data for propensity score modeling, we prefer using a nonparametric tree algorithm to logistic regression, which can be quite unwieldy when there are many auxiliary variables.

Once the response propensity scores are estimated, we will examine the scores to determine how to use the scores. If they do not vary too much, we can use them to directly calculate the nonresponse weight adjustment factor by the inverse of the score. However, if they vary too much, we need to control the variability of the adjusted weights to maintain a reasonable level of DEFF. One common way of doing it is to use weighting classes by forming them by grouping of respondents with similar propensity scores. If we take this route, we will use quintiles of the propensity scores to form between 5 to 10 weighting classes.[6] A good criterion to decide which method is to use

---

[5] There is a geocoding software package called ArcGIS that facilitates linking an address to a Census tract (or other geographic entities such as ZIP code), by which the Census tract-level ACS data can be retrieved.

[6] The literature recommends creation of 5 to 10 weighting classes. The smaller number of weighting classes can control the weight variability (and DEFF) better but the bias can be non-negligible, whereas the weight variability and bias move in opposite direction as the number of weighting classes increases. We will start with 10 weighting classes but may have to reduce the number if DEFF exceeds 1.5 with 10 weighting classes.

the assumed DEFF. If the first method produces a DEFF that is greater than the assumed value of 1.5, then we will use the second method (i.e., the weighting class method) to reduce the DEFF.

B.4 Tests of Procedures or Methods to be Undertaken

There is not a questionnaire currently being used that collects the information CDC wants to collect. For that reason, CDC funded a separate task in 2015 to bring together a panel of experts to brainstorm the best set of question topics and draft a questionnaire. That draft questionnaire was then updated and refined under this task by reviewing current literature and then sharing an updated version with a series of experts (see Section B.5). The resulting draft instrument and recruitment letter were then cognitively tested with 8 center directors who were selected to ensure a mixture of centers with diverse characteristics including; location (large metro, smaller metro, and non-metro); enrollment size; age range of children served; and type of administering agency (owner-administered centers, public and private centers, and a tribally administered center). The final version of the materials reflect the input provided by these respondents. With over 500 completes expected in the pilot survey, this task will produce a fully tested questionnaire, that will then be modified based on data quality findings to produce a high quality questionnaire for use in an ongoing surveillance system.

In many states, there are not any comprehensive lists of ECE centers; lists may exist of licensed centers, but the target population for this surveillance system are all legally operating centers. Thus, a major purpose of this pilot study is to determine the level of effort required to compile comprehensive lists for a state. We anticipate that states with many centers, those allowing many types of license exemption, and those with multiple tribal nations, are likely to take more effort to develop comprehensive lists than states that have a history of working closely with CDC on ECE childhood nutrition, physical activity, and obesity. For that reason, our pilot test will include one state that has been working with CDC on these topics, one large state, one with many types of exemptions, and one with multiple tribal nations.

B.5 Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

Westat, 1600 Research Blvd., Rockville MD 20850, will conduct the pilot study on behalf of CDC's Division on Nutrition, Physical Activity, and Obesity. Dr. David Marker will serve as Westat's Project Director and Mary Dingwall will oversee data collection. The lead statistician at Westat will be Dr. Hyunshik Lee. Their contact information is shown in the table below.

Name	Telephone Number	Email
David Marker	(301) 251-4398	DavidMarker@Westat.com
Mary Dingwall	(301) 738-3583	MaryDingwall@Westat.com
Hyunshik Lee	(301) 610-5112	HyunshikLee@Westat.com

REFERENCES

1. Kalton, G., and Flores-Cervantes, I., (2003) "Weighting Methods," Journal of Official Statistics, Vol. 19, No. 2, 81-97.
2. Little, R.J.A., (1986). “Survey Nonresponse Adjustments for Estimates of Means,” International Statistical Review, 54, 139-157.
3. Little, R.J.A. and Rubin, D.B. (2002). Statistical Analysis With Missing Data (2nd Ed.). New York: Wiley.