



An Employee-Owned  
Research Corporation

## Memo

**Date:** April 29, 2015

**To:** Chanchalat Chanhathasilpa

**From:** Roline Milfort, Adam Chu, Mamadou Diallo

**Subject:** Selection of States for the Erroneous Payments in Child Care Centers Study (EPICCS)

**Attachment:** SAS specifications for selecting the State sample

### 1. Introduction

The statistical team has completed the selection of States for participation in EPICCS. The 25 selected States (in alphabetical order) are:

1. Alabama	6. Florida	11. Louisiana	16. New York	21. Tennessee
2. Arkansas	7. Georgia	12. Maryland	17. North Carolina	22. Texas
3. California	8. Idaho	13. Minnesota	18. Ohio	23. Virginia
4. Colorado	9. Illinois	14. Missouri	19. Pennsylvania	24. Washington
5. Connecticut	10. Kansas	15. New Jersey	20. South Carolina	25. Wisconsin

This memo describes the process and result of the first-stage sample selection of the 25 States listed above. The sample was selected using systematic sampling with probability-proportionate-to-size (PPS) with the sampling measure of size (MOS) defined by the average daily attendance (ADA) of children enrolled in the child care centers for each state. The required ADA counts were obtained from the FNS National Data Bank (NDB) as of April 4, 2014.

### 2. Selection of the State Sample

The sample of 25 States was selected from the frame of 49 States and the District of Columbia shown in Table 1. Some States were included in the sample with certainty because of their large ADA relative to the other States. An iterative process was used to determine the certainty States. The iterative process consisted of the following steps:

## APPENDIX J: EPICCS STATE SAMPLE SELECTION MEMO

- a. Compute the initial probability of selecting a State  $p_s^{(1)}$  and treat as certainty any State with a probability of selection larger than 1, where  $p_s^{(1)} = \frac{25 * M_s}{\sum_{s=1}^{49} M_s}$ , and  $M_s$  is the ADA for State  $s$ .
- b. Remove the certainty States obtained at the previous step and compute the probability of selecting a State using the remaining states. Again treat any state with probability larger than 1 as a certainty.
- c. Cycle through the previous steps until no computed probability of selection is greater than 1.

It took three cycles to identify the 9 certainty States, which were: California, Florida, Georgia, Illinois, New York, North Carolina, Ohio, Pennsylvania, and Texas. As per the study plan, all States with initial probability of selection greater than 2/3 are treated as “certainties”.

The 16 non-certainty States in the sample were randomly selected among the 40 remaining States using systematic PPS sampling. States with higher probability of selection had a greater chance of selection. The randomness of State selection guarantees that the sampling theory can be applied and therefore measures of uncertainty (variance, confidence intervals, etc.) can be computed using the probability of selection and other information.

Region and ADA were used to order the frame with an alternating ordering for the variable ADA. Specifically, the variable ADA was sorted descending for the first region (Northeast), then ascending for the second region (Midwest), then descending for the third region (South), and finally ascending for the last region (West). The rationale for the alternating ordering is to have successive States selected within/across regions to be similar in terms of measure of size (ADA). If it later becomes necessary to pair States for variance estimation this alternating ordering may help to create more homogenous pairs of States (in terms of ADA).

Table 2 shows the selected States with the corresponding sampling probabilities and weights. Attachment 1 provides the program specifications for selecting the sample.

Table 1 State sampling frame for the EPICCS

State	Region	ADA	Final Probability	Final Certainty
Connecticut	1. NORTHEAST	15,388	0.204	0
Maine	1. NORTHEAST	4,413	0.058	0
Massachusetts	1. NORTHEAST	38,175	0.505	0
New Hampshire	1. NORTHEAST	7,879	0.104	0
New Jersey	1. NORTHEAST	61,184	0.810	0
New York	1. NORTHEAST	224,922	1.000	1
Pennsylvania	1. NORTHEAST	121,002	1.000	1
Rhode Island	1. NORTHEAST	6,918	0.092	0
Vermont	1. NORTHEAST	6,367	0.084	0
Illinois	2. MIDWEST	106,179	1.000	1
Indiana	2. MIDWEST	47,303	0.626	0
Iowa	2. MIDWEST	27,779	0.368	0
Kansas	2. MIDWEST	23,816	0.315	0
Michigan	2. MIDWEST	57,146	0.756	0
Minnesota	2. MIDWEST	27,915	0.370	0
Missouri	2. MIDWEST	53,571	0.709	0
Nebraska	2. MIDWEST	26,843	0.355	0
North Dakota	2. MIDWEST	7,976	0.106	0
Ohio	2. MIDWEST	98,742	1.000	1
South Dakota	2. MIDWEST	9,247	0.122	0
Wisconsin	2. MIDWEST	47,986	0.635	0
Alabama	3. SOUTH	38,442	0.509	0
Arkansas	3. SOUTH	70,391	0.932	0
Delaware	3. SOUTH	14,347	0.190	0
District of Columbia	3. SOUTH	5,453	0.072	0
Florida	3. SOUTH	231,571	1.000	1
Georgia	3. SOUTH	124,185	1.000	1
Kentucky	3. SOUTH	48,828	0.646	0
Louisiana	3. SOUTH	53,648	0.710	0
Maryland	3. SOUTH	46,973	0.622	0
Mississippi	3. SOUTH	38,098	0.504	0
North Carolina	3. SOUTH	108,665	1.000	1
Oklahoma	3. SOUTH	42,727	0.566	0
South Carolina	3. SOUTH	27,825	0.368	0
Tennessee	3. SOUTH	57,193	0.757	0
Texas	3. SOUTH	341,283	1.000	1
Virginia	3. SOUTH	54,845	0.726	0
West Virginia	3. SOUTH	19,864	0.263	0
Arizona	4. WEST	30,029	0.398	0
California	4. WEST	380,494	1.000	1
Colorado	4. WEST	24,008	0.318	0
Idaho	4. WEST	9,003	0.119	0
Montana	4. WEST	8,318	0.110	0
Nevada	4. WEST	12,420	0.164	0
New Mexico	4. WEST	23,137	0.306	0
Oregon	4. WEST	31,856	0.422	0
Utah	4. WEST	18,453	0.244	0
Washington	4. WEST	57,083	0.756	0
Wyoming	4. WEST	5,876	0.078	0
<b>Total</b>		<b>2,945,758</b>		<b>9</b>

Table 2 State sample for the EPICCS

STATE	REGION	ADA	SAMPLINGPROB	SAMPLINGWEIGHT	CERTAINTY
New Jersey	1. NORTHEAST	61184	0.810	1.235	0
Connecticut	1. NORTHEAST	15388	0.204	4.909	0
Kansas	2. MIDWEST	23816	0.315	3.172	0
Minnesota	2. MIDWEST	27915	0.370	2.706	0
Wisconsin	2. MIDWEST	47986	0.635	1.574	0
Missouri	2. MIDWEST	53571	0.709	1.41	0
Arkansas	3. SOUTH	70391	0.932	1.073	0
Tennessee	3. SOUTH	57193	0.757	1.321	0
Virginia	3. SOUTH	54845	0.726	1.377	0
Louisiana	3. SOUTH	53648	0.710	1.408	0
Maryland	3. SOUTH	46973	0.622	1.608	0
Alabama	3. SOUTH	38442	0.509	1.965	0
South Carolina	3. SOUTH	27825	0.368	2.715	0
Idaho	4. WEST	9003	0.119	8.391	0
Colorado	4. WEST	24008	0.318	3.147	0
Washington	4. WEST	57083	0.756	1.323	0
New York	1. NORTHEAST	224922	1	1	1
Pennsylvania	1. NORTHEAST	121002	1	1	1
Illinois	2. MIDWEST	106179	1	1	1
Ohio	2. MIDWEST	98742	1	1	1
Florida	3. SOUTH	231571	1	1	1
Georgia	3. SOUTH	124185	1	1	1
North Carolina	3. SOUTH	108665	1	1	1
Texas	3. SOUTH	341283	1	1	1
California	4. WEST	380494	1	1	1

## Attachment 1

### SAS specifications for selecting the State sample

The code to select the State sample was generated using SAS/BASE/STAT software, Version 9.3 of the SAS System for Windows.

First convert the input Excel file into a SAS file and call it EPICCS\_STATES\_FRAME. This file has 49 records, one for each of the 48 contiguous States and the District of Columbia, and three variables, STATE, REGION and ADA. Split the file into two parts, the first part called EPICCS\_STATES\_CERT contains the certainty States and the second part called EPICCS\_STATES\_NCERT contains the noncertainty States.

Using the file EPICCS\_STATES\_NCERT created above, run the following SAS procedure.

```
PROC SURVEYSELECT DATA=EPICCS_STATES_NCERT METHOD=PPS_SYS  
                SAMPSIZE=16 OUT=EPICCS_SAMPLE_NCERT SEED=521347;  
    ID STATE REGION;  
    SIZE ADA;  
RUN;
```

Append the two files EPICCS\_STATES\_CERT and EPICCS\_SAMPLE\_NCERT. For the records from the former file, set the sampling probabilities and weights to 1.