



PSU and Cluster Weights User Guide

This document describes how to use PSU and cluster weights that have been developed to allow construction of PSU- and cluster-level aggregate measures using the NSECE provider and household survey data. As we describe in this document, such aggregate measures can then be merged on to household survey data for use in analyses that exploit the provider cluster design of the NSECE sample, for example, to better understand the interaction of ECE supply and demand in local areas.

An informative illustration of the use of clusters in NSECE analysis is available in the brief, “Households’ Geographic Access to Center-based ECE.”

<http://www.acf.hhs.gov/programs/opre/research/project/national-survey-of-early-care-and-education-nsece-2010-2014>

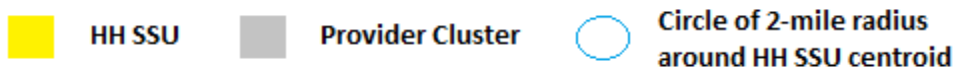
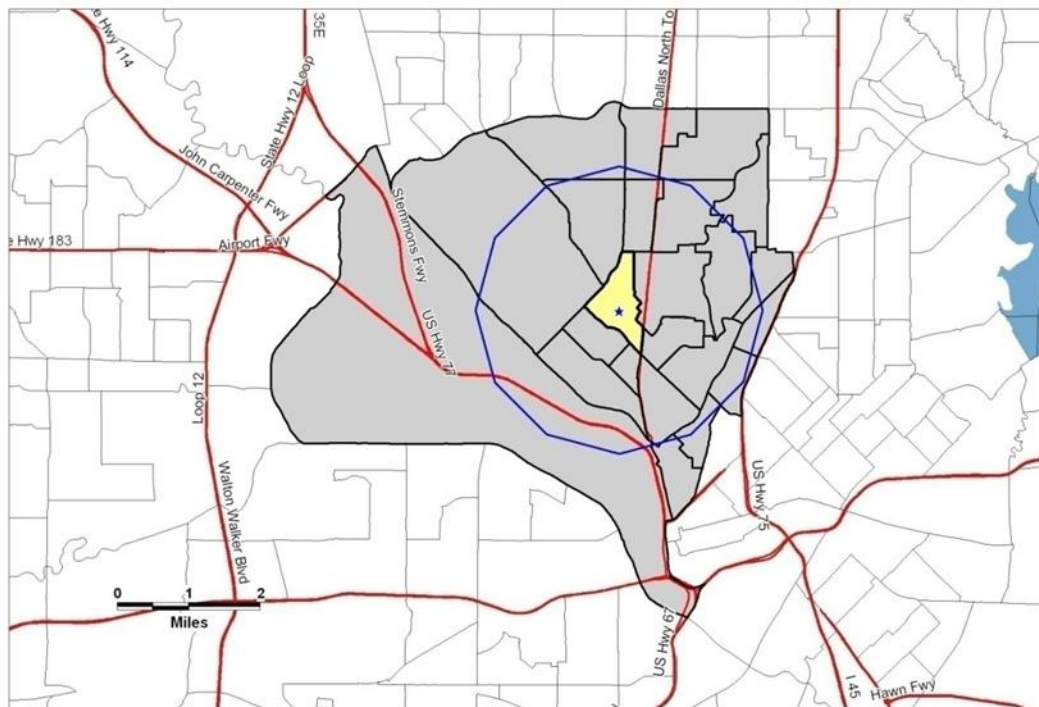
Last Updated: May 31, 2018

1 Background Information on PSU and Cluster

1.1 Recap of the Sampling Methodology

Figure 1 Example of Hypothetical HH SSU vs. provider cluster

Dallas County 0006.01



The NSECE sample has four major components: the household (HH) sample, the unlisted home-based provider (unlisted HB) sample, the listed home-based provider (listed HB) sample, and the center-based provider (CB) sample.

The HH sample was selected in three stages. The first stage selected primary sampling units (PSUs), which are single counties or clusters of contiguous counties. The second stage selected secondary sampling units (HH SSUs), which are single tracts or clusters of contiguous tracts, within selected PSUs. In Figure 1 above, HH SSU is the yellow tract. The third stage selected addresses within selected HH SSUs. The unlisted HB sample was the same as the HH sample. Unlisted HBs were identified by screening the addresses selected for the HH sample.

The listed HB sample and the CB sample use the same selected PSUs as the HH and unlisted HB sample used. However, instead of using HH SSUs, provider clusters are created. A two mile radius circle (blue circle in Figure 1) is drawn around the centroid of each selected HH SSU (* in Figure 1). The provider cluster is defined to be the cluster of all tracts either within or intersecting this circle. In Figure 1 above, the provider cluster is the grey area. Provider clusters could overlap, i.e. a tract could belong to both the provider cluster centered on HH SSU 1 and the provider cluster centered on HH SSU 2. After the provider clusters were defined, listed HB and CB providers within the provider clusters were pulled together and sampled. For more details on sampling, please see the NSECE Summary Data Collection and Methodology Report (OPRE Report # 2013-46).

1.2 Overlapping Provider Clusters

Provider clusters could overlap with each other. One situation is when the centroids of two selected HH SSUs are within four miles of each other. This will cause the two mile radius circles of the two selected HH SSUs to overlap; therefore, the provider clusters centered on these HH SSUs would also overlap. Another situation is when a tract intersects the two mile radius circles of two selected HH SSUs. This tract will then by definition belong to both provider clusters. Any provider completes within such a tract will also belong to both provider clusters. A third situation is when the two-mile radius circles do not overlap (the centroids are more than four miles apart), but both circles cross into the same census tract, which lies in both provider clusters.

A provider can also “belong” to multiple PSUs. This is when the tract in which the provider is located belongs to multiple provider clusters and one of the provider clusters is centered on a HH SSU selected from PSU A, while another of the provider clusters is centered on a HH SSU selected from PSU B. The provider is located only in one of the two PSUs, but it falls into provider clusters for more than one cluster.

2 Definition of Terms

PSU: The primary sampling unit for the HH and listed provider samples. Single counties or several contiguous counties.

HH SSU: The secondary sampling unit for the HH sample. Single tracts or several contiguous tracts. Yellow area in Figure 1.

Provider cluster: A cluster of contiguous tracts centered on the HH SSU. Used in sampling listed providers. Gray area plus yellow area in Figure 1.

PCLUSTER_ID: Identification number for HH SSUs and provider clusters. Can be used to link provider clusters with the HH SSU it is centered on. *Please note that for disclosure protection reasons, NSECE Level 2 researchers receive a simulated cluster id variable, L2_PCLUSTER_FALSE_ID. This variable effectively replaces PCLUSTER_ID in all discussions below. The NSECE team will link cluster or PSU aggregate variables from provider files to the Household file following the tutorial in section 4 below*

when requested by a research team. The actual values of PCLUSTER_ID for households or providers are not available to researchers.

PSU_ID: Identification number for PSUs.

3 The PSU and Cluster Weights

Unlike the main interview weights, which are used to create national estimates, the PSU weights and the cluster weights create summary statistics for sub-national areas. Specifically, the PSU weights are used to create estimates for each of the 219 selected PSUs. Similarly, the cluster weights are used to produce estimates for each of the 755 selected HH SSUs or provider clusters.

3.1 HH PSU and Cluster Weights

3.1.1 Weights Data File

The HH PSU and cluster weights data file will have the following variables:

HH_METH_CASEID

PSU_ID

PCLUSTER_ID

HH_WEIGHT_PSU: the HH PSU weight

HH_WEIGHT_PCLUSTER: the HH cluster weight

Each HH_METH_CASEID is linked to a unique PCLUSTER_ID and a unique PSU_ID. There should be no duplicates of HH_METH_CASEID.

3.1.2 Description of the PSU Weight

The HH PSU weight sums to the number of households with children under age 13 in each selected PSU. It can be used to create estimates that describe

1. Households with children under age 13 in each selected PSU. This is when the weights are used to create summary statistics for variables concerning the whole household.
2. All children under age 13 in households with children under age 13 in each selected PSU. This is when the weights are used to create summary statistics for variables that collected information on all children under age 13 in the household.
3. All household members (HHMs) age 13 or older in households with children under age 13 in each selected PSU. This is when the weight is used to create summary statistics for variables that collected information on all HHMs age 13 or older in the household.

4. All providers used by households with children under age 13 in each selected PSU. This is when the weights are used to create summary statistics for variables that collected information on all the providers used by the household.

3.1.3 Application of the PSU Weight

First, do a one-to-one merge of the HH PSU and cluster weights data file and the HH main data file by HH_METH_CASEID. The resulting data file should have 11629 rows. Sort the merged data file by PSU_ID. Set the weight variable to HH_WEIGHT_PSU. There are no strata or cluster variables. Generate the estimates for each PSU separately.

Sample SAS code for numeric variable of interest.

```
proc surveymeans data=HH_MERGED_DATA;  
  var VARIABLE_OF_INTEREST;  
  weight HH_WEIGHT_PSU;  
  by PSU_ID;  
run;
```

3.1.4 Description of the Cluster Weight

The HH cluster weight sums to the number of households with children under age 13 in each selected HH SSU. It can be used to create estimates that describe, for each selected HH SSU:

1. Households with children under age 13.
2. All children under age 13 in households with children under age 13 and over in households with children under 13.
3. All household members (HHMs) age 13 or older in households with children under age 13 in each selected HH SSU.
4. All ECE arrangements used by households with children under age 13.

3.1.5 Application of the Cluster Weight

First, do a one-to-one merge of the HH PSU and cluster weights data file and the HH main data file by HH_METH_CASEID. The resulting data file should have 11629 rows. Sort the merged data file by PCLUSTER_ID. Set the weight variable to HH_WEIGHT_PCLUSTER. There are no strata or cluster variables. Generate the estimates for each HH SSU separately.

Sample SAS code for numeric variable of interest.

```
proc surveymeans data=HH_MERGED_DATA;  
  var VARIABLE_OF_INTEREST;  
  weight HH_WEIGHT_PCLUSTER;  
  by PCLUSTER_ID;
```

```
run;
```

3.2 Unlisted HB PSU and Cluster Weights

3.2.1 Weights Data File

The unlisted HB PSU and cluster weights data file will have the following variables:

HB_METH_CASEID

PSU_ID

PCLUSTER_ID

HB_WEIGHT_PSU: the HB PSU weight

HB_WEIGHT_PCLUSTER: the HB cluster weight

Each HB_METH_CASEID is linked to a unique PCLUSTER_ID and a unique PSU_ID. There should be no duplicates of HB_METH_CASEID.

3.2.2 Description of the PSU Weight

The unlisted HB PSU weight sums to the number of unlisted HB providers in each selected PSU. It can be used to create estimates that describe

1. Unlisted HB providers in each selected PSU.

3.2.3 Application of the PSU Weight

First, subset the HB main data file with the condition TYPE="UL". This creates a dataset containing only unlisted HB cases. Then, do a one-to-one merge of the HB PSU and cluster weights data file and the subsetted HB main data file by HB_METH_CASEID. The resulting data file should have 2052 rows. Sort the merged data file by PSU_ID. Set the weight variable to HB_WEIGHT_PSU. There are no strata or cluster variables. Generate the estimates for each PSU separately.

Sample SAS code for numeric variable of interest.

```
proc surveymeans data=HB_MERGED_DATA;  
  var VARIABLE_OF_INTEREST;  
  weight HB_WEIGHT_PSU;  
  by PSU_ID;  
run;
```

3.2.4 Description of the Cluster Weight

The unlisted HB cluster weight sums to the number of unlisted HB providers in each selected provider cluster. It can be used to create estimates that describe

1. Unlisted HB providers in each selected provider cluster.

3.2.5 Application of the Cluster Weight

First, subset the HB main data file with the condition TYPE="UL". This creates a dataset containing only unlisted HB cases. Then, do a one-to-one merge of the HB SSU and cluster weights data file and the subsetted HB main data file by HB_METH_CASEID. The resulting data file should have 2052 rows. Sort the merged data file by PCLUSTER_ID. Set the weight variable to HB_WEIGHT_PCLUSTER. There are no strata or cluster variables. Generate the estimates for each provider cluster separately.

Sample SAS code for numeric variable of interest.

```
proc surveymeans data=HB_MERGED_DATA;  
  var VARIABLE_OF_INTEREST;  
  weight HB_WEIGHT_PCLUSTER;  
  by PCLUSTER_ID;  
run;
```

3.3 Listed HB PSU and Cluster Weights

3.3.1 Weights Data File

The listed HB PSU and cluster weights data file will have the following variables:

HB_METH_CASEID

PSU_ID

PCLUSTER_ID

HB_WEIGHT_PSU: the HB PSU weight

HB_WEIGHT_PCLUSTER: the HB cluster weight

A HB_METH_CASEID could be linked to multiple PCLUSTER_IDs and/or multiple PSU_IDs, resulting in multiple rows with the same HB_METH_CASEIDs. This is because a listed HB case could belong to multiple provider clusters or PSUs. For example, if a case belongs to both provider clusters A and B, then this case will appear as two rows in the dataset. One row would have the PCLUSTER_ID of provider cluster A and the corresponding cluster weight. The other row would have the PCLUSTER_ID of provider cluster B and the corresponding cluster weight.

3.3.2 Description of the PSU Weight

The listed HB PSU weight sums to the number of listed HB providers in each selected PSU. It can be used to create estimates that describe

1. Listed HB providers in each selected PSU.

3.3.3 Application of the PSU Weight

First, subset the HB main data file with the condition TYPE="L". This creates a dataset containing only listed HB cases. Then, do data de-duplication for the listed HB PSU and cluster weights data by HB_METH_CASEID and PSU_ID. This will delete duplicated rows due to a HB_METH_CASEID linked to multiple PCLUSTER_IDs in the same PSU. Next, do a many-to-one merge of the de-duplicated listed HB PSU and cluster weights data file and the subsetted HB main data file by HB_METH_CASEID. The resulting data file should have 3934 rows. Sort the merged data file by PSU_ID. Set the weight variable to HB_WEIGHT_PSU. There are no strata or cluster variables. Generate the estimates for each PSU separately.

Sample SAS code for numeric variable of interest.

```
proc surveymeans data=HB_MERGED_DATA;
  var VARIABLE_OF_INTEREST;
  weight HB_WEIGHT_PSU;
  by PSU_ID;
run;
```

3.3.4 Description of the Cluster Weight

The listed HB cluster weight sums to the number of listed HB providers in each selected provider cluster. It can be used to create estimates that describe

1. Listed HB providers in each selected provider cluster.

3.3.5 Application of the Cluster Weight

First, subset the HB main data file with the condition TYPE="L". This creates a dataset containing only listed HB cases. Then, do a many-to-one merge of the listed HB SSU and cluster weights data file and the subsetted HB main data file by HB_METH_CASEID. The resulting data file should have 3934 rows. Sort the merged data file by PCLUSTER_ID. Set the weight variable to HB_WEIGHT_PCLUSTER. There are no strata or cluster variables. Generate the estimates for each provider cluster separately.

Sample SAS code for numeric variable of interest.

```
proc surveymeans data=HB_MERGED_DATA;
  var VARIABLE_OF_INTEREST;
  weight HB_WEIGHT_PCLUSTER;
  by PCLUSTER_ID;
run;
```

3.4 CB PSU and Cluster Weights

3.4.1 Weights Data File

The CB PSU and cluster weights data file will have the following variables:

CB_METH_CASEID

PSU_ID

PCLUSTER_ID

CB_WEIGHT_PSU: the CB PSU weight

CB_WEIGHT_PCLUSTER: the CB cluster weight

STRATA: strata based on CB type of care

A CB_METH_CASEID could be linked to multiple PCLUSTER_IDs and/or multiple PSU_IDs, resulting in multiple rows with the same CB_METH_CASEIDs. This is because a CB case could belong to multiple provider clusters or PSUs. For example, if a case belongs to both provider clusters A and B, then this case will appear as two rows in the dataset. One row would have the PCLUSTER_ID of provider cluster A and the corresponding cluster weight. The other row would have the PCLUSTER_ID of provider cluster B and the corresponding cluster weight.

3.4.2 Description of the PSU Weight

The CB PSU weight sums to the total number of ECE centers that serve at least one child age five and under not yet in kindergarten in each selected PSU. It can be used to create estimates that describe:

1. ECE center that serve at least one child age five and under not yet in kindergarten in each selected PSU. This is when the weights are used to create summary statistics for variables concerning the whole center.

It cannot be used to create estimates that describe:

1. ECE age groups in each selected PSU. We cannot use the weights to create summary statistics for variables concerning only the selected age groups (variables starting with CB_F1_ and variables created from them).
2. ECE classrooms in each selected PSU. We cannot use the weights to create summary statistics for variables concerning only the selected classrooms (variables starting with CB_F3_ or CB_F8_ and variables created from them).

3.4.3 Application of the PSU Weight

First, perform a data de-duplication for the CB PSU and cluster weights data by CB_METH_CASEID and PSU_ID. This will delete duplicated rows due to a CB_METH_CASEID linked to multiple PCLUSTER_IDs in the same PSU. Next, do a many-to-one merge of the de-duplicated CB PSU and cluster weights data file and the CB main data file by CB_METH_CASEID. The resulting data file should have 8312 rows. Sort the merged data file by PSU_ID. Set the weight variable to CB_WEIGHT_PSU. Set the strata variable to STRATA. There is no cluster variable. Generate the estimates for each PSU separately.

Sample SAS code for numeric variable of interest.

```
proc surveymeans data=CB_MERGED_DATA;  
  var VARIABLE_OF_INTEREST;  
  weight CB_WEIGHT_PSU;  
  strata STRATA;  
  by PSU_ID;  
run;
```

3.4.4 Description of the Cluster Weight

The CB cluster weight sums to the total number of ECE centers in each selected provider cluster. It can be used to create estimates that describe:

1. ECE center in each selected provider cluster.

It cannot be used to create estimates that describe:

1. ECE age groups in each selected provider cluster.
2. ECE classrooms in each selected provider cluster.

3.4.5 Application of the Cluster Weight

First, do a many-to-one merge of the CB SSU and cluster weights data file and the CB main data file by CB_METH_CASEID. The resulting data file should have 9552 rows. Sort the merged data file by PCLUSTER_ID. Set the weight variable to CB_WEIGHT_PCLUSTER. Set the strata variable to STRATA. There is no cluster variable. Generate the estimates for each provider cluster separately.

Sample SAS code for numeric variable of interest.

```
proc surveymeans data=CB_MERGED_DATA;  
  var VARIABLE_OF_INTEREST;  
  weight CB_WEIGHT_PCLUSTER;  
  strata STRATA;  
  by PCLUSTER_ID;  
run;
```

3.5 WF PSU and Cluster Weights

3.5.1 Weights Data File

The WF PSU and cluster weights data file will have the following variables:

WF_METH_CASEID

PSU_ID

PCLUSTER_ID

WF_WEIGHT_PSU: the WF PSU weight

WF_WEIGHT_PCLUSTER: the WF cluster weight

STRATA: strata based on CB type of care

A WF_METH_CASEID could be linked to multiple PCLUSTER_IDs and/or multiple PSU_IDs, resulting in multiple rows with the same WF_METH_CASEIDs. This is because a WF case could belong to multiple provider clusters or PSUs. For example, if a case belongs to both provider clusters A and B, then this case will appear as two rows in the dataset. One row would have the PCLUSTER_ID of provider cluster A and the corresponding cluster weight. The other row would have the PCLUSTER_ID of provider cluster B and the corresponding cluster weight.

3.5.2 Description of the PSU Weight

The WF PSU weight sums to the total number of lead, regular, and assistant teachers and instructors and aids, working at least five hours a week in ECE centers in each selected PSU. It can be used to create estimates that describe:

1. Eligible CB staff in each selected PSU.

3.5.3 Application of the PSU Weight

First, conduct a data de-duplication for the WF PSU and cluster weights data by WF_METH_CASEID and PSU_ID. This will delete duplicated rows due to a WF_METH_CASEID linked to multiple PCLUSTER_IDs in the same PSU. Next, do a many-to-one merge of the de-duplicated WF PSU and cluster weights data file and the WF main data file by WF_METH_CASEID. The resulting data file should have 5556 rows. Sort the merged data file by PSU_ID. Set the weight variable to WF_WEIGHT_PSU. Set the strata variable to STRATA. There is no cluster variable. Generate the estimates for each PSU separately.

Sample SAS code for numeric variable of interest.

```
proc surveymeans data=WF_MERGED_DATA;  
  var VARIABLE_OF_INTEREST;  
  weight WF_WEIGHT_PSU;  
  strata STRATA;  
  by PSU_ID;  
run;
```

3.5.4 Description of the Cluster Weight

The WF PSU weight sums to the total number of lead, regular, and assistant teachers and instructors and aids, working at least five hours a week in ECE centers in each selected provider cluster. It can be used to create estimates that describe:

1. Eligible CB staff in each selected provider cluster.

3.5.5 Application of the Cluster Weight

First, do a many-to-one merge of the WF SSU and cluster weights data file and the WF main data file by WF_METH_CASEID. The resulting data file should have 5556 rows. Sort the merged data file by PCLUSTER_ID. Set the weight variable to WF_WEIGHT_PCLUSTER. Set the strata variable to STRATA. There is no cluster variable. Generate the estimates for each provider cluster separately.

Sample SAS code for numeric variable of interest.

```
proc surveymeans data=WF_MERGED_DATA;
  var VARIABLE_OF_INTEREST;
  weight WF_WEIGHT_PCLUSTER;
  strata STRATA;
  by PCLUSTER_ID;
run;
```

4 Tutorial for Analyzing PSU and Cluster Level Metrics Created from Provider Data

4.1 Analyzing PSU Level Provider Metrics

Step 1: Follow instructions in each of the “Application of the PSU Weight” sections to create your desired metrics for provider data at the PSU level. Generate a dataset with PSU_ID and the created provider metrics. This dataset should have 219 or less rows. Each row should have a unique PSU_ID.

Step 2: Do a one-to-one merge of the PSU_ID variable from the HH PSU and cluster weight data file to the HH main data by HH_METH_CASEID. This dataset should have 11629 rows. Each row should have a unique HH_METH_CASEID.

Step 3: Do a one-to-one merge of the metrics dataset and the merged HH main data by PSU_ID. Now you have a dataset with 11629 rows. Each row should have a unique HH_METH_CASEID.

Step 4: Set the weight variable to HH_WEIGHT, the strata variable to HH_VSTRATUM, and the cluster variable to HH_VPSU. Analyze the provider metrics as variables that describe the provider environment of the PSU each HH resides in.

Sample SAS code:

```
proc surveyfreq data=HH_MERGED_DATA;
  table PROVIDER_METRICS*ORIGINAL_HH_VARIABLE/row;
  weight HH_WEIGHT;
  strata HH_VSTRATUM;
  cluster HH_VPSU;
run;
```

4.2 Analyzing Cluster Level Provider Metrics

Step 1: Follow instructions in each of the “Application of the Cluster Weight” sections to create your desired metrics for provider data at the provider cluster level. Generate a dataset with PCLUSTER_ID

and the created provider metrics (keep also PSU_ID if you want to run step 1.5). This dataset should have 755 or less rows. Each row should have a unique PCLUSTER_ID.

Step 1.5: This is an optional step. The purpose of this step is to reduce the number of provider clusters with missing metrics by supplementing the cluster level metrics with the PSU level metrics.

Do step 1 of “Analyzing PSU Level Provider Metrics” to generate a dataset with PSU_ID and PSU level metrics.

Do a one-to-many merge of the PSU level metrics dataset to the provider cluster level metrics dataset by PSU_ID. The resulting dataset should have the same number of rows as the provider cluster level metrics dataset. Each row should have a unique PCLUSTER_ID.

Step 2: Do a one-to-one merge of the PCLUSTER_ID variable from the HH PSU and cluster weight data file to the HH main data by HH_METH_CASEID. This dataset should have 11629 rows. Each row should have a unique HH_METH_CASEID.

Step 3: Do a one-to-one merge of the metrics dataset and the merged HH main data by PCLUSTER_ID. Now you have a dataset with 11629 rows. Each row should have a unique HH_METH_CASEID.

Step 4: Set the weight variable to HH_WEIGHT, the strata variable to HH_VSTRATUM, and the cluster variable to HH_VPSU. Analyze the provider metrics as variables that describe the provider environment around the HH SSU each HH complete resides in.

Sample SAS code:

```
proc surveyfreq data=HH_MERGED_DATA;  
  table PROVIDER_METRICS*ORIGINAL_HH_VARIABLE/row;  
  weight HH_WEIGHT;  
  strata HH_VSTRATUM;  
  cluster HH_VPSU;  
run;
```

5 Summary of Cluster Weight Descriptions

3.1.4 HH	3.2.4 Unlisted HB	3.3.4 Listed HB	3.4.4 CB	3.5.2 WF
<p>The HH cluster weight sums to the number of households with children under age 13 in each selected HH SSU. It can be used to create estimates that describe, for each selected HH SSU:</p>	<p>The unlisted HB cluster weight sums to the number of unlisted HB providers in each selected provider cluster. It can be used to create estimates that describe</p>	<p>The listed HB cluster weight sums to the number of listed HB providers in each selected provider cluster. It can be used to create estimates that describe</p>	<p>The CB cluster weight sums to the total number of ECE centers in each selected provider cluster. It can be used to create estimates that describe:</p>	<p>The WF PSU weight sums to the total number of lead, regular, and assistant teachers and instructors and aids, working at least five hours a week in ECE centers in each selected provider cluster. It can be used to create estimates that describe:</p>
<ol style="list-style-type: none"> 1. Households with children under age 13. 2. All children under age 13 in households with children under 13. 3. All household members age 13 and over in households with children under 13. 4. All ECE arrangements used by children under age 13. 	<ol style="list-style-type: none"> 1. Unlisted HB providers in each selected provider cluster. 	<ol style="list-style-type: none"> 1. Listed HB providers in each selected provider cluster. 	<ol style="list-style-type: none"> 1. ECE center in each selected provider cluster. 	<ol style="list-style-type: none"> 1. Eligible CB staff in each selected provider cluster.

3.1.4 HH	3.2.4 Unlisted HB	3.3.4 Listed HB	3.4.4 CB	3.5.2 WF
			It cannot be used to create estimates that describe: 1. ECE age groups in each selected provider cluster. 2. ECE classrooms in each selected provider cluster.	

6 Discussion of limits

6.1 Estimation on Different Geographic Levels

To calculate estimates at the national level, one should use the main interview weights. Using the PSU weights or the cluster weights will introduce large coverage errors into the estimates, because the PSU weights and the cluster weights only calculate estimates for the selected PSUs and the selected clusters. The not selected PSUs and clusters are not taken into consideration.

To calculate estimates at the state level, one should also use the main interview weights. This is because all states in the U.S. were selected into the sample by design. Using the PSU weights or the cluster weights will introduce large coverage errors into the estimates, because the selected PSUs and the selected clusters only cover parts of each state.

County level estimates can only be calculated when a PSU is a single county. In this case, the PSU estimate is the same as the county estimate. When a PSU consists of several counties, one cannot apply the PSU weights to interview completes in only one county to calculate estimate for that county.

Census tract level estimates can only be calculated for the HH data when a HH SSU is a single county. In this case, the HH SSU estimate is the same as the tract estimate. When a HH SSU consists of several tracts, one cannot apply the cluster weights to the HH interview completes in only one tract to calculate estimate for that tract. A provider cluster consists of several tracts by default; therefore, we cannot calculate tract level estimates from the provider data.

6.2 Incorrect Ways to Analyze PSU and Cluster Metrics

When you have a PSU or cluster metrics dataset with 219 or 755 rows, you should not analyze the dataset on its own. This is because the PSU metrics dataset only contains estimates for the selected PSUs. There is no weight to apply to the 219 rows to generalize the analysis to the national level. Similarly, the cluster metrics dataset only contains estimates for the selected HH SSUs or provider clusters. There is no weight to apply to the 755 rows to generalize the analysis to the national level.

In the “Tutorial for Analyzing PSU and Cluster Level Metrics Created from Provider Data” section, we have talked about merging PSU or cluster metrics datasets onto the HH main data for analysis. It is also possible to merge metrics PSU or cluster metrics datasets onto the unlisted HB main data for analysis, since each unlisted HB complete is linked to only one PCLUSTER_ID and only one PSU_ID. However, since a listed HB, CB, and WF complete could link to multiple PCLUSTER_IDs and even to multiple PSU_IDs, users should not merge metrics PSU or cluster metrics datasets onto the listed HB, CB, and WF main data for analysis.