NATIONAL SURVEY OF OLDER AMERICANS ACT PARTICPANTS (NSOAAP)
FOURTEENTH NATIONAL SURVEY (2019)

1. SAMPLE SELECTION, WEIGHTING, AND VARIANCE ESTIMATION

The survey employed a two-stage sample design, first selecting a sample of Area Agencies on Aging (AAAs) in stage one and, in the second stage, a sample of clients for each service within each sampled AAA. The fourteenth national survey covered six services – Home Delivered Meals, Homemaker Services, Transportation, the Family Caregiver Support Program, Congregate Meals and Case Management.

Weighting of each service was done separately. Initially, base weights were computed by taking the inverse of the selection probability for each sampled client. Then the base weights were adjusted for nonresponse, followed by trimming of the extreme weights. Finally, a poststratification adjustment was made using available control totals from State Program Reports supplied by ACL. Fay's modified Balanced Repeated Replication (BRR) method was used for computation of the sampling variances of survey estimates.

Agency Selection

At the first stage of the two-stage design for the national survey, a stratified sample of 350 AAAs (allowing for a 25% non-response rate) was selected from the frame of 629 agencies. This sample size was increased from 325 in the thirteenth survey due to (1) the AAA response rate had been hovering at about 75% for the last three years (Surveys 10-13) whereas for Survey 9 and earlier surveys, the response rate was slightly more than 80%, and (2) the AAA sample for the 14^{th} NSOAAP was originally proposed as the baseline sample for a longitudinal expansion of the survey. In addition, the loss in the number of responding AAAs resulted in continually inflating the within-AAA client sample sizes over the past several years. This increase was becoming untenable as we were not achieving our target of 6,000 responding clients. From a statistical point of view, adding more clients within AAAs is less effective at providing additional information or precision in the estimates since clients within AAAs tend to be clustered.

The sampling frame was essentially the same as that used for the sixth through thirteenth national surveys, except for an agency added or removed from year to year. The agency measures of size (MOS19) were completely updated in 2019 using total client counts instead of the budget figures used since 2011. These client counts were taken from previous cycles as far back as 2011, although three-quarters of the client counts were from 2015-2018. Client counts were available for 567 of the 629 AAAs in the frame. For the remaining 62 AAAs with no client counts available from previous surveys, we calculated the predicted value for client counts based on the AAAs' annual budget figures which had been previously used as the measures of size and Census Division (listed below). Having client counts for all AAAs on the frame enabled the use of standard probability proportional to size (PPS) sampling of AAAs.

Census Division	States
1 (New England)	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
2 (Middle Atlantic)	New Jersey, New York, Pennsylvania
3 (East North Central)	Illinois, Indiana, Michigan, Ohio, Wisconsin
4 (West North Central)	Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota
5 (South Atlantic)	Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia
6 (East South Central)	Alabama, Kentucky, Mississippi, Tennessee
7 (West South Central)	Arkansas, Louisiana, Oklahoma, Texas
8 (Mountain)	Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming
9 (Pacific)	Alaska, California, Hawaii, Oregon, Washington

The AAA sample was selected independently within four strata, which were based on the four Census Regions (Northeast, Midwest, South, West). With the updated sampling (probability proportional to the total number of clients) within stratum (comprised of the four Census regions), we sampled 174 AAAs with certainty out of 350 sampled AAAs. The total sample size was allocated to the four strata as shown in the following table:

Table 1 Sampling strata and allocation of agencies into strata for the national sample.

STRAT19	Census Region	Frame N	Sum of MOS19	% of total MOS19	Allocation of AAA Sample (TARGET)
1	Northeast (NE)	171	641,200	25.62%	90
2	Midwest (MW)	104	662,220	26.46%	93
3	South (S)	229	604,519	24.15%	84
4	West (W)	125	595,127	23.77%	83
Total	Total	629	2,503,066	100.00%	350

Note that since the measures of size are more nearly uniformly distributed across region (the sampling strata) comparisons among estimates at the regional level will tend to be more precise.

Client Selection

Client samples by service type (Home Delivered Meals, Homemaker, Transportation, Caregiver Service, Congregate Meals, and Case Management) were drawn randomly within each sampled AAA. Before selecting the sample of clients, Westat obtained the total number of clients receiving each service within an agency by contacting either the sampled agency or the State Unit on Aging (SUA) for the state in which the sampled agency is located. Based on the total number of clients, line numbers from client master lists were sampled using a Westat software application that started with the total number of clients in each service by agency and randomly selected the matching line numbers for the sampled clients. In past surveys, the goal was to increase the number of clients selected by the inverse of the rates observed in the previous cycle of the national survey in order to meet the required sample size for each service. However, to continue to do so for the 14th National Survey would have resulted in the proposed numbers getting to be too large for some services. Thus, for the 14th National Survey, as mentioned earlier in this document, the sample size of AAAs was increased from 325 to 350 to increase the pool of clients from which to sample.

As a result of the updated sampling and the large number of AAAs selected with certainty, fixed-size client samples were selected from each agency (certainty and non-certainty) for each service as indicated in Table 2 below.

Table 2 **Within-AAA sample sizes for the six target services**

Service	Non-certainty Stratum
Family Caregiver	35
Home Delivered Meals	12
Homemaker Service	8
Transportation	20
Congregate Meals	16
Case Management	14

Selection Probability

The probability of selection of a client within a service can be mathematically expressed as follows. First, for non-certainty agencies, let

$P_{i \in h}$ = Probability of selection of agency i in stratum h,

$$= \frac{\text{Number of non-certainty agencies selected from the stratum}}{\text{Total number of non-certainty agencies in the stratum}}$$

$= \frac{m_h}{M_h}$, for agencies in a non-certainty stratum.

For certainty agencies, the probability of selection was 1 (that is, $P_{h=c} = 1$). Next, let

P_{ijs} = Probability of selection of client j in service s within agency i,

$$= \frac{\text{Number of clients selected from service } s \text{ in agency } i}{\text{Total number of clients in service } s \text{ in agency } i} = \frac{n_{is}}{N_{is}}.$$

Recall that n_{is} was fixed in advance for both certainty and non-certainty agencies by service, as shown in Table 2.

Thus, the overall probability of selection of client j in service s within agency i in stratum h was

$\pi_{ijs} = P_{i \in h} \times P_{ijs} = \frac{m_h}{M_h} \times \frac{n_{is}}{N_{is}}$ for the clients within non-certainty agencies,

$$= 1 \times \frac{n_{is}}{N_{is}} = \frac{n_{is}}{N_{is}} \qquad \text{for the clients within certainty agencies.}$$

Weighting

Weighting was done in four steps: calculation of base weights, nonresponse adjustment, trimming of extreme weights, and poststratification adjustments to known population control totals.

Base Weights

The base weight is the inverse of the overall selection probability of a client. The base weight of a client can be obtained by calculating the base weight for an agency and multiplying that weight by the within-agency-level base weight of a client in a service within that agency.

The base weight for an agency i can be expressed as

$$a_{i, i \in h} = \frac{1}{P_h} = \frac{M_h}{m_h} \qquad \text{for non-certainty agencies,}$$

$$= 1 \qquad \text{for certainty agencies,}$$

and the base weight for a client in a service within an agency can be expressed as

$$v_{ijs} = \frac{1}{P_{ijs}} = \frac{N_{is}}{n_{is}},$$

= the within-agency base weight of client j in service s within agency i.

Therefore, the overall base weight of a client within a service is

$$w_{ijs} = a_i \times v_{ijs} = \frac{1}{\pi_{ijs}},$$

$$= \frac{M_h}{m_h} \times \frac{N_{is}}{n_{is}} \qquad \text{for non-certainty agencies,}$$

$$= 1 \times \frac{N_{is}}{n_{is}} \qquad \text{for certainty agencies.}$$

Nonresponse Adjustment

Since not all sampled agencies and clients responded to the survey, the base weights had to be adjusted for nonresponse. The nonresponse adjustment was done in two steps by performing separate adjustments for agency-level and client-level nonresponse. The nonresponse adjustments were applied specific to each service group within Census region.

If m_{hs}^{r} denotes the number of agencies in stratum h that responded to the survey for service s, then the agency-level nonresponse adjustment was calculated as follows:

$$a_{is,i\in h}^{r} = \frac{M_h}{m_h} \times \frac{m_h}{m_{hs}^{r}} = \frac{M_h}{m_{hs}^{r}}$$

= the nonresponse adjusted weight of agency i for service s.

If n_{is}^{r} denotes the number of clients that responded for service s within agency i, then the client-level nonresponse adjustment was calculated as follows:

$$v_{ijs}^{r} = \frac{N_{is}}{n_{is}} \times \frac{n_{is}}{n_{is}^{r}} = \frac{N_{is}}{n_{is}^{r}},$$

= the nonresponse adjusted weight for client j for service s within agency i.

Therefore, the overall nonresponse-adjusted weight of client j for service s within agency i is

$$w_{ijs}^{r} = a_{is}^{r} \times v_{is}^{r} = \frac{M_h}{m_{hs}^{r}} \times \frac{N_{is}}{n_{is}^{r}} .$$

Trimming of Weights

To keep the variance of the survey estimates within an acceptable level, extreme weights were trimmed. The variability in weights was increased due to the adjustment of client nonresponse rates that varied substantially from agency to agency. In addition, since not all agencies provided all services, variability in sample size by service contributed to increased variability in the base weights. Since variability in the weights increases the variances of the survey estimates, those weights which were too high compared to the median base weight over all clients within a given service were trimmed to acceptable upper limits to reduce the overall variance of the survey estimates.

Initially, the acceptable upper limits were determined by using the median base weight within a service group such that weights larger than 4 times the median base weight in the service group were trimmed to be equal to 4 times the median base weight in the group. However, for all six services, this trimming rule was empirically shown to over-trim with respect to the percentiles of the distribution of all weights for that service. Thus, for Family Caregiver, Congregate Meals, Homemaker, Home Delivered Meals, and Transportation, the weights were trimmed at the 98th percentile. For Case Management the weights were trimmed at the 96th percentile. One effect of trimming weights is that estimated totals are reduced from what they would have been, had trimming not been applied to the weights. This loss in the sum of weights due to the trimming was adjusted in the final poststratification step described below. The trimmed, nonresponse adjusted weights will be denoted by w_{ijs}^{θ} in the following sections.

Poststratification Adjustment

The final step of weighting involved the benchmarking of the estimated number of clients in a service (based on the trimmed, nonresponse-adjusted weights) to the known total number of clients (control total) obtained from the AoA State Program Reports (SPR). The poststratification adjustment, or benchmarking, was done at the regional level, since reliable control totals were available at the regional level. Also, controlling for regioin in the adjustments has proven to be effective in producing more precise survey estimates.

The post-stratified weights (w_{ijs}^{p}) for service s were calculated by multiplying the trimmed, nonresponse-adjusted weights (w_{ijs}^{θ}) by the ratio of the known control total (N_s) to the estimated total $(\sum_{ij} w_{ijs}^{\theta})$ as follows:

$$w_{ijs}^{p} = w_{ijs}^{\theta} \times \frac{N_s}{\sum_{ij} w_{ijs}^{\theta}}$$

The poststratification adjustment described in this paragraph was applied to Home-delivered Meals, Homemaker Services, Congregate Meals, Case Management, and Family Caregiver. The adjustments for Transportation services were calculated somewhat differently and are described below.

Poststratification Adjustment for Transportation Service

For the Transportation service, control totals for the number of clients were not available. However, State Units on Aging (SUAs) did provide the number of one-way passenger trips in the State Program Reports (SPR). These SPR regional level trip counts were used for the purpose of estimating control totals for the number of clients receiving transportation services by region. The following summarizes the methodology used for constructing these estimated transportation client counts:

- The national survey asked respondents how many one-way trips per month they usually took using the AAA transportation service.

- An average annual per-person trip count by region was estimated from the survey data file using the trimmed, nonresponse-adjusted weights.

- By dividing the total trip count by the per-person average annual number of trips, Westat estimated the total number of persons who received transportation services by region.

The method of estimation explained above can be mathematically expressed as follows:

$$\hat{N}_s = \sum_g \hat{N}_{gs} = \sum_g \frac{T_g}{\bar{t}_g} = \sum_g \frac{T_g}{\dfrac{\sum_{ij} t_{ij} w_{ijs}^{\theta}}{\sum_{ij} w_{ijs}^{\theta}}} = \sum_g \frac{T_g}{\hat{T}_{gw}} \times \hat{N}_{gw} \text{ ,}$$

where

$\hat{N}_s$ is the final estimate of transportation client count,

$\hat{N}_{gs}$ is the final estimate of transportation client count in region g ,

T_g is the total number of one-way trips reported by the SUAs in region g ,

$\bar{t}_g = \dfrac{\sum_{ij,i\in g} t_{ij} w_{ijs}^{\theta}}{\sum_{ij,i\in g} w_{ijs}^{\theta}}$ is the per-person weighted average of annual number of trips in region g ,

t_{ij} is the number of annual one-way trips made by client j in agency i ,

$\hat{T}_{gw} = \sum_{ij,i \in g} t_{ij} w_{ijs}^{\theta}$ is an initial estimate of the total number of one-way trips in region g based on the trimmed, nonresponse-adjusted weights;

$\hat{N}_{gw} = \sum_{ij,i \in g} w_{ijs}^{\theta}$ is an initial estimate of the total number of transportation clients in region g based on the trimmed, nonresponse-adjusted weights.

The above estimator is widely known as a *Ratio Estimator* in the sample survey literature because the initial estimate of the total number of transportation clients ($\hat{N}_w$) is adjusted by the ratio of actual to estimated total number of one-way trips ($\frac{T}{\hat{T}_w}$).

Variance Estimation

Westat routinely uses replication-based variance estimation methods for computing sampling variances of the survey estimates derived from complex multi-stage sample designs. Westat's variance computation software, WesVar, is designed for this purpose. A version of balanced repeated replication (BRR) referred to as "Fay's method" was used to calculate the variances (and their square roots, the standard errors) of estimates derived from the NSOAAP. Implementation of BRR methods for variance estimation requires the use of a series of "replicate weights," each of which provides an alternative (replicate-specific) estimate of a characteristic of interest. The variability of the replicate estimates about the full-sample estimate of the same characteristic is then used to obtain the variance or standard error of the characteristic.

Let y_{ij} denote a survey characteristic (variable) for the j th respondent in the i th agency, and let w_{ij}^{p} denote the corresponding full-sample final weight. Further, let w_{ij}^{k} denote the kth replicate weight, where k = 1, 2, ..., K . The estimated total for the survey variable is given by the weighted sum

$$\hat{y} = \sum_{ij} w_{ij}^{p} y_{ij} \ .$$

The corresponding replicate estimates are given by the weighted sums

$$\hat{y}_k = \sum_{ij} w_{ij}^{k} y_{ij} \text{ , for } k = 1, 2, ..., K$$

The variance of the estimate $\hat{y}$ is then computed as:

$$\text{var}(\hat{y}) = \frac{1}{(1-.30)^2} \sum_{k=1}^{K} (\hat{y}_k - \hat{y})^2 ,$$

where the 0.30 in the above formula is referred to as "Fay's factor." The corresponding standard error is simply the square root of $\text{var}(\hat{y})$ as computed above.

The replicate weights, w_{ij}^k, required for variance estimation were derived from replicate-specific base weights and include all of the adjustments (e.g., nonresponse, trimming, and poststratification) used to develop the final full-sample weights, w_{ij}^p .

Replicates were formed by first creating variance strata and variance units. For non-certainty AAAs, variance strata were formed with two or three AAAs in each stratum, and each AAA was treated as a variance unit. For certainty AAAs, each AAA was treated as a variance stratum, and random groups of clients were formed as variance units within the stratum. This difference in forming variance strata for certainty and non-certainty AAAs was necessary to account for the fact that there was no first stage sampling variance for certainty AAAs. Under BRR, the replicates are formed in a balanced way by taking one variance unit from each variance stratum. However, a modified version of BRR called Fay's method was used for the AoA survey. Under the modified approach, the full-sample weights are adjusted or "perturbed" to define the required replicates, rather than taking one variance unit from each stratum. Further details on BRR and Fay's method, or replication methods in general, can be found in WesVar 5.1 User's Guide. The User's Guide is available without charge by emailing wesvar_tech_support@westat.com; see this link: https://www.westat.com/capability/information-systems-software/wesvar/wesvar-documentation. Note that the User's Guide is for WesVar 4.2, with an addendum for what is new in WesVar 5.1.

WesVar, SUDAAN, STATA, SAS, SPSS and other complex sample survey software packages can use replicate weights to compute variance estimates that fully account for the complex design used in the AoA national surveys.

2. SIGNIFICANCE TESTING OF THE DIFFERENCE BETWEEN TWO SURVEY CHARACTERISTICS

The statistic given below can be used to test whether the observed difference between two estimated proportions is statistically significant. This test can be used to check the significance of the difference either between an agency level and a national level characteristic or between characteristics estimated for two differentsubgroups, such as regions. The test statistic is

$$z = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{SE^2(\hat{p}_1) + SE^2(\hat{p}_2)}}$$

where, $\hat{p}_1$ and $\hat{p}_2$ are estimates of the two survey characteristics to be compared, and $SE^2(\hat{p}_1)$ and $SE^2(\hat{p}_2)$ are squares of the corresponding standard errors of the two estimates.

When the sample size (i.e., the number of valid responses in each comparison group) is 30 or more, the above test statistic will approximately follow a statistical distribution called the *normal distribution* and the difference will be considered significant at the 5% level of significance if $z > 1.96$. The interpretation of such a result is that the probability of obtaining a difference as large as the observed difference by chance alone is less than 5%.

However, if the number of valid responses in one of the groups is less than 30, then the above test statistic will follow a different statistical distribution called the *t-distribution* with $(n_1 + n_2 - 2)$ degrees of freedom, where n_1 and n_2 are the number of valid responses in the two groups. In this case, the critical value for the significance of a difference will depend on $(n_1 + n_2 - 2)$. The following table presents a rough indication of the critical values of the t distribution for a 5% level of significance for different values of $(n_1 + n_2 - 2)$. The computed value of z must be greater than the corresponding critical value for the difference between the two estimates to be considered significant.

Degrees of freedom, $(n_1 + n_2 - 2)$	Critical value of t distribution at the 5% level of significance
>58	1.96
30-58	2.05
25-29	2.06
20-24	2.08
15-19	2.13

For interested readers, more detailed tables of critical values of the normal, t, and other statistical distributions are available in standard textbooks on statistical methods.