A.O.A. THIRD NATIONAL SURVEY

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 survey covered three services this year – Home Delivered Meals, Transportation and the Family Caregiver Support Program.

Weighting of each service record 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. 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 310 AAAs (allowing for a 20% non-response) was selected from the frame of 649 agencies. The same sampling frame used for the first and second national surveys was used for this survey.

The AAA sample was selected independently within five budget size strata, which were created based on the square root of the total budget sizes of the AAAs. The AAA and client samples were proportionally allocated to the total of the square root of the budget sizes in each stratum. However, within a stratum the sample of AAAs was selected with equal probability, but sorted by Census region. This method was used instead of a direct probability proportional to size (PPS) sampling because in the earlier national surveys it was found that budget size was not necessarily well correlated with the total number of clients in each agency for every service. In the absence of any other information, budget size was still used in sample selection, but with less importance. First, the square root of the budget size was used to reduce the effect of large variation in budget sizes. Second, the sample was allocated at the stratum level proportional to the total of the square root of the budget size. This procedure gave a higher probability of selection to agencies with larger budget sizes, but the agencies with budget sizes within different ranges (size strata) received the same probability of selection. As in the prior surveys, some agencies were selected with certainty. The total sample size was allocated to the five strata as shown in the following table:

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

STRATUM	Square Root of Budget Size	Allocation of AAA Sample
Certainty	Greater than $5,199	40
Non-certainty Stratum 1	$2,117 - $5,199	60
Non-certainty Stratum 2	$1,481 - $2,116	70
Non-certainty Stratum 3	$995 - $1,480	82
Non-certainty Stratum 4	Less than $995	58

The forty agencies with the largest budget sizes were selected with certainty for the AAA sample. The remaining sample was then selected independently within each of the non-certainty strata. The implicit stratification (sorting) variables in the selection process were Census Division and state, meaning that the number of agencies in each Division or state was selected roughly in proportion to the total of the square root of budget of a Division or a state. Table 2 shows the agency distribution in the frame and in the sample by Census Region.

Table 2 Distributions of agencies in the universe and in the sample by region.

Census Region	Number of AAAs in the Frame	Number of AAAs in the Sample
Northeast	172	76
Midwest	121	70
South	233	106
West	123	58
Total	**649**	**310**

Client Selection

Client samples by service type (Home Delivered Meals, Transportation, Caregiver Service) were drawn randomly within each sampled AAA. The total number of clients receiving each service within an agency was obtained by contacting the sampled agencies before selecting the sample of clients. Based on the total number of clients, line numbers from client master lists were sampled using a Westat software application that took the total number of clients in each service by agency and randomly selected the matching line numbers for the selected clients. In the first two national surveys, a fixed, predetermined number of clients in a service were selected from each selected agency. However, in the current survey, it was decided to select a number of clients that is proportional to the total number of clients in a service offered by an agency. Thus, instead of a fixed sample size per service per agency, a sampling fraction was provided for each agency by service by size group combination (see Table 3). The software then calculated the sample size for the particular cell by multiplying the sampling fraction by the total client size in the cell. The program then selected the sample of line numbers equal to the derived sample size.

The calculated sample sizes were modified for the agencies with client populations that were either too small or too large. The final sample size was derived as follows:

Sample Size = Minimum of (8, Total Clients), if the computed sample size < 8
= 50, if the computed sample size > 50
= the computed sample size, otherwise.

That means, if the total number of clients in the cell was >8, then a sample of at least 8 clients was selected. If the total number of clients in the cell was <8, then all clients in the service in that agency were selected. On the other hand, a maximum of 50 clients in a service were selected from an agency regardless of the size of the agency.

Table 3 Sampling fractions by size group used for the three target services

Service	Size Group				
	0 (Certainty Stratum)	1 (Non-certainty Stratum 1)	2 (Non-certainty Stratum 2)	3 (Non-certainty Stratum 3)	4 (Non-certainty Stratum 4)
Family Caregiver Support Program (FC)	2.2%	3.5%	3.3%	5.3%	11.8%
Home Delivered Meals (PM)	0.09%	1.73%	1.8%	1.8%	3.5%
Transportation (PT)	4.6%	6.0%	8.0%	7.0%	10.7%

Selection Probability

The probability of selection of a client within a service can be mathematically expressed as follows. First,

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

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

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

For certainty agencies, the probability of selection was 1 (i.e., $P_{h=c} = 1$).

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 calculated by multiplying the appropriate sampling fraction in Table 3 by the total number of clients receiving service s in agency i.

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

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

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

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 of 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 of an agency i can be expressed as

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

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

and the base weight of 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.

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

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

Trimming of Weights

To keep the variance of the survey estimates within an acceptable level, extreme weights were trimmed. The sample design was set up to select clients within a service with equal probability so that the base weights of all clients within a service would be roughly equal. This would have been the case if the measure of size used in selecting the agencies (i.e., each agency's annual budget) was perfectly correlated with the number of clients in a service and if there had been no nonresponse. But in reality, this correlation was not high, and there was some nonresponse. Some agencies had larger budgets due to larger client sizes in some services but smaller numbers of clients in other services. Similarly, some agencies had smaller budgets but relatively larger numbers of clients in a particular service. This contributed to increased variability in the selection probabilities and subsequently in the base weights. Moreover, the variability in weights was increased further due to the adjustment of client nonresponse rates that varied substantially from agency to agency. Since variability in the weights increases the variances of the survey estimates, those weights which were too high compared to the median base weight were trimmed to upper acceptable limits to reduce the variance of the weights.

The upper acceptable limits were determined by using the median base weight within a service group. 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. 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 distributed to the weights of other clients (whose weights were not trimmed) in the same trimming cell defined by stratum and Census Division. This adjustment ensured that the sum of the weights remained the same before and after trimming, but variability in the weights was reduced. In other words, this adjustment made a compromise between the reduction in variance and the increase in bias due to trimming. 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.

The post-stratified weights (w_{ijs}^{p}) for the service s were calculated by multiplying the trimmed, nonresponse-adjusted weights (w_{ijs}^{θ}) by the ratio of the known control total (N_s) and 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 and Caregiver services. The adjustment for Transportation services is described below.

<u>Poststratification Adjustment for Transportation Service</u>

For the Transportation service, the control totals 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 purposes of computing control totals for the number of clients receiving transportation services by region. The following summarizes the methodology used for constructing these transportation client counts:

- The national survey asked respondents how many one-way trips per month they usually took using the AAA transportation service. To ensure proper identification of AAA-funded transportation programs, the computer assisted telephone interviewing (CATI) software allowed the interviewers to prompt the respondents with the specific name of the transportation service, which the provider had supplied to Westat during the client sampling stage.

- 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, we estimated the total number of persons who received transportation services by region.

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

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

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 ,

$\bar{t}_g = \frac{\sum_{ij,i \in g} t_{ij} w_{ijs}^{\theta}}{\sum_{ij,i \in g} w_{ijs}^{\theta}}$ 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 AoA national survey. 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 was given by the weighted sum

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

The corresponding replicate estimates were 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 $\hat{y}$ was 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 and poststratification) used to develop the full-sample weights, w_{ij}^{p} .

Replicates were formed first by 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 4.0 User's Guide, (www.Westat.com).

WesVar, SUDAAN, STATA and other complex sample survey software packages can use replicate weights to compute variance estimates.

2. SIGNIFICANCE TESTING OF THE DIFFERENCE BETWEEN TWO ESTIMATES

The statistic given below can be used to test whether the difference between two estimates of proportions is statistically significant. This test can be used to check the significance of the difference either between an agency level estimate and a national level estimate or between estimates for two different agencies. The test statistic is

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

where, $\hat{p}_1$ and $\hat{p}_2$ are the two survey estimates 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 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.