Technical Note on Source and Reliability of the Estimates for the 1963 Survey of the Aged

SOURCE OF THE DATA

In 1962 the Social Security Administration of the Department of Health, Education, and Welfare undertook a nationwide survey of the socio-economic characteristics of the aged, with the Bureau of the Census responsible for collecting and tabulating the information. Collection was carried out in 1963.

SURVEY DESIGN

1. Interview unit

The basic interview unit for the Survey was an "aged unit," which was defined as a married couple living together, either member of which was aged 62 or older, or a nonmarried person (including persons whose spouse had a usual residence elsewhere) who was aged 62 or older.

2. Sample design

A representative multistage area probability sample of the universe was used as the basis for the Survey. (The universe was composed of the civilian population aged 62 and over residing in the 50 States and the District of Columbia.) Ultimate sample units consisted of a representative subsample (one-half) of the Current Population Survey (CPS) sample[1] and the full Quarterly Household Survey (QHS), to create the sample for the 1963 Survey of the Aged. The ultimate sample units in the 1963 Survey sample, therefore, were selected after the following stages of sampling:

a. The standard metropolitan statistical areas and counties of the United States were grouped into about 1,900 primary sampling units (PSU).

b. These primary sampling units were then grouped into strata of one or more primary sampling units that are relatively homogeneous according to socio-economic characteristics. (There were 357 strata for the CPS and 333 for the QHS. The 333 represent an earlier phase of the evolution of the first-stage design of the CPS.)

c. Within each of the strata a single primary sampling unit was selected to represent the stratum. The 357 area CPS design is composed of 701 counties and independent cities and the 333 area QHS design of 641 counties and independent cities—with very substantial overlap between the two sets.

d. Within each of the primary sampling units a sample of housing units was selected. Selections were made from housing units with addresses in the 1960 Census listing books, building permit records, and current listings in area segments. A procedure was also used to provide a sample of units missed in the Census and for additional new units that would not be covered in the building permit records.

e. Within the sample units about 8,500 aged units consisting of about 11,000 aged persons was the expected sample size for the 1963 Survey of the Aged.

---

[1] For a complete description of the CPS sample see Bureau of the Census, *The Current Population Survey—A Report on Methodology*, Technical Paper No. 7, 1963. The QHS sample design is similar to the CPS design.

3. Survey collection stages

Because of the amount and complexity of the information being requested, the field survey was conducted in two stages. In the first stage, begun early in January 1963, respondents were identified and the Survey was explained to them. They were asked to provide their social security account number and such identifying information (not already available from the CPS or QHS interview) as is usually obtained on an application for an account number. Respondents were then given a questionnaire to complete and hold for an interviewer to pick up at a subsequent visit. In the second stage, completed in February 1963, the interviewer reviewed the answers on the self-administered form and filled in a second questionnaire relating to additional topics. Altogether, useful questionnaires were completed for 7,515 aged units, a completion rate of about 88 percent.

Persons in institutions were included (at half the sampling ratio used for the aged units in households). Only a limited amount of information—primarily on income and medical care—was requested. Where feasible, the answers were obtained directly from the respondent; in other cases, personnel of the institution and/or hospital records provided the needed detail.

4. Nature of information

Information was collected on such topics as income by source, work experience, assets and liabilities, health care costs, health insurance coverage, and living arrangements, as well as other facets of socio-economic status of persons aged 62 and over. Information in this detail will be available for the first time for a representative sample of all aged persons in the United States rather than only of OASDI beneficiaries.

The first-stage questionnaire covered health insurance, medical care costs, assets and debts, and income. The follow-up interview obtained more detail on these subjects and included additional questions on other subjects such as home tenure, living arrangements, housing and food expenses for those living alone, and on labor-force participation and work experience, as well as special questions for recent widows.

The information obtained from these two questionnaires was supplemented by information on household composition and family income from the CPS and QHS interviews as well as the Social Security Administration's record data described below.

5. Match with social security records

All cases were checked against the Social Security Administration's National Employee Index and other records to determine if the individual respondent had an account number or, by cross reference, if he had any possible claim status. All cases with a social security account number or a possible claim were then further screened to determine if a claim had been filed. Information was abstracted on type of benefit, primary insurance amount, benefits received during the survey year, year of first benefit, and other factors. Of the 8,378 units screened against OASDI records, positive identification as beneficiaries or nonbeneficiaries was completed on all but about 10, for which there was no evidence of a claim. Benefit record data were compiled on all but five of the 5,253 units identified as beneficiaries. Anyone who had received at least one benefit by the end of 1962 on an existing claim was classified as a beneficiary.

ESTIMATION

The estimates presented in this report are therefore derived from both the field collection and the OASDI program information. The basic data for each unit were weighted as follows:

1. Adjustment for noninterview

Some of the sample units did not provide usable questionnaires. For most households, however, there was some limited information that could be utilized in the noninterview adjustment process. Interviewed units having characteristics similar to those of noninterviewed units were selected at random and given a weighting factor of 2 to adjust for units not interviewed. The characteristics used in identifying similarities between interviewed and noninterviewed units were geographic area, size of aged unit (1 or 2 persons), age and race of the head of the unit, and sex for one-person units.

2. First-stage ratio estimation

The first stage of ratio estimation takes into account differences at the time of the last Census in the distribution by race and residence of the population estimated from the sample PSU's and that of the total population in each of the four major regions of the country. This stage of estimation has the effect of reducing somewhat the contribution to sampling variability arising from the selection of sample areas in the first stage of sampling.

3. Second-stage ratio estimation

The second-stage ratio estimation used the results of the 1963 Survey of the Aged after the noninterview adjustment and the first-stage ratio estimation to provide distribution of characteristics within age and race groups. Independent estimates of the civilian population aged 62 and over by race, sex, and age groups were then multiplied by the distributions derived from the Survey to create the estimates shown in this report. The number of OASDI beneficiaries calculated in this way was found to be less than 2 percent below the Social Security Administration estimate of the number with benefits in current-payment status and within 5 percent of the number with benefits in force—that is, on the rolls, whether or not a benefit had ever been received. At the end of 1962, more than 400,000 of the 14.5 million persons aged 62 or over with benefits in force were not actually receiving payments.

RELIABILITY OF THE ESTIMATES

Since the estimates in this report are based on a sample, they may differ somewhat from the figures that would have been obtained if all aged persons in the United States had been surveyed using the same schedules, instructions, and interviewers. As in any survey work, the results are subject to errors of response and of reporting as well as being subject to sampling variability. The standard error is primarily a measure of sampling variability, that is, of the variations that occur by chance because a sample rather than the whole of the population is surveyed. As calculated for this report, the standard error also partially measures the effect of response and enumeration errors but does not measure any systematic biases in the data. The chances are about 68 out of 100 that an estimate from the sample would differ from a complete census figure by less than the standard error. The chances are about 95 out of 100 that the difference would be less than twice the standard error.

1. Sampling variability of estimated numbers

The figures presented in table A are approximations to the standard errors of estimates of aged units and aged persons shown in this report. In order to derive standard errors that would be applicable to a wide variety of items and could be prepared at a moderate cost, a number of approximations were required. As a result, the table of standard errors provides an indication of the order of magnitude of the standard errors rather than the precise standard error for any specific item.

TABLE A.—Standard errors of estimated numbers of aged units or aged persons (68 chances out of 100)

Level of estimate (in thousands)	Standard error (in thousands)
50	13
100	18
250	28
500	40
1,000	54
2,500	85
5,000	120
7,500	130
10,000	140

2. Sampling variability of estimated percentages

The reliability of an estimated percentage, computed by using sample data for both numerator and denominator, depends upon both the size of the percentage and the size of the total on which the percentage is based. Estimated percentages are relatively more reliable than the corrsponding absolute estimates of the numerator of the percentage, particularly if the percentage is 50 percent or greater. Table B shows the standard errors of estimated percentages of aged units and aged persons.

TABLE B.—Standard error of estimated percentages of aged units or aged persons (68 chances out of 100)

Estimated percentage	Base of percentage (in thousands)					
	100	250	500	1,000	5,000	10,000
2 or 98	2.9	1.4	1.2	0.7	0.4	0.3
5 or 95	4.5	2.6	1.8	1.3	0.6	0.4
10 or 90	5.8	3.2	2.5	1.7	0.7	0.5
25 or 75	8.1	5.0	3.5	2.5	1.1	0.8
50	9.4	5.8	4.0	2.8	1.2	1.0

3. Sampling variability of estimated means

A rough approximation to the standard error of an estimated mean can be calculated from the distribution from which it was obtained by the following formula:

$$\sigma'_{\bar{x}} \doteq 80 \sqrt{\frac{\sum\limits_{i=1}^{c} n_i x_i^2}{n^2} - \frac{\bar{x}^2}{n}}$$

where $\sigma'_{\bar{x}}$ is the standard error of the estimated mean

n_i is the weighted number of cases reported for the ith class

$n = \sum\limits_{i=1}^{c} n_i$ ——the weighted total number of cases from which responses were obtained in all the c classes

x_i is the midpoint of the ith class

$\bar{x}$ is the mean of the distribution

Approximate standard errors of selected means for medical costs have been calculated through this formula by way of illustration. They are shown in table C.

Example: Table 1 of the article on medical care costs of the aged (pages 3-8 of the July issue of the *Bulletin*) shows the percentage distribution and mean medical costs incurred by married couples aged 65 and over, in the first column. In the formula for $\sigma'_{\bar{x}}$ given above, the value of n is equal to 4,765,000 (87.5% of the total 5,445,000); and $\bar{x}$ is \$442 (the entry labeled "mean cost"). The values of n_i are the numbers derived from the percentages multiplied by 4,765,000. The values of x_i were taken as the midpoint of the intervals shown in the stub. Hence, for the row labeled "none incurred," x_i has the value 0; for the next row, x_i has the value of \$50, and for the following row the value of x_i is \$150, etc. For the last row (\$2,500+), x_i has been estimated by subtracting the sum of $n_i x_i$ up to this point from $n_{\bar{x}}$, and dividing by n_i in this open-end interval (\$3,167).

The application of the formula for $\sigma'_{\bar{x}}$, then, shows that the estimated standard error of the mean medical costs incurred by couples aged 65 and over is about \$21. This means that the chances are about 68 out of 100 that the difference due to sampling variability between the estimated mean and the figure that would have been obtained in a complete census is less than \$21. The chances are about 95 out of 100 that the difference is less than \$42 or twice the standard error.

In comparing the results for two classes of the population, the difference between the two means will frequently be subject to greater sampling variability than other types of comparisons because the sample estimate of a mean is affected significantly by even a few extremely high values. In many cases, therefore, if comparisons are to be made of summary measures of two groups, it may be more useful to compare the medians rather than the means, since the medians are not affected as much by a few extreme observations.

4. Sampling variability of estimated medians

The variability of a median depends on the size of the base of the distribution and on the distribution on which the median is based. An approximate method for measuring the reliability of an estimated median is to determine an interval about the estimated median such that

TABLE C.—Illustrative examples of approximate standard errors calculated for selected means

Units aged 65 and over	Mean	Standard error of mean
Medical costs incurred by:		
Married couples	\$442	\$21
Nonmarried men	260	29
Nonmarried women	282	18
Married couples:		
In hospital	1,220	72
Not in hospital	233	10
Short-stay hospital costs incurred by:		
Married couples	588	46
With health insurance	612	55
Without health insurance	520	87

there is a stated degree of confidence that the true median lies within the interval.

As the first step in estimating the upper and lower limits of the interval about the median (that is, the confidence limits), compute one-half the number reporting the characteristic on which the median is based; designate this as N/2. Determine the standard error of N/2. Subtract this standard error from N/2. Cumulate the frequencies (in the table on which the median is based) up to the interval containing the difference between N/2 and its standard error and, by linear interpolation, obtain a value corresponding to this number.

In a similar manner, add the standard error to N/2, cumulate the frequencies in the table, and obtain a value corresponding to this number. The chances are about 68 out of 100 that a median based on results of a complete census would lie between these two values. The range for 95 chances out of 100 can be computed in a similar manner by multiplying the standard error by 2 before subtracting and adding to N/2. Interpolation to obtain the values corresponding to these numbers gives the confidence limits for the median.

5. Errors of response and nonreporting

In addition to sampling variability, as in any survey work, the results are also subject to errors of response and nonreporting. In many cases the data were based on memory rather than on records. In most income and expenditure data derived from field surveys the memory factor probably produces underestimates because of the tendency to forget minor or irregular sources of income and outlays. There are indications, however, that the tendency to underestimate income was less in this Survey. Other errors of reporting result from misrepresentation or misunderstanding as to the scope of a concept.

Incomplete responses to questions were handled in a variety of ways, depending on the question. Every effort, short of mechanical imputation, was made to obtain for each schedule a total income and a total medical expense figure, each built up from a detailed series of questions. In the case of income, for example, when an asset was reported and there was no entry for income accruing from assets of that type, income at the rate of 4 percent was recorded. If, on the other hand, the respondent reported on most income items but failed to make an entry (of an amount, "None," or "Don't Know") for certain infrequent income sources, such as unemployment insurance or individual annuities, this was tabulated as a zero entry. In the case of medical care, if the cost of care by doctors, dentists, and care in hospitals was recorded, but there was no entry at all for "Other" (miscellaneous) medical care, this item too was tabulated as a zero.

In addition to the results available from the match against the social security records, a series of comparisons with other reports on the number receiving income from specified sources is in process. Data on size of income, amount of assets, health insurance coverage, and hospital utilization are also being compared with those yielded by other field surveys. The results of these comparisons will be published in the detailed report on the Survey.