# INVESTIGATING DISCREPANCIES BETWEEN SOCIAL SECURITY ADMINISTRATION AND CURRENT POPULATION SURVEY BENEFIT DATA FOR 1972

Denton R. Vaughan Social Security Administration and Robert E. Yuskavage\* Bureau of Economic Analysis

This paper is the first in a series which will examine errors in the reporting of 1972 social security benefit income in the March 1973 Current Population Survey (CPS). Earlier studies conducted at Social Security by Bixby, Haber, and Finegar in connection with the 1963 and 1968 surveys of the aged [1,3,4] and by Projector -Bretz [9] for the March 1971 CPS have documented the quality of social security benefit reporting in the CPS and similar surveys. By-and-large, this previous work has emphasized the demographic and programmatic characteristics associated with misreporting. Although some mention has been made of the actual sources of reporting error [1,3], to our knowledge, presentation of a detailed analysis of this sort has not been undertaken. We intend to explore the actual sources of reporting error in some depth.

## DELIMITING THE PROBLEM OF ERROR

While the causes of misreporting in the CPS are, no doubt, many and varied, we consider the following to be among the more important:

- 1. failing or refusing to report benefit income or attributing it to another source;
- 2. reporting benefit income when it was not received;
- 3. reporting income from another source together with benefit income;
- 4. combining the benefits for two or more persons in the record of a single recipient;
- 5. misspecifying the elements needed to derive an annual benefit figure for the survey, such as, the dollar value of the monthly benefit, the number of months benefits were received, deductions, etc.;
- 6. estimating the annual amount, including more or less haphazard guesses, because of ignorance or uncooperativeness on the part of the respondent; and,
- 7. processing errors, from those made in recording information in the original interview through creation of the machine readable file.

The present analysis will focus primarily on the fifth type of error which results when the elements needed to derive the annual social security benefit amount have been misspecified. The third and sixth sources will also receive some attention. In all, six possibilities are considered:

- 1. mistiming benefit increases,
- 2. misspecifying the number of months of benefits,
- ignoring benefit increases,
  mishandling Supplemental Medical Insurance

(SMI) deductions,

- 5. Social Security-Railroad Retirement dual recipiency, 1/ and
- 6. rounding errors.

# THE DATA BASE AND STUDY UNIVERSE

The analysis draws on a data file which was obtained by matching the March 1973 CPS to Social Security earnings and benefit records.2/ Because of problems associated with the reporting of social security income for spouses and minor children in the CPS, we have focused our attention on persons 18 years or older who were the sole SSA beneficiary in their <u>recipiency</u> <u>unit.3</u>/ Among this group, we will further limit ourselves to recipients with social security income amounts in both the SSA and CPS record. Furthermore, individuals were also excluded if their CPS benefit income had to be imputed due to refusal or nonresponse.4/

Estimates for the number of beneficiaries and total benefits in the overall SSA-CPS universe age 18 and older and in the study universe were obtained using sample weights which incorporate some adjustments for nonmatches and mismatches. Population controls used in the weighting include adjustments for institutionalized, overseas, and decedent beneficiaries not eligible for interview. The derivation of the sample (initial raking) weights is described in detail elsewhere [15].

The study universe is significant in both its size and diversity. It consists of 9.9 million beneficiaries who received \$15.1 billion in benefits, each about 40 percent of the respective 1972 totals for the age 18 and older CPS-eligible beneficiary universe. As indicated in figure 1, the beneficiaries included in the study tend to be somewhat older and have a considerably greater representation of females than the out-of-scope segment. While retired worker recipients constitute the dominant beneficiary type in both groups, the proportion of widowed recipients was considerably higher, and that of spouse beneficiaries, considerably lower, in the study universe.

Figure 2 classifies study universe recipients and gross discrepant income by the direction of the observed SSA-CPS discrepancy. It shows that more than half of the recipients (54 percent) were "CPS overreporters," that is, the amount reported in the CPS was greater than in the SSA record. About two-fifths of the recipients were "CPS underreporters" (39 percent), with the amount in the CPS record understating the SSA benefit amount. For the remaining 7 percent, the survey and administrative amounts agreed exactly (1 percent) or fell within a narrow tolerance (\$10 or less). The gross dollar discrepancy associated with the misreporting in the CPS amounted to \$2.2





\* Totals do not add to 100 percent because unknown age and sex have not been included.

Figure 2. --Distribution of Cases in the Study Universe and Discrepant Income by Direction of Discrepancy



billion, or 15 percent of all benefit income paid to members of the study universe. Nearly twothirds of this discrepancy was overreported.

It should be noted that this preponderance of overreporting in the study population is not characteristic of the overall benefit reporting error found in the CPS [2, 9] or in other surveys of the beneficiary population carried out in a similar setting [1, 3, 4]. At the present time, all that can be said is that the net underreporting known to exist in the 1973 March Supplement data [2] does not stem from that part of the beneficiary population under consideration here.

## DETECTING SOURCES OF REPORTING ERROR

While the sources of disagreement between survey and administrative information may lie with either the survey or the administrative record system (and, in some instances, neither may be correct), in a great majority of cases, the administrative data are likely to yield a superior estimate of the "true" amount. Consequently, for the purposes of this paper, the information in the administrative record will be used as the criteria for evaluating the benefit data reported in the CPS.

<u>Methods of Detection.</u>— Two general approaches were used to detect the source of SSA-CPS benefit discrepancies.<u>5</u>/ For errors stemming from misspecifying months of benefits, mistiming or ignoring benefit increases, and mishandling SMI deductions, we attempted to reproduce the observed discrepancies by holding constant one or more of the known elements from the administrative system. Primarily because we considered it likely that the monthly benefit amount would be rounded in deriving the annual benefit income figure for the CPS (at a minimum, to the nearest dollar), we did not expect to be able to exactly replicate the observed discrepancies. Consequently, in testing for a match between the observed and simulated differences, a tolerance of \$6 to \$10 was permitted.

Discrepancies due to Social Security-Railroad Retirement dual recipiency and rounding errors were treated differently. It is not possible to distinguish between Social Security and Railroad Retirement income amounts in the CPS when the beneficiary receives payments from both sources. However, both the SSA and CPS records indicate the occurrence of dual recipiency. Consequently, discrepancies of CPS overreporters whose SSA record disclosed dual recipiency were attributed to the inclusion of Railroad Retirement income in the CPS. The mean value of the discrepancies detected in this fashion, when disaggregated by type of annuitant, closely approximates the corresponding average Railroad Retirement benefit for 1972.6/

CPS benefit amounts were tested for possible rounding errors if their ending digits were '000', '500', '00', or '50'. It was assumed that amounts ending in '000' might have been rounded to the

nearest thousand dollars, those ending in '500', to the nearest 500 dollars, and so forth. Observed discrepancies were attributed to rounding if differences of \$1 to \$500 were associated with ending digits of '000'. if errors of \$1 to \$250 were associated with ending digits of '500', and so on for amounts ending in '00', or '50'.

Multiple Handling Explanations.--In а considerable number of cases it proved possible to attribute an individual discrepancy to more than one source of error. In other words, a particular discrepancy was subject to alternative explanations. In order to simplify the analysis, a single source of error was assigned such cases based on priorities established by the overall prevalence of each category of error. Thus, timing errors were given first priority, ignoring the benefit increase was given second priority, months benefits received, third, and dual recipiency, fourth priority. Primary SMI errors were exempted from the criteria and assigned unconditionally, since they often seemed to be confused with small discrepancies also attributable to timing problems. As a matter of judgment, we felt that this resulted in a more plausible final distribution of the different types of error. Rounding errors were also exempted from the frequency criteria. Since the tests associated with detecting rounding errors were much less precise than those employed for the other possible sources of error, competing explanations were always given priority over rounding.

## FINDINGS

Table 1 displays the percentage distribution of all explained discrepant cases and respective discrepant benefit income for overreporters and underreporters by type of error, as assigned by the priority procedure discussed above. Sixtynine percent of all discrepant cases were explained by one or more of six hypothesized discrepancy models. The error models proved to be slightly better at detecting errors of overreporting (70 percent) than of underreporting (66 percent). The explained discrepancies accounted for proportionately less of the gross dollar misreporting, just over half the total. A somewhat greater share of the dollar discrepancy was explained among underreporters than overreporters.

Overreporters. -- Turning now to the sources of explained discrepancy for overreporters, we note the importance of errors in specifying the timing of benefit increases. Beneficiaries were granted a 20 percent across-the-board benefit increase in September 1972, effective in their October checks [11, 16]. This error model assumes that the respondent remembered the benefit increase but mistimed its occurrence. CPS overreporters, in committing this type of mistake, proceeded as though the higher benefit amount had been in effect four months or more. Misspecifying the timing of the benefit increase was, by far, the most frequent source of discrepancy detected among overreporters. Seventy-six percent of explained overreporter cases and 47 percent of the

	Discrepant Cases (in thousands)			Gross Dollar Discrepancy (in millions)		
Type of Primary Discrepancy	Total	Over- repor- ters	Under- repor- ters	Total	Over- repor- ters	Under- repor- ters
Total	9,227	5,328	3,899	2,233	1,456	777
Unexplaimed: Number or Amount Percent	2,896 31.4	1,586 29.8	1,311 33.6	1,075 48.1	717 49.2	358 46.0
Explained: Number or Amount Percent	6,331 68.6	3,742 70.2	2,588 66.4	1,158 51.9	739 50.8	419 54.0
PERCENT OF EXPLAINED						
Total	100.0	100.0	100.0	100.0	100.0	100.0
Mistiming Benefit Increase Misspecifying Months of Benefits Ignoring Benefit Increase Mishandling SMI* Deductions Railroad Recipiency Rounding Error	55.7 19.0 13.0 9.3 2.1 0.9	76.2 3.1 14.4 2.1 3.6 0.7	26.0 42.0 11.0 19.6 - 1.3	34.7 31.1 13.0 3.4 17.2 0.6	46.7 7.6 17.5 0.7 26.9 0.7	13.6 72.6 5.3 8.1 - 0.5

Table 1.--Percent Distribution of Explained Discrepant Cases and Associated Discrepant Income by Type of Primary Discrepancy

Note: Percents may not add to 100 due to rounding. \*Supplemental Medical Insurance

associated explained discrepant income was attributed to this error source.

A related type of discrepancy was produced by completely ignoring the effect of the benefit increase. In this instance, the respondent apparently failed to recall the benefit increase at all and proceeded as though the post-increase amount had been in effect for each month benefits were received. This kind of error accounts for an additional 14 percent of the explained overreporter discrepancies and 13 percent of the explained overreporter income discrepancy. More than 90 percent of the explained overreporting errors and almost 65 percent of the associated discrepant income were attributable to these two sources.

<u>Underreporters.</u>— The relative importance of the six error types was somewhat different for underreporters. Since nearly 90 percent of the recipients in the study universe received benefits for a full 12 months, errors in specifying the number of months of benefits were more likely to result in an underreport than an overreport. In fact, this kind of mistake was the most frequent source of explained discrepancy among underreporters. It was responsible for forty-two percent of the explained underreporter cases and nearly 73 percent of the associated underreported income.

Twenty-six percent of the explained underreports were associated with possible errors in specifying

timing of the benefit increases. Another 11 percent apparently resulted from ignoring the benefit increase. Together these two error sources accounted for about 20 percent of explained underreported SSA benefit income.

Supplemental Medical Insurance, a voluntary complement to the Medicare Health Insurance program, covers virtually all Old Age, Survivor, and Disability Insurance (OASDI) recipients age 65 and older [13]. In 1972, monthly premiums of slightly less than \$6 were deducted from their benefit checks [12]. The CPS interviewer was instructed to determine if the respondent was an SMI enrollee and, if so, to add the total premium to the net annual benefit amount. This adjustment provided the opportunity to make one of two kinds of errors: an overstatement of the CPS amount, for those who were not SMI enrollees but had an amount corresponding to the premium added to their actual benefits, or an understatement, for those who were SMI enrollees but did not have the premium amount added to their net benefit. Since nearly fourfifths of the study population was age 65 or older and, therefore, generally exposed only to the second kind of mistake, one would expect SMI errors to be associated primarily with underreporting in the CPS. In fact, mistakes in handling SMI premium deductions were attributed to nearly 20 percent of the explained underreports. However, since the errors were relatively small (about \$68 on the average), they accounted for only 8 percent of explained underreported benefits. Furthermore, mishandling SMI deductions was not an important source of reporting error among overreporters.

Dual Recipiency and Rounding. -- The two remaining sources of discrepancy, dissimilar in both nature and method of detection from the previous four, are Social Security - Railroad Retirement dual recipiency and rounding in the CPS reported amount. Dual recipiency was associated with only 4 percent of explained overreporter cases, but, since the discrepancies tended to be quite large, it accounted for 27 percent of explained overreported income. Rounding errors, on the other hand, proved to be a relatively unimportant source of discrepancy for both over- and underreporters. They were associated with only about 1 percent of the explained discrepant cases and less than 1 percent of the explained discrepant income.

Dollar Value of Explained and Unexplained Discrepancies .-- Reflecting the finding noted earlier that the proportion of explained misreported benefit income was less than the proportion of explained discrepant cases, table 2 shows that the dollar size of explained reporting errors was considerably smaller than that of the residual or unexplained discrepancies. Threequarters of the explained discrepancies were less than \$200, while this was so for only 39 percent of the unexplained differences. The average explained dollar difference (\$183) was slightly less than half that for the unexplained cases (\$371). Furthermore, the ratio of mean explained to unexplained error was considerably smaller for overreporters than underreporters (44 percent vs. 59 percent).

## SUMMARY AND FUTURE PLANS

We have shown that six relatively simple error models, when applied to the March 1973 CPS, explain somewhat more than two-thirds of the cases involving misreported benefit amounts and about half of the associated gross discrepant benefit income. However, the results also indicate that the error models were at a comparative disadvantage in detecting relatively large-sized discrepancies, particularly among overreporters.

In the future, we will develop and report on new models designed to detect the sources of these larger discrepancies. Another major task will involve the extension of the present analysis to multi-beneficiary units. In pursuing this facet of the research, we should also uncover the sources of CPS net underreporting. With the expectation of gaining useful insights into the background variables related to misreporting, we also intend to undertake analysis of the demographic and programmatic characteristics of the beneficiaries with explained and unexplained errors. In particular, size of discrepancy will be taken into account.

#### FOOTNOTES

\* The authors would like to express their thanks to Wendy Alvey, Ben Bridges, Dan Radner, and Fritz Scheuren for their helpful comments on earlier versions of this paper and to Beth Kilss and Helen Kearney for their assistance with the charts and tables.

# Table 2.--Percentage Distribution of Explained and Residual Discrepancies of CPS Overreporters and Underreporters by Size of Discrepancy

Size of Discrepancy (in dollars)	Total		Overreporters		Underreporters	
	Explained	Residual	Explained	Residual	Explained	Residual
Total Number	6,331	2,896	3,742	1,586	2,588	1,310
Percent of Total	100.0	100.0	100.0	100.0	100.0	100.0
1 to 99	45.1	20.4	38.1	7.6	55.2	35.8
100 to 199	29.9	18.9	33.2	15.4	25.2	23.2
200 to 299	14.2	17.0	18.5	21.3	8.0	11.8
300  to  499	5.3	21.2	5.3	29.2	5.3	11.6
500 to 999	3.2	15.5	2.0	17.6	4.9	12.9
1000 to 1499	1.5	4.7	1.6	5.7	1.3	3.6
1500 or More	0.8	2.1	1.2	3.1	0.3	0.9
Mean Discrepancy						
(in dollars)	183	371	197	452	162	273

(Numbers in Thousands)

Note: Detail may not add to total due to rounding.

- 1/ While the reporting of Railroad Retirement in the CPS is, in fact, a source of discrepancy between the survey amount and the social security administrative figure, it does not actually constitute misreporting in the CPS. The social security record simply does not include Railroad Retirement benefits.
- 2/ At the 1975 ASA meetings in Atlanta, several papers were presented which focused on the conceptual and reporting differences among the linked CPS, IRS, and SSA data sets for calendar year 1972. Some of the preliminary analyses presented included comparisons between matched CPS and IRS income information [5] and similar comparisons for SSA and IRS wage data [8]. For more information on the basic study, see [7], which appears elsewhere in these 1976 Proceedings. See also[14].
- 3/ The particular type of recipiency unit we will be using is known as the <u>dependency</u> <u>unit</u>, a classification routinely used at Social Security to arrange CPS household members in kinship groups consisting of persons generally considered interdependent under social insurance programs. Administratively, it is analogous to the subgroup of CPS household or family members that would be considered when determining eligibility and benefits for a single disabled or retired worker and his or her dependents or survivors [6].
- 4/ A small number of recipients with an SSA age of 18 or older (less than 30 sample cases) are also excluded because their age in the CPS was reported to be under 14, and, hence, income information was not obtained for them during the interview.
- 5/ A detailed description of the methods used to detect the individual sources of error is available from the authors on request. Their mailing address is: Division of Economic and Long-Range Studies, Office of Research and Statistics, Social Security Administration, 1875 Connecticut Avenue, N.W., Washington, D.C. 20009.
- 6/ For a description of the dual recipient population and benefit levels for 1972, see [10].

## REFERENCES

- [1] Bixby, L. E., Finegar, W., Grad, S., Kolodrubetz, W., Lauriat, P., and Murray, J. <u>Demographic and Economic Characteristics of the Aged: 1968 Social Security Survey</u>, Social Security Administration, Washington, D.C., 1975.
- [2] Bureau of the Census. <u>Current Population</u> <u>Reports</u>, "Money Income of Families and Persons in the United States," Series P-60, no. 90, Government Printing Office, Washington, D.C., 1973.
- [3] Epstein, L. E and Murray, J. <u>The Aged</u>, <u>Population of the United States: the 1963</u> <u>Social Security Survey of the Aged</u>, Social

Security Administration, Washington, D.C., 1967.

- [4] Haber, L. "Evaluating Response Error in the Reporting of the Income of the Aged: Benefit Income," <u>1966 American Statistical</u> <u>Association Proceedings, Social Statistics</u> Section, 1967.
- [5] Herriot, R. and Spiers, E. "Measuring the Impact on Income Statistics of Reporting Differences between the Current Population Survey and Administrative Sources," <u>1975</u> <u>American Statistical Association</u> <u>Proceedings</u>, Social Statistics Section, 1976, pp. 147-158.
- [6] Johnston, M. P. "Derivation of STATS Units," <u>Studies in Income</u> <u>Distribution</u>, <u>Report no. 1</u>, by D. Projector, M. T. Millea and K. Dymond, Social Security Administration, 1975, pp. 59-76.
- [7] Kilss, B., and Alvey, W. "Further Exploration of CPS-IRS-SSA Wage Reporting Differences for 1972," <u>1976</u> <u>American Statistical Association</u> <u>Proceedings</u>, <u>Social Statistics Section</u>.
- [8] Millea, M. T. and Kilss, B. "Exploration of Differences between Linked Social Security and Internal Revenue Service Wage Data for 1972," <u>1975 American Statistical Association Proceedings, Social Statistics Section</u>, 1976, pp. 138-146.
- [9] Projector, D. and Bretz, J. <u>Measurement of</u> <u>Transfer</u> <u>Income</u>, Social Security Administration, 1973.
- [10] Railroad Retirement Board. "Dual (RRA-SSA) and Special Guarantee Benefits as of December 31, 1972," <u>Research and Statistics Notes</u>, no. 7-73, April 1973.
- [11] Shulman, H. "OASDI Benefit Amounts under 1972 Amendments," <u>Social Security Bulletin</u>, 1973, vol. 36, no. 1, p. 30.
- [12] Social Security Administration. <u>Social</u> <u>Security Bulletin</u>, <u>Annual Statistical</u> <u>Supplement</u>, 1972.
- [13] Social Security Administration. <u>Medicare:</u> <u>Health Insurance for the Aged, Section 2,</u> <u>Enrollment, 1972</u>, Washington, D.C., 1975.
- [14] Scheuren, F., <u>et. al. Studies from</u> <u>Interagency Data Linkages</u>, Report No. 4, Social Security Administration, Washington, D.C., 1975.
- [15] Social Security Administration, <u>Studies from</u> <u>Interagency Data Linkages</u>, Report No. 8, Washington, D.C., (in preparation).
- [16] U.S. Senate Committee on Finance. Social Security Amendments of 1972, Government Printing Office, Washington, D.C., 1972.