EVALUATION OF WAGE AND SALARY INCOME REPORTING ON THE 1978 INCOME SURVEY DEVELOPMENT PROGRAM TEST PANEL

Angela Feldman, Charles Nelson, and John Coder, U.S. Bureau of the Census

INTRODUCTION

The Income Survey Development Program (ISDP) is a research and development program undertaken jointly by the Department of Health and Human Services and the Bureau of the Census. The objective of this program is to plan a comprehensive, continuing, household survey covering income, both cash and noncash components, and a broad base of other information related to participation and eligibility in a wide range of government assistance programs. This survey, it is hoped, will permit us to overcome some of the constraints of collecting income and related data on the March Current Population Survey (CPS) and thus provide an improved data base for program analysis and simulation. Work, both planned and in progress under the ISDP program, is directed toward establishing a framework for future surveys.

This paper focuses on a comparison between two estimates of annual wage and salary income derived in different ways. One estimate was derived by summing wage and salary income reported in four separate quarterly interviews. The other comparable estimate which was considered the "benchmark" for our comparisons was an annual figure collected in a fifth interview stressing the use of W-2 forms or other record sources. Although considerable case-by-case variation was found, overall means and medians derived from these two sources were found to be "similar," indicating a situation of compensating errors. More consistent reporting was noted for those respondents who worked the entire year, and those who used their W-2 records for reporting annual earnings.

SAMPLE AND GENERAL DESCRIPTION OF DATA COLLECTION PROCEDURES

The 1978 research panel was the second phase of ISDP field work. In addition to testing of several methods for collecting information on quarterly work experience, income, and related information for each household member 16 years old and over, the 1978 research panel was conducted for research of problems associated with a longitudinal survey and the theoretical aspects of multiframe sampling. The survey consisted of approximately 2,350 households selected from the general population within 60 selected ISDP primary sampling units. The sample consisted of a nationally representative area frame of about 1,950 households and a nationally representative list frame of about 400 current recipients of Federal Supplemental Security Income.

One of the major purposes of the initial April interview was to develop a detailed income profile for each household member. In the second and later interviews, data from previous visits were used to shorten and simplify the updating of the income and work experience information. Various supplements were added to the core of income and labor force questions for the July, October, and January interviews. The July questionnaire included a supplemental module covering data needed to estimate eligibility for some of the major government transfer programs. In October, the supplemental module concerned disability status, receipt of benefits from human service programs, marital history, educational attainment and other relevant personal history items. In January, the supplemental questions covered a variety of topics including wealth data for input to a life cycle earnings and savings model.

The final quarterly interview took place in April 1979. This interview was devoted to collecting data on taxes and income sources for which records are available or which appear on tax returns. Use of records to collect various sources of income was encouraged.

In the interviewed households information was collected for all household members 16 years old or over. Proxy interviews were accepted for members not present at the time of the interview. All interviews were conducted by personal visit by Census Bureau interviewers. Telephone callbacks could be used to obtain information not available in the initial personal interview.

COLLECTION OF QUARTERLY WAGE OR SALARY AND RELATED JOB INFORMATION

A large section of the questionnaire was designed to record detailed information about work experience over the three-month reference period including details on the occupation, industry, and earnings from each wage or salary job and each business and details on farm-related work, receipts, and expenses. This quarterly employment and earnings section was completed for each person indicating a "job or business" during the three-month period. Subsections were provided for jobs, nonfarm self-employed businesses or professional practices, and farms. Iterations were provided so that data for up to three different jobs and three separate nonfarm businesses could be recorded.

The items in the work experience portion of the questionnaire dealing with jobs were designed to provide several basic variables associated with quarterly earnings as well as general information about the nature of the job.

The section for recording amounts of wage or salary income received from a specific job was completed after this basic data had been recorded. There were several possible methods for recording the monthly wage and salary from a job. The method (or path through the questionnaire) that was used was dependent on two factors: 1) the availability and willingness of the respondent to use employee records, and 2) the method by which the respondent was paid, e.g., hourly wage, annual salary, monthly salary, commissions, etc. Workers who had employer records were asked separate questions concerning their quarterly earnings. Those who earned the same amount every payday during the quarter were asked only to provide that amount. Those who were not paid the same every payday were asked to provide: 1) the number of paychecks they received monthly, and 2) the amount of each paycheck. Those who were unwilling or unable to use employer records followed a path dependent on their method of payment. For workers paid by the hour the following information was

obtained: 1) hourly pay rate as of the end of the reference period, 2) overtime pay rate, 3) monthly hours of overtime worked, and 4) monthly amounts earned from tips, bonuses, or commissions. Salaried workers were asked: 1) salary as of the last day of the reference period, 2) overtime salary, and 3) monthly hours of overtime worked. Workers paid by other methods were simply asked to report their earnings for each month. In all cases, the ultimate goal was to derive monthly earnings for each job. These estimates were derived in the data processing phase using algorithms developed for each path in the wage or salary amounts area.

COLLECTION OF ANNUAL WAGE OR SALARY DATA

During the 1979 final quarterly interview, the following information was obtained for each of the sample person's 1978 employers: 1) whether or not the respondent had and was willing to use a W-2 form, 2) annual earnings according to the respondent's W-2 form, or an estimate if the W-2 was not available, and 3) annual payroll deductions for Federal income tax, State and local income tax, and Social Security or Railroad Retirement taxes.

MATCHING OF QUARTERLY DATA FILES

Before any analysis of earnings reporting could begin, five separate quarterly data files had to be created and matched together. These files consisted of a record for each sample person showing selected social and demographic characteristics and details concerning the job and earnings. Matching of the quarterly files was accomplished by first matching households using a control number unique to each sample household and then matching persons within households using person number, age, and sex.

A total of 4,401 persons were eligible for matching. After matching of the five data files, a total of 3,126 had been matched across all interviews. Of the 1,275 persons lost in matching only 193 were person nonmatches. The remaining 1,082 were lost because their household became a noninterview in one of the months after the initial visit. The complete household noninterview rate was 6.6 percent for the first interview in April 1978 and 15 percent at the end of the fifth interview period in April 1979. No intensive effort was made to resolve the 193 unmatched individuals. Causes of these nonmatches are most likely related to improperly coded identifiers such as household numbers, person numbers, etc.

QUARTERLY INCOME NONRESPONSE

Of the 2,135 jobs that were available for analysis, 754 or about 35 percent were excluded from our analysis because one or more of their quarterly income amounts were not derivable due to missing data problems. Of the 1,381 jobs that remained, 228 or about 17 percent had to be excluded because an annual estimate could not be provided during the fifth interview. Our final analysis consisted then of 1,153 jobs, or about 54 percent of the jobs that were originally available.

Table 1 shows the pattern of income nonresponse by type of interview. About 69 percent of the respondents who had some income nonresponse were only missing earnings information. Some of these, however, involved jobs which were held only in that one quarter. Type of interview seemed to have some effect on the percentage of income nonresponse. About 39 percent of the jobs for which a proxy interview was accepted in all four quarters had to be excluded from the final analysis because of income nonresponse. This compares to a figure of 28 percent for cases in which respondents were self-interviewed in all four quarters.

The implications for this particular missing data problem are significant and especially problematic in the terms of the longitudinal nature of the survey. First, a detailed look at the reasons for nonresponse should be started so that the problem can be better understood and solutions sought to lower the number of occurrences. Second, methods must be developed for between quarter consistency checks, edits, and, if needed, imputations of the missing data so that a longitudinal record can be developed. We have bearly scratched the surface in this area of research.

FINDINGS

A summary of the findings of this analysis of earnings reporting from the 1978 ISDP is shown in Tables 2, 3, and 4. Shown in Table 2 is a comparison of annual earnings distributions derived from the two estimation methods, by use of W-2 records in reporting annual earnings during the fifth interview. The underlined diagonal refers to the number of cases in which the two estimation methods produced earnings at the same point in the two distributions. Overall, 55 percent of the cases were on the diagonal. The percentage of cases on the diagonal for those who did not use W-2 forms in reporting annual earnings was 45 percent. Among those who used W-2 forms 61 percent were on the diagonal. This higher percentage of cases on the diagonal can most probably be attributed to two factors: 1) the rounding of annual estimates in cases where the W-2 form is not used, and 2) the probability that the same type of respondents who use W-2 records in reporting annual earnings are more likely to use records in reporting quarterly earnings.

Shown in Tables 3 and 4 are comparisons of median earnings, and percent deviations between annual estimates derived by the two estimation methods by selected factors thought to be potentially important in the reporting of these two earnings amounts. Overall the median quarterly derived estimate of 1978 earnings per job was $5,395 or about 94 percent of the median based on the annual estimate that stressed the use of W-2 forms. The following is a summary of the results, by the selected factors, as shown on Tables 3 and 4.

Method of Payment.--The ratio of median quarterly summed earnings to the median annual estimate was higher among hourly paid workers (1.01) than salaried workers (.93). Overall though, reporting quality would seem to be better among salaried workers, as shown by the fact that 52 percent of the summed quarterly earnings for salaried jobs were within 10 percent of the annual estimates compared to a figure of 33.5 percent for

hourly jobs. The higher ratio of the medians for hourly jobs can be attributed to what may be called the effect of compensating errors. While 32 percent of the summed quarterly income values for salaried jobs were less than the annual estimate by more than 10 percent, only 16 percent were more than 10 percent greater than the annual estimate. This phenomenon is not true, however, for hourly wage jobs. For these jobs about 34 percent of the summed quarterly estimates were more than 10 percent greater than the annual estimate and 32 percent were more than 10 percent less than the annual estimate.

Frequency of Payment.--The ratio of quarterly summed mean earnings to the median annual estimate was higher for workers paid weekly and biweekly (.98 and 1.01) than those paid monthly and bimonthly (.92 and .93).

Most of this difference in ratios can be attributed to the fact that most hourly workers are paid either weekly or biweekly, and most salaried workers are paid either monthly or bimonthly. Thus, as one would expect, the ratio of quarterly to annual medians among those paid weekly and biweekly are very close to the ratio for hourly paid workers (1.01) and the ratios of those paid monthly and bimonthly are very close to the ratio for salaried workers (.93).

Number of Months with Job.--Those who worked at the same job the entire 12 months had a very high ratio of summed quarterly to annual earnings estimate (.97), and a relatively high percentage of quarterly totals within 10 percent of annual estimates (61 percent). The reporting of quarterly income was considerably worse for respondents who did not work the entire quarter as shown by the ratio of summed quarterly to annual median (1.12) and the low percentage of quarterly totals within 10 percent of the annual estimate (17 percent). Most noticeable in this group is the percentage of cases in which summed quarterly earnings were more than 50 percent higher than the annual estimate (25 percent). In looking at these quarterly records, it was found that many were cases in which the respondent held more than one job during the year, and the interviewer inadvertently used the same job number for both the original and the new job. Since the quarterly earnings were summed by job number, this problem resulted in distorted quarterly income figures for these people.

Use of W-2 Records for Annual Estimates.--The ratio of median quarterly summed earnings to the median annual estimate was .97 among respondents who used records in reporting annual amounts, and 1.14 for those who did not. It would seem that the reporting quality of quarterly income was better among those people who used records in reporting their annual earnings. This makes some sense, as one would expect that the type of person who used W-2 records in reporting annual earnings would also take more care in reporting accurate quarterly earnings. Of course, this effect is somewhat mitigated by the fact that there is certain to be some rounding in the reporting of annual income in cases where a W-2 form is not used. It would also appear that there is a positive correlation between level of income and use of W-2 forms in estimating annual income. The median reported earnings jobs in which a W-2 form was used was $7,454, which is considerably higher than the median for jobs in which the W-2 form was not used ($2,275).

Respondent Type.--The ratio of summed quarterly median to annual estimated median was .94 for respondents who were interviewed by proxy in all four quarters. As one might expect, there was a higher percentage of summed quarterly earnings within 10 percent of the annual estimate among respondents who were self-interviewed all four quarters (43 percent) than there was for respondents who were interviewed by proxy all four quarters (31 percent).

SPECIAL PROBLEM CASES

Of the 1,153 jobs available for analysis, the summed quarterly earnings were more than 30 percent different from the annual estimate in 383 cases, or about 33 percent of the time. A number of these quarterly records were reviewed on a case-by-case basis to ascertain if there was any one problem, or group of problems, that was causing these large discrepancies to occur.

In looking at these cases, it was found that there were three main reasons for these substantial differences between the two methods of estimating annual earnings. They were:

1. Interviewer or keying errors.--These errors took many forms. Typical cases included: A) the transposition of digits, so that, for instance, a $15,000 annual salary would become $51,000, and B) the addition of an extra digit or the truncation of the last digit, so that, for instance, a $2.40 hourly wage would become $24.00, or a $200 weekly paycheck would become $20. Perhaps the most common of these errors, however, occurred among workers who were paid biweekly. For these people, their biweekly salary was supposed to be entered in the amount box. There was a tendency by the interviewers to record their weekly salary instead. It would seem that the effects of these problems could be alleviated by using edits that would include consistency checks between quarters to make sure that earnings from the same job do not fluctuate greatly because of an easily corrected error.

2. Reporting take home pay instead of gross pay.--During the April 1979 roundup interview, in addition to reporting their annual pay from each employer according to their W-2 form, respondents were also asked to estimate deductions from that pay, so that the respondent's take home pay could be derived. Of the respondents who substantially underreported their quarterly earnings, about half the cases could be traced to the reporting of take home pay instead of gross pay. The effects of this problem could be lessened by stressing to the interviewers during training to probe in all questionable cases to make sure that proper amounts are reported.

3. Using the same job number for more than one job.--A large majority of the cases in which quarterly summed earnings were substantially greater than the annual estimate could be traced to this problem. As was mentioned earlier, there were particular problems in the reporting of income in cases where the respondent changed jobs during the year. In these cases, there was a tendency to assign the original job number to the new jobs. This problem could also be alleviated by using edits that check for consistency between quarters.

CONCLUSIONS AND RECOMMENDATIONS

This study has uncovered some important aspects regarding the collection of monthly earnings data in a household survey using a 3-month recall period but we have a long way to go. The importance of continued evaluation and analysis of the earnings data cannot be underestimated. The accuracy of modeling and simulation of tax and transfer programs will be affected more by the quality of the earnings data than perhaps any other single data item. Several significant problems have been found as the result of this study which point up the need for improvement. While the summary measures showed some surprising consistency, significant case-by-case variation was present. Significant nonresponse problems were also detected and the initial magnitudes established. These are symptoms of several serious data collection and processing problems each of which will require analysis and the development of many new, innovative data collection systems.

TABLE 1. SUMMARY OF QUARTERLY INCOME NONRESPONSE BY TYPE OF INTERVIEW: 1978 ISDP

	Total	Self All Quarterly Interviews	Proxy All Quarterly Interviews	Some Self-Some Proxy Interviews
Total	2,135	808	338	989
Income NA in at least one quarter	754	224	131	399
Income NA in only one quarter	519	154	81	284
NA in first quarter	90	26	20	44
NA in second quarter	168	53	20	95
NA in third quarter	125	40	9	76
NA in fourth quarter	136	35	32	69
Income NA in two quarters	136	41	20	75
Income NA in three quarters	65	20	19	26
Income NA in all four quarters	34	9	11	14
Annual income NA in fifth interview	228	75	36	117
Total used for analysis	1,153	509	171	473

TABLE 2. ANNUAL EARNINGS DERIVED FROM QUARTERLY INTERVIEWS AND ANNUAL REPORTED EARNINGS FROM INTERVIEW PERIOD NO. 5 BY USE OF RECORDS: 1978 ISDP

Annual Reported Earnings	Annual Earnings Derived from Quarterly Interviews											
	Total	Less than $500	$500 to $999	$1,000 to $1,999	$2,000 to $2,999	$3,000 to $4,999	$5,000 to $7,499	$7,500 to $9,999	$10,000 to $14,999	$15,000 to $19,999	$20,000 to $24,999	$25,000 and over
Total	1,153	129	88	132	99	109	123	116	159	104	52	42
Less than $500	148	84	30	19	9	4	2	0	0	0	0	0
$500 to $999	103	32	34	24	8	3	1	1	0	0	0	0
$1,000 to $1,999	114	9	19	54	20	6	5	1	0	0	0	0
$2,000 to $2,999	83	2	2	19	43	14	2	1	0	0	0	0
$3,000 to $4,999	97	0	1	13	12	53	13	5	0	0	0	0
$5,000 to $7,499	116	1	1	2	6	24	67	14	1	0	0	0
$7,500 to $9,999	110	0	1	1	1	3	27	59	17	1	0	0
$10,000 to $14,999	163	0	0	0	0	2	4	32	107	17	1	0
$15,000 to $19,999	110	0	0	0	0	0	2	2	30	65	11	0
$20,000 to $24,999	58	0	0	0	0	0	0	0	4	19	30	5
$25,000 and over	51	1	0	0	0	0	0	1	0	2	10	37

TABLE 2. ANNUAL EARNINGS DERIVED FROM QUARTERLY INTERVIEWS AND ANNUAL REPORTED EARNINGS FROM INTERVIEW PERIOD NO. 5 BY USE OF RECORDS: 1978 ISDP

Annual Reported Earnings	Annual Earnings Derived from Quarterly Interviews											
	Total	Less than $500	$500 to $999	$1,000 to $1,999	$2,000 to $2,999	$3,000 to $4,999	$5,000 to $7,499	$7,500 to $9,999	$10,000 to $14,999	$15,000 to $19,999	$20,000 to $24,999	$25,000 and over
Total, Used W-2.....	731	57	40	67	55	68	87	85	115	83	42	32
Less than $500.....	55	36	10	5	2	2	0	0	0	0	0	0
$500 to $999.....	43	13	16	10	3	0	1	0	0	0	0	0
$1,000 to $1,999...	64	6	11	31	11	1	3	1	0	0	0	0
$2,000 to $2,999...	54	1	1	13	27	10	1	1	0	0	0	0
$3,000 to $4,999...	68	0	1	7	8	40	8	4	0	0	0	0
$5,000 to $7,499...	83	1	1	0	3	14	53	10	1	0	0	0
$7,500 to $9,999...	82	0	0	1	1	1	17	51	11	0	0	0
$10,000 to $14,999.	114	0	0	0	0	0	2	18	80	13	1	0
$15,000 to $19,999.	86	0	0	0	0	0	2	0	20	55	9	0
$20,000 to $24,999.	46	0	0	0	0	0	0	0	3	14	25	4
$25,000 and over...	36	0	0	0	0	0	0	0	0	1	7	28
Total, Did not use W-2.....	422	72	48	65	44	41	36	31	44	21	10	10
Less than $500.....	93	48	20	14	7	2	2	0	0	0	0	0
$500 to $999.....	60	19	18	14	5	3	0	1	0	0	0	0
$1,000 to $1,999...	50	3	8	23	9	5	2	0	0	0	0	0
$2,000 to $2,999...	29	1	1	6	16	4	1	0	0	0	0	0
$3,000 to $4,999...	29	0	0	6	4	13	5	1	0	0	0	0
$5,000 to $7,499...	33	0	0	2	3	10	14	4	0	0	0	0
$7,500 to $9,999...	28	0	1	0	0	2	10	8	6	1	0	0
$10,000 to $14,999.	49	0	0	0	0	2	2	14	27	4	0	0
$15,000 to $19,999.	24	0	0	0	0	0	0	2	10	10	2	0
$20,000 to $24,999.	12	0	0	0	0	0	0	0	1	5	5	1
$25,000 and over...	15	1	0	0	0	0	0	1	0	1	3	9

TABLE 3. PERCENT DEVIATION OF QUARTERLY EARNINGS FROM ANNUAL REPORTED VALUES: 1978 ISDP

	Total	Salaried	Hourly Wage	Held Job 12 Months	Held Job Less Than 12 Months	All Self	All Proxy	Some Self, Some Proxy
Total.....	1,153	314	786	556	597	509	171	473
Less than -50 percent.....	83	10	69	3	80	34	9	40
-30 to -50 percent.....	79	23	52	10	69	30	17	32
-15 to -30 percent.....	126	43	78	52	74	48	24	54
-10 to -15 percent.....	84	26	56	54	30	34	17	33
- 5 to -10 percent.....	102	29	66	80	22	50	14	38
- 5 to + 5 percent.....	272	115	149	220	52	133	29	110
+ 5 to +10 percent.....	69	19	48	41	28	36	10	23
+10 to +15 percent.....	41	9	29	19	22	25	2	14
+15 to +30 percent.....	77	19	54	39	38	34	11	32
+30 to +50 percent.....	52	8	40	21	31	19	11	22
+50 percent or more.....	168	13	145	17	151	66	27	75

TABLE 4. COMPARISON OF ANNUAL EARNINGS DERIVED FROM QUARTERLY INTERVIEWS AND ANNUAL EARNINGS REPORTED FROM INTERVIEW NO. 5: 1978 ISDP EARNINGS

	Number	Median from Quarterly Interivews	Median Annual Earnings	Ratio	Mean Absolute Difference
Total	1,153	5,395	5,768	.94	1,384
Time Job Held					
Under 12 months, total	597	1,742	1,558	1.12	1,325
3 months or less	230	652	713	.91	750
4 to 6 months	187	1,874	1,644	1.14	1,299
7 to 11 months	180	4,906	4,656	1.05	2,088
12 months	556	11,735	12,100	.97	1,447
How Paid					
Hourly wage	786	3,423	3,394	1.01	1,162
Salaried	314	11,230	12,070	.93	1,931
Other pay	53	4,749	5,468	.87	1,435
How Often Paid					
Every week	583	4,124	4,225	.98	1,309
Every 2 weeks	309	7,749	7,662	1.01	1,432
Once a month	120	7,499	8,181	.92	1,440
Twice a month	111	8,976	9,624	.93	1,438
Other	28	411	349	1.18	2,071
Weeks Missed Without Pay					
No weeks missed	888	6,276	6,637	.95	1,427
One or more weeks missed	265	3,370	3,405	.99	1,240
Type of Respondent					
All self	509	5,581	5,956	.94	1,167
All proxy	171	5,249	4,768	1.10	1,826
Some self, some proxy	473	5,224	5,482	.95	1,459
Use of Records for Annual Estimate					
Used W-2	731	7,255	7,454	.97	1,352
Did not use W-2	422	2,590	2,275	1.14	1,439