Volume Title: Studies in Income and Wealth

Volume Author/Editor: Conference on Research in Income and Wealth

Volume Publisher: NBER

Volume ISBN: 0-870-14168-6

Volume URL: http://www.nber.org/books/unkn51-1

Publication Date: 1951

Chapter Title: Coordination of Old-Age and Survivors Insurance Wage Data with Those from Other Sources

Chapter Author: Benjamin J. Mandel

Chapter URL: http://www.nber.org/chapters/c5729

Chapter pages in book: (p. 464 - 480)

Part VIII

Coordination of Old-Age

The opinions expressed in this article do not necessarily reflect the views of the Bureau of Old-Age and Survivors Insurance.

and Survivors Insurance Wage Data with Those from Other Sources

Benjamin J. Mandel
BUREAU OF OLD-AGE AND SURVIVORS INSURANCE

A Need for Coordination Studies

As is generally known, national wage and salary statistics come from various sources. Aggregates compiled by the Department of Commerce, Office of Business Economics, are released as a component of the official national income estimates. The Bureaus of Old-Age and Survivors Insurance and of Employment Security compile aggregates of wages and salaries received in employments covered by the social security program, and the Bureau of Internal Revenue compiles aggregates reported on income tax returns. Distributions of wages and salaries by size are prepared from income surveys by the Bureau of the Census and by a few

other governmental agencies, including the Bureaus of Agricultural Economics and of Labor Statistics, and the Federal Reserve Board; from income tax returns by the Bureau of Internal Revenue; and from employers' reports on taxable earnings of individual workers by the Bureau of Old-Age and Survivors Insurance. Wages and salaries classified by size are expected from the 1950 Census of Population.

Fundamental differences in these wage and salary figures stem from differences in coverage, definitions, and methods of compilation. The OASI wage data are confined to wages and salaries in covered employments, and employers report only the first $3,000 paid to an employee in a year. Census wage and salary data, on the contrary, cover all gainful employment. Unlike the OASI figures, which are based on employers' quarterly wage reports, however, they are compiled from information on earnings reported by respondents interviewed at home who often have to rely on their memory of wages and salaries received during an entire year. Moreover, the information is often obtained from a member of the household who does not have the full facts about the wage earner's income. Census wage figures, furthermore, are generally recorded up to $10,000, and earnings of that amount or more are recorded as $10,000 and over. The BIR wage and salary figures, although not limited to any specific employments, inevitably exclude the wages and salaries received by the majority of individuals whose earnings are below $500 a year (or some other amount prescribed by law). Furthermore, when the income is reported on joint husband-wife returns, it is difficult to distinguish the earnings of each.

In view of this variety of type, scope, and source of national wage and salary data, the published statistics are bound to differ even for the same employment groups or similar sectors of the economy. For intelligent use of the various wage series, information is needed on the extent of their comparability and their differences. As yet, the interrelations among the series have not been analyzed comprehensively.

The goal of studies should be to accumulate quantitative in-

formation that would make it possible to construct annually a complete and reliable distribution of individuals or families by wage and salary earnings. If these size distributions are reliable they should, of course, yield aggregates approximating the official estimates of wage and salary payments. The studies should also yield data that will account for differences between completely adjusted size distribution series and size distributions of wages derived directly from income tax returns, OASI wage records, and special field surveys. As a byproduct proper coordination should furnish information from one series to supplement that from another; for example, earnings in noncovered employments should be made available for workers in industries covered by social security.

B General Methods of Wage Reconciliation Studies

From Census Bureau income surveys, BOASI employer reports, and BIR income tax returns the amounts reported for the same individual in each file for the same year can be compared directly and differences analyzed. In such a matching study it is, of course, necessary first of all to identify positively the same individual in the different files. The next step is to compare the wage figures for identical individuals in the various sources. Differences in the coverage, concepts, and reliability of the information raise technical problems. For example, there is no immediate answer to the question whether all or only part of the earnings reported on an income tax return is covered under OASI. Moreover, even if coverage is known to be the same, but the reported amounts for an individual differ significantly, there is no direct evidence which figure is more nearly correct.

C Recent Experience

The BOASI, in cooperation with other agencies, has made several pilot wage coordination studies in the last few years.[1] Valuable

[1] The records of these agencies, of course, are confidential and proper safeguards were established to avoid disclosure.

information, which can be used as a guide in the future, is thus available on the nature of the problems, the procedures, potential results, and estimated costs of wage coordination studies. This paper describes briefly each study, summarizing the procedures, problems, and results, and discusses the value of wage coordination studies and suggests the types that should be undertaken in the future.

Of the 5 studies on which data are available, 4 were small pilot studies, each involving several hundred cases; the fifth, on a larger scale, was a coordination project dealing with 1944 income tax data.

1 *Census and Old-Age and Survivors Insurance Study, 1944*

In January 1945 the names of 458 persons 14 and older on a list of 200 households in Atlanta and Buffalo from the Census records for the April 1944 Monthly Report on the Labor Force were checked against OASI records to determine whether it was possible to identify individuals satisfactorily by the name, year of birth, and address—types of information readily available from Census surveys and in the OASI records. As only about 22 percent of the names were definitely identified, and a large proportion of the remaining names showed presumptive covered employment in the Census record, it was concluded that satisfactory identification could not be established by relying only on the name, year of birth, and address.

2 *Identification of Veterans in Old-Age and Survivors Insurance Files*

This conclusion was confirmed by another but successful pilot identification study. As part of a study on health and disability by the National Research Council it was necessary to identify a list of World War II veterans in the OASI records. From an initial sample of 275 names supplied by the National Research Council, 199 individuals, 72 percent, were positively identified from the name and full date of birth—largely because the day and month as well as the year of birth were given, items of infor-

mation not available in the earlier Census-OASI check. Since double-checking established the fact that the remaining 28 percent of the persons were definitely not represented in the OASI records it was concluded that matching on the basis of the complete date of birth, not just year of birth, and name yields fairly satisfactory results.

3 *Census and Old-Age and Survivors Insurance Pilot Study, 1946*

The first two studies had limited objectives; there was no attempt to match wage and salary or other types of information. The pilot wage coordination study of records from the 1946 Census income survey and the records maintained by the BOASI brought to bear considerable detailed information. It had three purposes: (1) to test the problems and procedures in identifying individuals from a great deal more information than in the two earlier studies; (2) to match wage information and to develop methods for determining areas of difference between the Census and OASI wage data; (3) to develop a method for adjusting Census income survey wage distributions on the basis of OASI wage data, and vice versa.

The Census Bureau selected a systematic random sample of schedules containing the names of 249 individuals 14 years and older from the file of schedules for an area enumerated in connection with the April 1947 Consumer Income Supplement (1946 incomes) to the Survey of the Labor Force, Population, and Housing. Since the sample was taken from the file for a single area, not the United States as a whole, figures based on it cannot be considered to have significance for the entire country.

a *Identification of individuals*

In identifying individuals in the OASI records, the name, address, and day, month, and year of birth, as shown on the Census schedule, were relied on. When given, the mother's and father's names also were used. Of the 249 names in the sample, 59 were shown on the Census schedule to have reported no civilian work in 1946 and could not be found in the OASI files. Practically all

were of school age, married women, or elderly; consequently, it was assumed that they were actually not in the labor force and were definitely unidentifiable.

The other 190 names were considered potentially identifiable. When checked against the OASI records, 161, or 85 percent, were definitely identified. The remaining 29 names, or 15 percent, could not be identified, although some did civilian work in 1946, according to the Census schedule. The high rate of identification is attributable primarily to the fact that all the relevant identifying information on the Census schedule was accessible for matching with corresponding OASI records; moreover, the area canvassed contained a high proportion of workers in covered employment.

b *Comparison of wage and salary data*

An individual's wages could be considered for matching only if the employment coverage of the two records was the same. It was, therefore, necessary to determine first whether the 161 workers whose names had been identified had received wages only from covered employment. This could be determined quite accurately in practically all instances from the information recorded on the Census schedule on the number of employers during the year, the class of worker, the major industry, and the major occupation of the person in the job at which most money was earned in 1946. In one or two instances exact agreement on the amount of wages was considered a definite indication of covered employment only, regardless of the information on the class of worker and industry on the Census schedule. In 19 cases it was assumed that the worker had covered employment only because the OASI records showed some wages and the Census schedule showed none. Of the 161 persons whose names had been identified in the OASI records 127 were classified as having had only covered employment in 1946. Of the remaining 34 individuals, 26 did not show wages or salaries in either record, and 8 showed noncovered employment. All 34 were excluded from the wage-matching study (Accounting Table).

Accounting Table

Distribution of 249 Census Individuals in 1946 Pilot Coordination Study by Selected Classifications

Classification	*Number of individuals*	% *of total*
Total in study	249	100.0
Not in labor force and not identifiable	59	23.7
In labor force, identifiable	190	76.3
In identification study	190	100.0
Not identified	29	15.3
Identified	161	84.7
Total in wage matching study	161	100.0
No 1946 wages in either OASI or Census files	26	16.1
Noncovered or both covered and noncovered employment	8	5.0
Covered employment only	127	78.9
Total in wage comparison study	127	100.0
Satisfactory agreement	43	33.9
Implied agreement	7	5.5
Disagreement	77	60.6
Total wage disagreements	77	100.0
Census estimate: weekly wage multiplied by 50 or 52 weeks	44	57.1
Other Census estimate	10	13.0
No Census wages but some OASI wages	19	24.7
No OASI wages but some Census wages	4	5.2

After separating those with covered employment only from the remaining group of identified individuals, the two sets of wage figures could be compared for 127 wage earners. To afford a basis for determining the degree of agreement and disagreement 'agreement' had to be defined. Bona fide differences in wages could exist because taxable wages under the OASI program are limited to the first $3,000 a year. Three categories of agreement or disagreement were set up: *Satisfactory agreement* was defined as a difference of less than $100 in annual wages or salary. The $100 dividing line was selected arbitrarily as being a relatively small difference when the data are classified by broad wage groups. *Implied agreement* was defined as a Census excess of $100 or more over the OASI wages, provided the OASI records showed taxable wages of $3,000 (because of the $3,000 limitation under the OASI program) and only covered employment

was indicated by the Census record. *Disagreement* included differences of plus or minus $100 or more when the OASI record showed earnings of less than $3,000.

Comparison of the two wage records for each of the 127 individuals for whom the Census schedule indicated covered employment only showed that there was 'satisfactory agreement' on the wages of 43 individuals, or 34 percent of the total (Accounting Table). The average Census wage for these 43 workers was $1,435 in 1946, ranging from $55 to $3,000. Agreement within the dollar on 19 of the 43 wage reports suggests that these individuals, or the persons giving the Census enumerator information about them, used some record of earnings, such as their income tax.

'Implied agreement' was found in the wages for 7 individuals or about 5 percent of the total. The OASI record showed exactly $3,000 and the Census record more than $3,000. Adding the number of workers in the 'satisfactory' and the 'implied agreement' categories gave 50 wage records in agreement or nearly 40 percent of the 127 cases in the wage comparisons.

The remaining 77 wage records were classified as disagreements, because in each case the difference was $100 or larger and the OASI record showed wages under $3,000. In 44 cases the Census record showed a lower wage; in fact, in 19 it did not show any wages. In the other 33 cases the Census wages were higher than the OASI wages; among them the OASI record did not show any wages in 4 instances.

Since in all these cases of disagreement it was fairly definitely established that only covered employment was involved, both sets of records were studied in detail to determine the most probable reasons; the actual reasons could have been established only by reinterview with the individuals themselves. As the worker's complete wage record under OASI for 1937–46 could be examined, his previous levels of earnings could be compared with that for 1946. The OASI wage record showed also the number of employers and the industry from which the wages were received. From the Census record the sex, age, major industry, and part-time work status classifications could be

checked. By combining these items of information, circumstantial evidence on the most likely reasons for differences could be obtained.

The most important probable reason seemed to be that the Census enumerator or the respondent estimated the annual wages on the income schedule by multiplying the current weekly wage by 50 or 52 weeks. Dividing the annual wages by 50 or 52 indicated the further probability that in many cases weekly wages were rounded to the nearest $5 or $10. The Census Bureau instructs the enumerator to get reasonable estimates if actual amounts are not ascertainable. The inference that in most of these instances weekly wages were multiplied by the number of weeks worked was tested by determining whether the Census wages were exact multiples of 50 or 52, whether full or part-time employment was reported, and whether the 1946 level in the OASI wage record was consistent with the level of preceding years. While it is quite possible for annual wages to be exact multiples of 50 or 52, the proportion of such cases in the Census survey seemed to be considerably higher than experience with many individual social security annual wage records indicates. In 44 of the 77 cases of disagreement, or nearly three-fifths, the Census wage report showed figures that were exact multiples of 50 or 52. For example, there were 3 at $1,800, which is exactly $36 a week for 50 weeks; 3 at $2,000, which is exactly $40 weekly for 50 weeks; and 3 each at $1,560 and $2,340, which are weekly wages of $30 and $45 respectively, for 52 weeks. A net excess of Census wages for these 44 workers suggested that the overstatement was probably due partly to the fact that some workers did not actually have 50 or 52 full weeks of employment in the year and partly to the upward rounding of weekly wages.

A second, but relatively less important, reason for differences as far as this sample was concerned may have been that the respondent guessed at the worker's wages, or estimated them by methods other than multiplying by 50 or 52 weeks. This reason was deduced largely by comparing the OASI and Census wages for 1946 with OASI wages for preceding years. The 1946 OASI wages were more consistent with the wage levels for preceding

years than the 1946 Census wages. It is also possible that some of these differences are due to conceptual differences. For example, OASI covers wages in kind, and the respondent might not consider such payments as wages in his reply to the Census enumerator.

A third, and also relatively insignificant presumptive reason, was the absence of wages in the OASI record for a small percentage of individuals because of incomplete or incorrect reporting of wages by employers. Wages on about 1 percent of all the wage reports for 1946 were still not posted to the workers' accounts at the time of the study. The 4 cases for which the Census report showed wages in covered employment but the OASI record did not show any wages could be explained by this presumption.

c *Adjusting Census wage distributions*

The OASI wage records can thus be a useful source for adjusting respondents' errors or estimates in the Census data, at least for annual wages in covered employment. On the other hand, the Census data can be useful for adjusting OASI wage records that do not show total earnings because of the $3,000 taxable wage provision or are incomplete because of unidentified or tardily processed wage reports.

A method was developed for basing such adjustments on the findings in this pilot study. It was necessary to determine which of the two sets of 127 comparable individual wage figures was more acceptable as a basis for adjusting the other. For the group of 43 wage reports where agreement was satisfactory, it would make little or no difference whether the Census or the OASI figures were taken as the standard. The 7 Census figures in the implied agreement group were more acceptable, since the OASI figures were limited to the first $3,000. On the other hand, among the 77 disagreements, OASI figures were generally considered more acceptable for the basic series in all except the 4 cases where Census but not OASI wages were recorded. While this greater confidence in the OASI figures seems justified for the foregoing reasons, the Census estimate may have more accurately represented workers' actual earnings in a few instances.

From the 127 more nearly correct wage figures, a frequency distribution by wage groups was set up. A synthesis of Census and OASI wage figures, it was then compared with the corresponding original Census wage distribution for the same 127 individuals, and ratios were computed for each wage group. It

Annual Wages & Salaries	Original Census Series	Adj. Series (OASI & Census data)	Ratio (Adj. ÷ Census original)
Total	127	127	1.00
0	19	0	...
$1–499	5	20	4.00
500–999	13	17	1.31
1,000–1,499	19	23	1.21
1,500–1,999	35	31	0.89
2,000–2,499	18	16	0.89
2,500–2,999	6	9	1.50
3,000 & over	12	11	0.92

seemed reasonable to assume that if the sample was representative, the ratios could be relied upon for adjusting the Census or the OASI wage distributions for workers in covered employments. In the absence of other information, the same adjustment ratios might be used for workers in employments not covered by OASI, on the assumption that the bias would be relatively small. To test the validity of the latter assumption a wage coordination study might be made with the BIR records for workers in noncovered employments, although the technical problems previously mentioned and explained in greater detail later would be encountered.

Obviously a Census-OASI wage coordination study would not furnish adjustment ratios for groups above the $3,000 and over group, because OASI wages are limited to the first $3,000. For a complete adjustment, therefore, it seems essential to bring into the comparison income tax records for identical individuals with earnings of $3,000 or more a year.

Even though the sample was not representative of the United States, and quantitative conclusions cannot be drawn from the results, the table suggests two principal tendencies as far as the workers in the pilot sample are concerned. First, the Census survey understated the number of persons with wage earnings.

Secondly, in the particular area covered by the test it appears to have understated most the number for workers in the under $1,500 class intervals.

4 *Income Tax and Old-Age and Survivors Insurance Wage Coordination Study, 1944*

An attempt by the BOASI in 1945–46 to match 1944 wage information from income tax returns with its own wage records yielded some information on the problems of coordinating other wage sources with income tax records. As this study was designed primarily to get, on a sample basis, information on the marital status and number of dependents of workers in covered employment, the approach was quite different from the one followed in the Census-OASI 1946 wage matching study.

To get a sample representative of all persons who reported a social security account number on their income tax returns, the sample was chosen from the regular BIR sample used for tabulations. Approximately 19,000 returns were selected by taking those with an account number having the digit 2 or 7 in the sixth place of the nine-digit number.

a *Identification of individuals*

Even in the first stage, identifying the same individuals in the two files, difficult problems were encountered on which only limited effort could be expended because of cost. Experience indicated that, in addition to the account number, some supplementary information, at least the person's name or age, is essential for satisfactory identification. For example, from the sample of about 19,000 income tax returns, as many as 1,100 were rejected because on the basis of control records kept in the BOASI the account number on the return had never been issued.[2] In some cases the number reported, even though it had been issued, was probably not the correct number of the individual filing the return, so that identification was incorrect. Moreover, since

2 The social security numbering system permits issuing a maximum of one billion unduplicated account numbers. Since about 100 million numbers have been assigned, a person unable to remember his number might give a number not yet issued.

the income tax sample was a sample of returns, not individuals, there were some joint returns of two or more persons with incomes; they had to be rejected. Even the joint single-income returns remaining were difficult to handle, since there was no assurance that the single account number listed on the return represented the individual whose wages were being matched. The conclusion was that identification by the social security account number alone, though relatively less time consuming than other methods, is unsatisfactory, especially when income tax returns are the source because numerous joint returns with two or more incomes are bound to give rise to complications.

b *Comparing wages*

Even though identification was considered unsatisfactory for wage adjustment purposes, the coordination study was not abandoned inasmuch as it could still serve some of its original purposes. The matching of wage information was attempted only as a byproduct. By the time the wage matching stage was reached, the sample was reduced from its original size of 19,000 to about 13,000 by the exclusion of the noncomparable groups already mentioned, such as joint returns with two or more incomes and returns showing nonexistent account numbers.

One major difficulty encountered in the wage matching operations was the lack of information in the income tax file to show whether the wages for identified individuals were received from covered employments only or from noncovered as well. As indicated above, to determine satisfactorily agreement or disagreement about wages, comparisons must be confined to individuals employed in covered industries. Another problem was the unavailability of the complete OASI wage records for some workers, due to late reporting by some employers. (This problem can generally be overcome, as was done in the 1946 Census-OASI study, by a supplementary search of the wage records.)

Even if it were possible to overcome the problems of identification and of determining employment status, the results from this particular income tax coordination study could not be considered representative of all wage and salary workers in the

country, because the sample originated from the BIR records on income tax filers. It was, therefore, subject to the qualifications inherent in that population, such as the exclusion of individuals who should have filed an income tax return but did not and individuals who were not required to file because of the $500 tax exemption in 1944. An additional bias was introduced into the original sample of 19,000 by the elimination of about 6,000 returns of unknown characteristics.

c *Evaluation of results*

Of the 13,000 individual wage records in the matching study, 11,000 or 85 percent reported wages on income tax returns either exceeding or equaling those in the OASI records. It was impossible to determine the number that agreed exactly or the amount of the difference for those reporting larger wages. Nothing definite could be concluded about the wages for this group because wages reported for income tax purposes could exceed OASI wages without real disagreement. Some or all the wages reported on tax returns may have been earned in noncovered employment, or some reports of OASI wages may have been delayed. Despite this and other difficulties already mentioned, one significant conclusion was drawn from the wage comparisons for the remaining 2,000 or 15 percent of the cases where OASI wages exceeded those reported on tax returns. After allowing for the fact that OASI wages may actually exceed tax return wages because of conceptual and other differences, such as the inclusion of wages in kind in the OASI record, and after allowing for a reasonable proportion of mismatched records, the tax return wages appeared to be understated for approximately 10 percent of the 13,000 individual records studied. The average understatement in 1944 was approximately $300.

5 *California Employment Security and Old-Age and Survivors Insurance Wage Data, 1947*

The most recent wage coordination study was in connection with a service furnished by the BOASI to the California Department of Labor in its statistical program. A random sample of

805 was selected from a total file of about 40,000 California punch cards representing a 1 percent sample of workers with unemployment insurance wage credits in 1947. An attempt was made to identify these 805 workers in the OASI files by the social security account number, then to match their wage records under both the employment security and the OASI programs. The purpose was twofold: to check again whether the social security account number alone, without name, was adequate for identification and to determine the extent to which the wage records under the two systems agreed.

a *Identification of individuals*

Identification was regarded as having been established when both the account number and the wages agreed for at least one quarter in the year. Unlike the Income Tax Study, agreement on account number alone was not considered satisfactory. In the initial attempt at identification, about 100 or 12 percent of the 805 accounts, while matched to an account number in the OASI files, appeared to be mismatched when further comparisons were made with quarterly wages.[3] In these 12 percent of the cases no single 1947 quarterly wage amount agreed; therefore they were considered presumptively mismatched because it is reasonable to assume that wage information furnished under the two programs by the same employer should have agreed for at least one quarter.

The 705 cases, 88 percent of the sample, for which the account number and the wages for at least one quarter agreed were considered positively identified.

b *Extent of agreement on wages*

As in the Census study, satisfactory agreement on wages was defined as a difference of less than $100 in annual wages. On this basis, 628 or nearly 90 percent of the 1947 wage figures for identical workers agreed. Of the remaining 77 cases, the OASI wages exceeded the California wages in 42 instances; the Cali-

[3] Subsequent checks with the California agency reduced the number still unidentified to 64, or 8 percent of the total.

fornia wages exceeded the OASI wages in 35 instances. From previous experience with OASI and state employment security wage data, these differences can be ascribed to such reasons as differences in the wages reported by the same employers under the different programs; delayed wage reports by some employers under either or both programs; coverage differences; and inconsistencies in reporting the state of employment by some employers having establishments in more than one state. On the basis of this comparison study, which indicated 90 percent agreement on wages, it was concluded that fairly substantial agreement exists in the wages for identical individuals as recorded in the OASI and California State Labor Department files.

D Summary and Conclusions

Three conclusions may be drawn from the experience with these pilot coordination studies of wage and salary data. Foremost is the evidence that while there are numerous technical problems to overcome, reasonable, fairly reliable, and workable methods are available for coordinating wage data in OASI, Census, and BIR records. Secondly, there is no single highly accurate and complete national series on wage size distributions. For developing such a series and for a more intelligent and complete utilization of the various wage size statistics, a continuing 3-way Census-OASI-BIR wage coordination study based on a reliable sample is essential. All the evidence points to the fact that the wage coordination studies, if properly conducted, will produce worth while results. Finally, to get the best results most efficiently, the starting point should be the Census file which aims at including the entire wage earning population. The match should be first with the corresponding OASI records, then with the corresponding income tax records.