

THE STATISTICAL ADEQUACY OF EMPLOYERS' OCCUPATIONAL RECORDS

KATHERINE D. WOOD *

THE LARGE number of agencies, both governmental and private, now collecting and publishing occupational statistics is evidence of the wide recognition of the importance of such information. The purposes for which occupational data are gathered, however, vary widely, as do the methods of collection. The Department of Labor, for example, is interested in occupational differentials in wage rates; the United States Employment Service is concerned with the demand for and supply of workers with different occupational qualifications; the insurance companies are concerned with death rates for different occupations because of their importance in determining risks. The Social Security Board has a potential interest in occupational statistics for the purpose of determining occupational differentials in connection with payments of both old-age insurance and unemployment compensation.

The needs of the Board are described as potential rather than immediate, because the first steps in the administration of the Social Security Act have necessarily been carried on without reference to occupational differences. As the administration of the act continues and matures and as need develops for greater precision in forecasting demands upon reserves, the importance of occupational information may become so great as to warrant increased emphasis upon occupational statistics. Furthermore, it is possible that the occupational records of employers subject to old-age insurance may some day prove a valuable source of current occupational information for other Government agencies. Occupational information collected by the States for purposes of unemployment compensation is based on employers' records and so also involves the problem of evaluating employer records as a source of occupational data.

Looking to the future, the Bureau of Old-Age Insurance has recently made a sample study to determine the adequacy, for statistical purposes,

* Bureau of Old-Age Insurance, Analysis Division. Mrs. Wood is the representative of the Social Security Board on the Technical Sub-Committee of the Committee of the Central Statistical Board and the American Statistical Association on Occupational Classification.

of the occupational records of employers. One of the means for testing the adequacy of these records is to compare them with occupational information from other recognized sources—for example, occupations reported by members of the family in a census type of enumeration, one of the most common methods of collecting occupational information. The comparability of occupational information obtained by the census method with that from employers' records is of particular importance to the Social Security Board because at the present time studies of death rates or the incidence of unemployment by occupation must be based upon occupational data provided by the census of population. The occupation of the wage earner at the time of death or at age 65 is, however, obtained from the employer,¹ as are the occupational data obtained by the States in the administration of unemployment compensation. For several reasons it might be expected that the information obtained from the employer might not be comparable with that from a census, in which information is supplied by the family. It is well known, for example, that a considerable amount of occupational upgrading exists in the returns obtained from the family and that the employer may have occupational concepts entirely different from those of the workers or their families. Moreover, vague or unsatisfactory occupational terms which are difficult if not impossible to code properly are frequently obtained in a census. All these factors may affect the base figures to such an extent that the measurement of death rates or incidence of unemployment will be far from accurate if dependent upon occupational data obtained by the two different methods.

The Philadelphia Study

An answer to this question of the comparability of data obtained from these two sources can be determined only when the occupational designations from each source are available in such a manner as to permit a comparison of the two returns

¹ Such information in death claims is sometimes also obtained from the undertaker, but the report of the employer is considered to be more reliable.

for the same person and presumably for the same job. Such a comparison was recently made by the Bureau of Old-Age Insurance in cooperation with the Industrial Research Department of the University of Pennsylvania.²

Table 1.—Extent of agreement between occupation reported by member of worker's household and that reported by his employer, according to person interviewed, Philadelphia sample, 1938¹

Person interviewed	Number			Percent		
	Total	Similar	Different	Total	Similar	Different
All persons.....	4,516	2,914	1,602	100.0	64.5	35.5
Worker.....	648	450	198	100.0	69.4	30.6
Spouse.....	1,835	1,193	642	100.0	65.0	35.0
Other family member.....	1,420	897	523	100.0	62.9	37.1
Other.....	86	42	44	100.0	48.8	51.2
Not specified.....	521	332	189	100.0	63.7	36.3

¹ Differences measured by use of a code of 233 items.

In the summer of 1938 a household survey of employment and unemployment was made in Philadelphia, covering a sample of approximately 10 percent of all the households in the city. The schedule of the survey included questions as to the present occupation and industry of all employed workers in these households; for approximately 10,000 employed persons in this sample, excluding persons in domestic service and certain other occupations not covered by old-age insurance, the name and address of the employer was also obtained. The 10,000 wage earners were employed by over 3,000 employers. Since time and expense limited the number of employers who could be reached, a first selection was made of employers who had at least 10 workers represented in the sample; a random sample of the employers having from 1 to 10 workers among the 10,000 was then chosen in order to include some smaller employers in the study.³ The final sample for which the two occupational designations were obtained consisted of approximately 4,500 wage earners employed by over 400 employers. These employers represented a wide range of industries, but the exclusion of certain occupational groups not covered by the Social Secu-

² The transcription of records and other clerical work on this study was done by the WPA Area Statistical Office in Philadelphia. Miss Claire Casey assisted in the coding and tabulation of the results and in the analysis of the reasons for the differences found. Mr. Leo K. Frankel, Jr., assisted in the interviewing of employers.

³ The number of workers in the sample reporting themselves as employed by a given employer should not be confused with the number of persons on an employer's pay roll; it is probably true, however, that the employers having over 10 workers in the sample were relatively large firms.

Act, especially domestic workers in private homes, agricultural workers, and government employees, limits the number of industries somewhat, as does also the proportionate overrepresentation of relatively large firms in the sample. For these reasons, and because of its limited size, the sample is not representative of the entire city, although many occupations are nevertheless represented in sufficient numbers to permit statistical analysis.

The terminology used by employers in describing the occupation of their employees differed frequently from that used by the workers or members of their families. Many of these differences in the basic information are interesting but not necessarily of statistical significance as measured by the coded result. An employer may use an occupational designation different from that used by the worker or his family, but the meaning of the terms used may not be materially different, and both may be included in the same code. In this study, for example, a person returned by the household as an "electrician" was called a "first-class wireman" by the employer; a "metal polisher" was described by the employer as a "buffer," and a "wood finisher" as a "hand sander." In each instance both of the occupational titles received the same occupational code. A fairly detailed code of 233 items, developed by the Industrial Research Department of the University of Pennsylvania for the Philadelphia area, was used in this study. To test the extent to which differences in occupational designations were due to a specific code, use was made also of the social-economic classification developed by the Bureau of the Census⁴ and used exclusively in the 1937 unemployment census. The resulting differences (see tables 1 and 2) show the importance of the code itself in the measurement of differences.⁵

When the more detailed of the two codes was applied to both sets of occupational designations, 35.5 percent of the individuals received differing designations. When the social-economic code of only 9 broad occupational groups was used, the designations differed for 21.7 percent of the individuals. The differences resulting from the application of the more detailed code of 233 items

⁴ United States Department of Commerce, Bureau of the Census, *A Social-Economic Grouping of Gainful Workers in the United States, 1930-1938*.

⁵ The exclusion of certain occupations, such as those in domestic service in private families, means that some of the occupational codes in each classification have not been used. This does not, however, affect the differences for the particular occupations included. If the complete range of occupations had been included, the total differences might have been somewhat altered.

is perhaps less surprising than those resulting from the use of the code of only 9 major groups.

The extent of these statistical differences is sufficiently large to arouse interest. Why were one-third of 880 individuals reported in the household survey as skilled workers reported differently by their employers, many of them as semiskilled? Were they skilled workers who had taken jobs of an apparently lower occupational rank, or was there conspicuous upgrading on the part of the workers or their families in giving the occupation? Was there perhaps a downgrading on the part of employers? If they were skilled workers in semi-skilled jobs, then the occupational statistics resulting from household enumeration may be said to reflect the potential labor supply in terms of usual occupations, whereas the employers' reports show those occupations in which the workers are actually engaged. If either upgrading by the family or downgrading by the employer is the explanation, then one or the other of the returns is actually incorrect. Which of these, or a combination of these and possibly other, factors accounted for the existing differences can be answered decisively only by much more thorough investigation. The differences for each of the main social-economic classes show a possible lack of comparability even when occupational returns are classified in such broad groups.

The application of the more detailed occupational code gives greater insight into the explanation of such differences. The wide variation in the extent of the differences for many of the 233 occupational titles is conspicuous. Frequent differences arose in the designation of persons reported by the household as machinists, mechanics, electricians, plumbers, welders, compositors, linotype and monotype operators, engineers and firemen (stationary), cutters, foremen, accountants and auditors, bookkeepers, office-appliance operators, secretaries, shipping and receiving clerks, stenographers, typists, railroad switchmen, flagmen and yardmen, technical engineers, and surveyors. At the other extreme, few differences appeared in designation of persons reported by the family as operatives in textile and clothing manufacture, including the knitters, loom fixers and others; the cigarette and tobacco workers; watchmen and guards; conductors and motormen; railroad trainmen, firemen, conductors and engineers; taxi and bus drivers; elevator operators; laundry

workers (not domestic); and waiters (not domestic).

The percentage of differences seemed to be proportionately high among certain of the skilled occupations and for some of the so-called white collar occupations such as accountants, secretaries, and stenographers, in which the possibility of either downgrading by the employer or upgrading by the family is clearly inherent. An inspection of the employers' returns for those individuals indicated the likelihood of frequent upgrading on the part of the household. In some cases, however, a reasonable doubt might arise as to whether it was not actually a skilled worker who was required by an employer for a certain job even though the job itself was coded as semiskilled.

Since no job analysis was made to determine whether the employer's designation accurately described the particular jobs in which workers were reported, an attempt was made to determine the reasons for the differences by a comparison of

Table 2.—Extent of agreement between occupation reported by member of worker's household and that reported by his employer, by social-economic group, Philadelphia sample, 1938¹

Social-economic group	Number			Percent ²		
	Total	Simi- lar	Differ- ent	Total	Simi- lar	Differ- ent
All groups.....	4,516	3,537	979	100.0	78.3	21.7
Professional.....	111	72	39	100.0	64.9	35.1
Farmers (owners and tenants), Proprietors, managers, and officials (except farmers).....	111	60	45	100.0	59.5	40.5
Clerks and kindred workers.....	1,214	1,009	205	100.0	83.1	16.9
Skilled workers and foremen.....	880	578	302	100.0	65.7	34.3
Semiskilled workers.....	1,709	1,480	229	100.0	83.7	16.3
Farm laborers.....	4	2	2			
Other laborers.....	245	172	73	100.0	70.2	29.8
Servant classes.....	182	158	24	100.0	86.8	13.2

¹ Differences measured by use of a code of 9 major occupational groups.

² Percentages calculated on totals of 25 or more cases.

the two occupational returns for each individual for whom differences arose. This involved considerable judgment; while the results can hardly be called statistically reliable, they give some indication of the proportion of differences which can be explained and the relative importance of some of the reasons for differences. Clearly the code system itself is partly responsible. The use of the more detailed code of 233 items is estimated to account for 13 percent of the total differences. The tendency of workers or their families to upgrade their occupational level is estimated to account for approximately 18 percent. Other ex-

planations which appeared significant were, in the order of their relative importance: (1) terminology peculiar to a particular employer or plant; (2) the existence of more than one reasonable possibility in the description of the individual or his job, such, for example, as a "foreman" who might also have been a "molder" by occupation; (3) unintentional downgrading by workers or their families; (4) downgrading by the employer; and (5) returns which were too vague or unsatisfactory to code accurately. In the final analysis it was possible to give reasonable explanations for only half of the differences. A job description or more complete information would be necessary to determine the reason for many of the differences in these doubtful cases and to decide which of the two sources gave the more accurate description of the individual in his current job.

Present Limitation of Occupational Statistics

This study of some 5,000 individuals for whom two independent occupational reports were available clearly indicates the need for further consideration of methods for improving occupation returns. A serious question may be raised as to the value of certain occupational data unless greater surety of their accuracy can be established. Even though the 20 percent difference does not prove that 20 percent of the household entries were inaccurate, reasonable doubt as to their accuracy certainly exists. If, when a code of only nine categories is used, one-fifth of the families' occupational designations are actually inaccurate then the value of the results is certainly in question. It might be argued that, in studies in which the information is similarly collected, the bias is generally in the same direction and therefore that for some purposes the statistical data describing a group of wage earners are adequate and the comparisons valid for different series of occupational data collected under similar conditions. Yet more emphasis should undoubtedly be placed upon the necessity for improving the returns, no matter what their source, and upon standardization of occupational terminology, so that a certain title in common usage has uniformly the same meaning and is so used in designating a given job no matter where the job occurs. Standardization of terminology is, of course, much less simple than it sounds, partly because the language habits of individuals are strong and because the

meaning of the same word may vary in different areas. Even among employers, usage of uniform terminology would be difficult to achieve because of the large number of existing job titles sometimes used for identical or only slightly varying work. In spite of these difficulties, much improvement in occupational statistics might be achieved by greater emphasis upon the quality of the original material.

From the standpoint of the social security program, the present study clearly shows that occupational data secured from employers' records cannot be measured against data obtained from household surveys and more particularly from the census of population except with great caution in interpretation. Particularly in such measurements as death rates and the incidence of unemployment by occupation, care should be taken to allow for possible differences in the original data. If the number of workers reporting occupations classified as skilled is overstated in a census, then reports from employers as to the number of skilled workers currently unemployed will result in a fictitiously low rate of unemployment for skilled workers. Similar distortion will follow throughout the occupational range, and the measurement of unemployment may be distressingly inaccurate for certain occupations.

Recognition of the limitations of most occupational statistics is perhaps the first principle to learn in their use. Only agencies which go into a plant and actually determine the duties and skill required by a certain job can be sure of reasonably good results for a detailed occupational classification, particularly one that attempts to classify workers by skill. In this connection it might be noted that the answer to a question in the present study as to the duties of the worker was of great assistance in coding the occupational entries. Even brief answers by the employer solved many difficult coding problems. It is not practical, however, to ask such a question in all studies, and it would be of little aid to query some members of a family or the worker himself if there was a deliberate desire to misstate occupation.

Consideration of methods to improve the accuracy of occupational statistics warrants increasing attention if occupational data continue to be of increasing importance to Government agencies and private research organizations.