

The Development of a New Geographic Coding System for the Continuous Work History Sample

*by Linda M. Dill, Barry V. Bye, and Cheryl I. Williams**

This article describes the statistical development of the geographic coding system used to identify worker location for the Continuous Work History Sample. The new system—which is planned for implementation for data year 1993—will provide more accurate geographic distributions of workers within a residence concept than the old system could provide within an employer location concept. The article also presents the results of a pilot study that tested the operational aspects of the new system. The results provide some preliminary estimates of the effect of the revised codes on the geographic distribution of workers.

*Division of Earnings Statistics and Analysis, Office of Research and Statistics, Social Security Administration.

To provide statistical data needed for the agency's planning and policy analysis, the Social Security Administration (SSA) has, since the early days of the its operations, collected information from employers about their business location. Employers who operated more than one place of business (multiunit employers) were asked to use—on a voluntary basis—the Establishment Reporting Plan (ERP). Under this plan, the employer gave SSA a list showing the location, industrial activity, and approximate number of employees of each establishment. All other employers reported as single unit entities and were classified under one primary geographic location. This information was used in connection with the employer's annual wage reports to code workers included in the Continuous Work History Sample (CWHS)—a major source of Social Security program and workforce data—by geography (Smith 1989).

For three decades, many multiunit employers elected the ERP to report wages. By the mid-1970's, signs of declining participation became evident. In 1978, when SSA changed from quarterly to annual wage reporting and asked employers to use magnetic media for their reports, a further significant decline in reporting occurred. Geographic classification of workers included in the CWHS was affected by these declines. As employers moved from the ERP to single unit reporting, a significant "superficial" shift of workers' geographic location to States in which corporate headquarters were located was noted. The decline in ERP participation led to its discontinuance in 1991.

The most promising alternative to the ERP may be the use of employee residence address data from the W-2 annual wage reporting forms for employers reporting by magnetic media. These data should more accurately identify the geographic location of workers for larger firms than those now obtained from SSA's employer-based coding files. Geographic codes for workers employed by smaller firms that are not using magnetic media would continue to be obtained from SSA's Single Unit Code File. Until a new system is in

place for 1993 data, geographic data for the CWS files will be derived from the most current coding files.

This article describes the new geographic coding system that will be implemented for 1993 data. Also presented are the results of a pilot study that indicate the kind of changes one might expect.

Background

Each employer whose business is subject to the Federal Insurance Contributions Act (FICA) must complete a Form SS-4, Application for Employer Identification Number (EIN). For statistical purposes, SSA has coded geographic information from the SS-4 since the first 3 million employers were enumerated in 1936-37. Although the Internal Revenue Service (IRS) began assigning these numbers in 1950, the geographic coding operation has remained with SSA.

The geographic classification used by SSA for coding employers is based on the *Employer-Worker Classification Geographic Code Manual*, which is published and maintained by SSA's Office of Research and Statistics.¹ The geographic code is a 5-digit code. The first two digits identify a major geographic area (a State, the District of Columbia, a U.S. territory) and the last three digits identify the county or county equivalent (that is, Alaska borough or Census area, independent city, municipality, or parish). The scope of geographic coding includes the United States and outlying areas (American Samoa, Guam, Northern Mariana Islands, Puerto Rico, and the Virgin Islands). Employers who have military (APO/FPO) or foreign addresses are also included.

Single unit employers, who primarily operate only one place of business, are assigned one geographic code. It is based on the primary physical location of the business when it is provided by the employer; otherwise, the mailing address is used. Although most employers are single establishment firms, a small number of employers are large, multiestablishment firms.

These multiunit employers have more than one business location in different geographic areas. To facilitate the processing of their wage reports, SSA, in 1943, instituted the voluntary Establishment Reporting Plan. The ERP was also designed to provide for the collection and classification of byproduct statistics on various employer characteristics, including the primary geographic location of employers by their establishment or other reporting unit (Dill 1992).

Prior to 1991, multiestablishment employers (100 employees or more; 50 or more prior to 1984) who chose the ERP to report wages grouped their employees by establishment within the annual wage report, identified each group by an employer-assigned establishment number, and filed a master list of these numbers (Form SSA-5019, Establishment Reporting Plan, List of Establishments or Reporting Units) with the Social Security Administration. Each establishment (or reporting unit) was individually classified by its primary physical location. Multiunit employers who did not participate in the plan were considered to be a single unit for classification purposes because information on the physical location of each individual establishment was not available. The geographic code, therefore, was usually based on the physical location of the home office.

Decline in ERP Participation

For many years, ERP participation was quite successful. However, indications in the mid-1970's showed declining participation, and the changeover from quarterly to annual wage reporting of FICA covered earnings (beginning with tax year 1978) resulted in a significant drop in participation. Attempts were made to revitalize employer participation in the ERP, but the structure of the annual wage reporting system placed constraints on administration of the plan. Difficulties associated with annual reporting in the early years delayed the receipt of the data needed to identify and correct

GEOGRAPHIC DATA FROM THE CWS

ANNUAL STATISTICAL SUPPLEMENT
TO THE SOCIAL SECURITY BULLETIN

EARNINGS AND EMPLOYMENT DATA
FOR WAGE AND SALARY WORKERS
COVERED UNDER SOCIAL SECURITY,
BY STATE AND COUNTY

STATE STATISTICS

REGIONAL WORK FORCE
CHARACTERISTICS AND MIGRATION
DATA: A HANDBOOK ON THE
SOCIAL SECURITY

CONTINUOUS WORK HISTORY
SAMPLE AND ITS APPLICATION

COUNTY BUSINESS PATTERNS

ERP reporting problems in a timely manner. In addition, SSA's solicitation of employer participation in magnetic media wage reporting resulted in further erosion of the plan.

The deterioration of multiunit reporting has had an impact on geographic classification of workers in the CWS. There have been significant shifts of workers to States where the headquarters for large companies are located. These shifts are merely artifacts of the decline in ERP participation and do not represent real changes in the geography of the workforce.

In 1991, the Office of Research and Statistics discontinued the ERP based on the results of a study conducted to analyze the deterioration of the plan's effectiveness due to the decline in participation and the relevant impact on geographic data in the CWS (Dill, Enis, and Williams 1991). Establishment Reporting Plan participation was analyzed at four points in time (1971, 1976, 1981, and 1986). The results showed a decline in employer participation from 1971 to 1986 of almost 81 percent. Also, total jobs reported under an establishment showed an overall decrease of 71.1 percent for the same period. This

decline was attributed primarily to the onset of annual wage reporting in 1978 and magnetic media reporting.

The deterioration of establishment reporting generated major negative impacts on geographic data in the CWHS because multiunit firms were being classified as single unit entities under home office addresses. These home offices were generally located in major cities or in the State of Delaware where many large companies incorporate for tax purposes. For example, the Bureau of Labor Statistics² reported total average annual employment for New York State in 1981 that was more than 5 million fewer jobs than the total jobs reported in the CWHS for the same year. For Delaware, total jobs reported in the CWHS for 1981 were almost double the annual average employment reported by the Bureau of Labor Statistics.

The impact of the decline in establishment reporting was also analyzed by examining the change in State codes for jobs that were reported by establishment in 1976, but not in 1981, under the same Employee Identification Number. These jobs represented about 20 percent of all jobs in 1976. Of these jobs, only one-half remained geographically coded in the same State for both years. More than 50 percent of the changes resulted in net inflows to California, Illinois, Michigan, New Jersey, New York, Ohio, Pennsylvania, and Texas. The principal reason for the large shift of jobs to these States is that they contain many of the major cities in which the home offices of companies are located. As stated in the first part of the analysis, this is due primarily to multiunit firms being classified as single unit entities under their home office address.

Coding Files

The geographic data obtained from the SS-4 and ERP coding are maintained in two data files. The Single Unit Code File, which is a historical file, contains one record for each EIN identified from either the SS-4

received and coded by SSA or from Statewide information accreted during annual wage processing when no record of the form's receipt was in the existing file. Excluded from the Single Unit Code File are household employers,³ nonemploying "6-million series" EIN's (for example, trust funds, fiduciaries, and estates that are assigned a 6 in the third position of the EIN), and inactive employers whose records were purged from the files in the early 1970's. The Multiunit Code File contains one record for each establishment of participating ERP employers identified from the form 5019 and coded by SSA. Each year, the job records in the CWHS that contained employer identification numbers are matched to the coding files to associate geographic codes with individual workers.

New Geographic Coding System

The new geographic coding system for the CWHS institutes the beginning of a major shift from an employment location code to an employee residence code. This change in coding concept is necessary because it is no longer possible to identify in SSA records the location of employment for many workers. At the same time, residence data are available for about 70 percent of workers whose W-2 reports are received on magnetic media (tape, cartridge, disk, or direct electronic transmission) from employers.

The process by which W-2 data are posted to SSA's Master Earnings File involves the development of a data file called the Common W-2 Record, which contains W-2 data from both magnetic media and paper reports. The Common W-2 magnetic media records contain the address of the worker as transmitted by the employer. SSA prepares an electronic record for W-2 data from paper reports by optical scanning, but the address of the worker is not one of the items scanned. Thus, the Common W-2 Record obtained from paper reports does not contain the worker's address.

Beginning with 1993 W-2 reports, cases included in the 1-percent CWHS will be extracted from the Common W-2 Record on a flow basis. The resulting data will be saved and matched to the 1-percent sample cases actually posted to the Master Earnings File. (For data year 1993, the posted cases were extracted throughout 1994.) For cases with W-2 address data available, geographic code of residence will be obtained from the ZIP Code and first five characters of the city name by utilizing the algorithm⁴ from the SSA Self-Employed Geographic Coding File. This file is used to assign geographic codes to address information received from the IRS for self-employed workers. Magnetic media W-2 records that cannot be assigned a code (estimated to be about 10 percent of the magnetic media cases) and all paper W-2 records will continue to be coded from the employer-based Single Unit Code File.

Geographic coding for 1991 and 1992 data was performed using the present system, which relies wholly on the Single Unit and Multiunit Code Files. For 1993 data, the new hybrid system will use employer-based geographic code only from the single unit file because information in the multiunit file is being replaced with more accurate geographic data for multi-establishment employers in the single unit file. This replacement will be done by coordinating SSA's Single Unit Code File with Census' Standard Statistical Establishment List (updated with 1987 Economic Census data). Each multiestablishment employer will be assigned a primary geographic code, if a common code representative of all establishments can be determined based on the location of 75 percent of the workers. Otherwise, the employer will be assigned an unclassified code.

Pilot Study

A pilot study of the new coding system was conducted to test the operational aspects of the system and to obtain some preliminary estimates of

the effect of the revised codes on the estimated geographic distribution of workers and FICA taxable wages. Specific research questions included:

- What proportion of the magnetic media W-2 addresses would yield a valid code under the new coding system?
- How many sample cases would be different under the revised geographic code, compared with the old system?
- Which States would have the largest gross differences?
- Which States would have the largest net differences?
- How different would 1991 published data have been had the new coding system been in effect?

Details of the pilot study sample design and selection are given in the appendix.

The analysis of the new geographic coding system is twofold. First, the gross changes in geographic codes for magnetic media cases were examined in order to describe the characteristics of the jobs that are most affected by the new system. The analysis focuses on changes by State, coverage group, and industry group. Second, net changes in the distribution of jobs by State are presented. Estimates were made of the actual State distributions based on both the old code and the new code from the 1991 Employee-Employer (EE-ER) File. To construct the "original" 1991 FICA actual State totals under the old coding system, wage and salary worker counts and FICA taxable wages for each State under the employer-based coding approach were totaled and adjusted by inflation factors to match control totals for the 1991 State and county publication.⁵ To construct 1991 FICA actual State totals under the new W-2 residence coding approach, the total estimated and weighted wage and salary worker count and FICA taxable wage were computed for each State and again inflated to published totals.

Analysis of Coded Magnetic Media Cases

Tables 1–3 show worker count and the percentage of change (that is, no change, any change, State change, county change only) in geographic location under the new coding system by State, coverage group, and industry major group, respectively. The overall percent of change for all workers was 58.3 percent—32.2 percent at the State level and 26.1 percent at the county level only.

States.—Over two-thirds of the States (including the District of Columbia) show a greater than 50 percent change in geographic location under residence coding (table 1). For five of these States—Arkansas, Connecticut, Delaware, Kansas and the District of Columbia—the change in State location alone was greater than 50 percent. These changes can be attributed to either residence State codes replacing State codes of home offices and Delaware corporations or residence State codes for contiguous States replacing employer-based State codes that may or may not have correctly identified the employee's actual work location. Table 1 also includes geographic change for foreign locations, ships at sea, outlying areas, and the unknown. All of the latter showed a high percentage of change at the State equivalent level. For example, 94.0 percent of those workers whose State code was unknown under the employer-based geographic system were assigned a State code under residence coding. The percentage of change at the county level for most States was under 30 percent. Most workers who shift to different counties within the same State under residence coding would be commuters—that is, workers who reside outside their county-of-work.

Appendix table 1 shows the count and percentage distribution of wage and salary workers by new State code for States coded under the old coding system that showed large shifts of workers to other States under the new coding system. These States are

Connecticut, Delaware, Illinois, Maryland, Michigan, New York, Tennessee, Texas, and the District of Columbia. The majority of the worker shifts to different States was generated by the movement from home office address coding of large corporate and Federal Government employers under the old employer-based geographic coding system to residence address coding under the new system. Home office address coding is especially prominent in those States, such as Illinois, Maryland, Michigan, New York, Ohio, and Texas where large cities are located and the District of Columbia. For example, over one-third of the workers previously coded under New York moved to other States—27.7 percent of which were not contiguous States. A second factor contributing to shifts of workers is commuting. For example, one-third of the workers previously coded under the District of Columbia were recoded to neighboring States—Maryland (19.3 percent) and Virginia (13.4 percent). However, the movement of workers commuting to contiguous States can be confounded with that of workers shifting from the home office State to the State of residence where the place of work may be located. A third factor is the incorporation of businesses in Delaware that are located in other States. Under the new geographic code, only 31.4 percent of the workers remained coded under Delaware. The majority of the workers were evenly reclassified under the other States.

Coverage groups.—Federal civilian employees and military personnel showed a high percentage of State change—70.6 percent and 52.6 percent, respectively—under the new geographic coding system (table 2). Shifts of Federal civilian workers to other States under residence coding can be attributed to either movement to contiguous States or movement from home office States of the Federal Government. There is a high percentage of change in State coding for military employees because under the old coding system all military personnel were assigned a special State code

Table 1.—Percent of change in geographic location under new residence coding system, by State¹

Previous location	Worker count (in hundreds)	No change	Any change	State change	County change only
All.....	1,301,131	41.7	58.3	32.2	26.1
Alabama.....	14,522	51.4	48.6	18.7	29.9
Alaska.....	1,825	59.1	40.9	18.8	22.1
Arizona.....	14,402	53.7	46.3	22.7	23.6
Arkansas.....	14,453	27.9	72.1	56.5	15.6
California.....	147,448	49.9	50.1	18.3	31.9
Colorado.....	18,416	31.1	68.9	39.6	29.3
Connecticut.....	22,722	26.6	73.4	59.7	13.8
District of Columbia.....	13,931	25.8	74.2	74.2	...
Delaware.....	6,852	27.7	72.3	68.6	3.7
Florida.....	52,874	51.3	48.7	25.8	22.9
Georgia.....	32,352	34.6	65.4	25.3	40.2
Hawaii.....	5,852	77.3	22.7	12.3	10.4
Idaho.....	3,965	53.6	46.4	19.7	26.7
Illinois.....	66,391	35.7	64.3	38.8	25.5
Indiana.....	23,022	48.0	52.0	18.6	33.4
Iowa.....	11,048	39.1	60.9	39.2	21.7
Kansas.....	13,201	34.3	65.7	54.9	10.9
Kentucky.....	13,165	39.5	60.5	28.7	31.9
Louisiana.....	14,935	41.0	59.0	34.8	24.2
Maine.....	4,320	41.1	58.9	46.3	12.6
Maryland.....	23,456	36.7	63.3	34.6	28.8
Massachusetts.....	36,790	41.1	58.9	27.0	32.0
Michigan.....	54,853	37.1	62.9	37.2	25.7
Minnesota.....	27,078	43.2	56.8	32.9	23.8
Mississippi.....	7,152	48.3	51.7	26.1	25.6
Missouri.....	24,987	38.3	61.7	40.9	20.8
Montana.....	2,314	43.9	56.1	46.5	9.6
Nebraska.....	7,066	53.5	46.5	34.5	12.1
Nevada.....	6,917	59.8	40.2	23.3	16.9
New Jersey.....	41,367	43.0	57.0	27.6	29.4
New Hampshire.....	3,869	34.5	65.5	37.0	28.5
New Mexico.....	4,787	51.5	48.5	19.1	29.5
New York.....	121,295	29.5	70.5	37.6	32.9
North Carolina.....	29,403	53.2	46.8	19.5	27.3
North Dakota.....	1,573	73.2	26.8	10.7	16.1
Ohio.....	64,166	41.0	59.0	37.3	21.8
Oklahoma.....	11,545	37.5	62.5	35.2	27.3
Oregon.....	11,725	53.6	46.4	16.5	29.9
Pennsylvania.....	53,289	47.7	52.3	22.1	30.2
Rhode Island.....	5,032	45.5	54.5	33.3	21.2
South Carolina.....	13,676	43.9	56.1	29.2	26.9
South Dakota.....	1,810	37.6	62.4	26.1	36.3
Tennessee.....	27,820	34.9	65.1	45.3	19.9
Texas.....	81,602	44.3	55.7	28.4	27.3
Utah.....	7,158	62.5	37.5	17.1	20.3
Vermont.....	2,118	41.1	58.9	34.1	24.8
Virginia.....	26,000	32.3	67.7	23.9	43.8
Washington.....	20,703	59.6	40.4	17.9	22.5
West Virginia.....	4,933	41.5	58.5	44.3	14.2
Wisconsin.....	25,518	34.2	65.8	28.3	37.5
Wyoming.....	1,380	58.9	41.1	25.2	15.9
Foreign.....	2,017	54.5	45.5	45.5	...
Guam.....	206	...	100.0	100.0	...
Puerto Rico.....	2,009	21.5	78.5	71.0	7.5
Ships at sea.....	74	9.1	90.9	90.9	...
Virgin Islands.....	180	.6	99.4	99.4	...
Unknown.....	9,473	6.0	94.0	94.0	...

¹For workers reported on magnetic media based on pilot study sample estimates.

(03) that is not representative of any geographic location. (The same applies to the military reserves, which was assigned State code 04.)

Table 2 also shows that household workers had the highest percentage of change at the county level. This is because residence codes provide household workers with a county location that is not available from the employer record. During the coding of the EE-ER File under the employer-based coding system, household workers (which are identified by their Type of Employment Code from the detail portion of the annual wage report) are in most cases assigned a Statewide geographic code based on the first two digits (Internal Revenue District) of the EIN, which may or may not represent the State where the employer is located. This method of assigning a Statewide geographic code to household workers is done when there is no State and county coded employer record in the corresponding Single Unit Code File. Currently, most household employers are not in the single unit file because their form SS-4 does not go through the geographic coding process at SSA and their annual wage reporting record is not included in the single unit file update process. (Beginning with data year 1992, household employers will be identified from their annual wage reports and accreted to the Single Unit Code File. However, a Statewide geographic code will still be assigned based on the first two digits of the EIN.

Industry major groups.—Table 3 shows a greater than 50 percent change in geographic location for several of the industry major groups (major group code is in parenthesis):

- Military (03)
- Metal/oil and gas mining (10 and 13)
- Manufacturing (20, 28, 30, 31, 32, 37, and 38)
- U.S. Postal Service (43)
- Air transportation/transportation services (45 and 47)

- General merchandise stores (53)
- Holding/other investment offices (67)

Except for military personnel, the shift of most workers under these major groups to other States can be attributed to the movement from the employer-based geographic codes that identify home office locations of large corporations (including the U.S. Postal Service) to residence-based State codes, which may or may not be a contiguous State.

Table 4 provides examples of industry major groups with large shifts of workers to other States under the new geographic coding system. These major groups are food manufacturing (20), auto manufacturing (37), retail general merchandise stores (53), and eating and drinking establishments (58). The table lists for each major group the seven States under the old geographic code that showed the largest concentration of workers under that particular major group and the percentage of geographic change at the State level (that is, same State, moved to another State) for these workers under the new geographic code. The percentage of workers that moved to another State was further divided and identified by whether or not it was a contiguous State, along with a list of

the surrounding States for each previous State location.

Table 4 shows a pattern of worker movement under resident coding that differs greatly among the industry major groups. In food manufacturing industries, States in which large corporate home offices are located are identified by considerable shifts of workers to noncontiguous States under the new residence coding system. For example, Illinois, Ohio, and Texas showed more than 60 percent of workers now located elsewhere. In contrast, for California, Idaho, and Pennsylvania the two coding systems show large majorities in the same State and, where shifts occur, the new State locations are primarily contiguous States.

Compared with the food industry, geographic shifts across States were less frequent in auto manufacturing. Under each State, except for Indiana (51.4 percent), over 60 percent of the workers did not move to another State. Of the workers that moved to another State, an average of 57 percent moved to contiguous States and an average of 43 percent moved to noncontiguous States.

For retail general merchandise stores, only two States showed a high percentage of workers remaining within the same State—North Carolina (93.4 percent) and Virginia (51.1). The re-

Table 2.—Percent of change in geographic location under new residence coding system, by coverage group¹

Coverage group	Worker count (in hundreds)	No change	Any change	State change	County change only
All.....	1,301,131	41.7	58.3	32.2	26.1
Agriculture.....	10,451	50.0	50.0	20.8	29.2
Household.....	2,133	4.4	95.7	4.4	91.3
Federal civilian.....	36,020	29.4	70.6	55.6	15.0
State/local.....	201,525	65.6	34.4	2.7	31.7
Military.....	28,037	47.4	52.6	52.6	...
Reserves.....	11,840	68.4	31.6	31.6	...
National Guard.....	(2)	(2)	(2)	(2)	(2)
Religious.....	(2)	(2)	(2)	(2)	(2)
Nonprofit.....	77,790	64.5	35.6	9.0	26.6
Reported tips.....	21,244	41.4	58.6	34.9	23.8
Other.....	911,610	34.5	65.5	39.3	26.3

¹ For workers reported on magnetic media based on pilot study sample estimates.

² Sample size too small to be shown.

Table 3.—Percent change in geographic location under new residence coding system, by industry major group¹

Industry major group	Worker count (in hundreds)	No change	Any change	State change	County change only
All.....	1,301,131	41.7	58.3	32.2	26.1
00 Unknown.....	142,505	16.7	83.3	43.5	39.7
01 Agricultural production.....	20,824	42.3	57.7	14.5	43.2
03 Armed Forces.....	28,037	47.4	52.6	52.6	...
04 Reserves.....	12,057	67.2	32.8	32.8	...
07 Agricultural services.....	2,088	34.5	65.5	47.5	18.1
08 Forestry.....	45	36.8	63.2	...	63.2
09 Fishing/hunting.....	158	54.5	45.5	34.5	11.1
10 Metal mining.....	1,063	26.2	73.8	54.1	19.7
12 Coal mining.....	443	39.4	60.6	31.6	29.0
13 Oil/gas mining.....	2,264	19.3	80.7	61.9	18.8
14 Nonmetal mining, except fuels.....	612	31.5	68.5	41.4	27.1
15 General building contractors.....	5,950	32.7	67.3	41.2	26.2
16 Heavy construction.....	6,443	27.8	72.2	37.5	34.7
17 Special trade contractors.....	11,136	43.3	56.7	26.5	30.1
20 Food manufacturing.....	28,416	27.4	72.6	54.5	18.0
21 Tobacco manufacturing.....	1,905	34.6	65.4	38.7	26.7
22 Textile mill products manufacturing.....	8,327	44.9	55.1	24.0	31.0
23 Apparel/textile products manufacturing.....	5,978	51.5	48.5	26.1	22.4
24 Lumber/wood products manufacturing.....	4,080	36.0	64.0	41.5	22.5
25 Furniture/fixtures manufacturing.....	4,255	39.3	60.7	32.3	28.4
26 Paper/allied products manufacturing.....	4,640	43.8	56.2	41.8	14.4
27 Printing/publishing.....	12,539	48.2	51.8	22.4	29.4
28 Chemical/allied products manufacturing.....	8,948	28.6	71.4	54.0	17.5
29 Petroleum/coal products manufacturing.....	3,039	18.7	81.3	49.3	32.1
30 Rubber/plastics products manufacturing.....	7,765	33.2	66.8	53.0	13.8
31 Leather/leather products manufacturing.....	5,651	20.6	79.4	70.0	9.4
32 Stone/clay/glass products manufacturing.....	3,690	28.6	71.4	54.6	16.8
33 Primary metal manufacturing.....	6,254	51.8	48.2	32.0	16.2
34 Fabricated metal products manufacturing.....	8,751	52.2	47.8	25.0	22.8
35 Industrial machinery/equip. manufacturing.....	17,272	42.1	57.9	36.4	21.5
36 Electronic/electrical equip. manufacturing.....	12,601	38.5	61.5	42.3	19.2
37 Transportation equip. manufacturing.....	13,772	22.8	77.2	59.7	17.5
38 Instruments/related products manufacturing.....	9,452	28.5	71.5	54.3	17.2
39 Miscellaneous manufacturing.....	5,064	50.5	49.5	33.6	15.9
40 Railroad transportation.....	257	22.9	77.1	19.2	57.8
41 Local/interurban transit.....	4,258	61.5	38.5	15.7	22.7
42 Trucking/warehousing.....	15,011	25.7	74.3	20.8	53.6
43 U.S. Postal Service.....	9,419	6.1	93.9	93.9	...
44 Water transportation.....	1,034	37.2	62.8	36.5	26.3
45 Air transportation.....	10,340	20.5	79.5	53.6	25.9
46 Pipelines, except natural gas.....	374	33.7	66.3	39.5	26.8
47 Transportation services.....	4,267	22.0	78.0	63.8	14.3
48 Communications.....	14,605	23.9	76.1	39.7	36.4
49 Electric/gas/sanitary services.....	9,223	54.5	45.5	7.8	37.7
50 Wholesale durable goods.....	26,811	28.5	71.5	46.7	24.8

¹ See footnote at end of table.

Table 3.—Percent of change in geographic location under new residence coding system, by industry major group¹— *Cont.*

Industry major group	Worker count (in hundreds)	No change	Any change	State change	County change only
51 Wholesale nondurable goods.....	19,126	41.3	58.7	35.1	23.6
52 Retail building/garden supplies.....	4,357	36.9	63.1	39.6	23.5
53 General merchandise stores.....	29,578	27.1	72.9	60.4	12.4
54 Food stores.....	30,846	39.6	60.4	24.8	35.6
55 Auto dealers/gas stations.....	5,381	53.6	46.4	19.1	27.3
56 Apparel stores.....	12,963	27.5	72.5	47.6	24.9
57 Homefurnishings stores.....	4,028	42.0	58.0	26.4	31.6
58 Bars and restaurants.....	92,810	39.2	60.8	42.2	18.6
59 Miscellaneous retail.....	18,834	35.8	64.2	36.4	27.8
60 Depository institutions.....	20,924	49.2	50.8	7.0	43.8
61 Nondepository institutions.....	3,010	44.0	56.0	30.7	25.3
62 Security/commodity brokers.....	3,156	23.9	76.1	40.9	35.3
63 Insurance carriers.....	20,657	35.8	64.2	44.0	20.2
64 Insurance agents/services.....	4,397	29.5	70.5	45.3	25.2
65 Real estate.....	13,525	39.3	60.7	36.0	24.6
67 Holding/investment offices.....	19,350	24.1	75.9	63.1	12.9
70 Hotels/lodging places.....	17,689	50.0	50.0	32.2	17.7
72 Personal services.....	10,963	50.5	49.5	18.3	31.3
73 Business services.....	73,016	36.5	63.5	40.3	23.3
75 Automotive services/parking.....	3,499	50.2	49.8	26.6	23.3
76 Miscellaneous repair services.....	1,221	48.4	51.6	13.1	38.5
78 Motion pictures.....	7,273	54.8	45.2	33.8	11.4
79 Amusement/recreation services.....	8,264	69.9	30.1	13.2	16.8
80 Health services.....	66,571	65.2	34.8	8.0	26.9
81 Legal services.....	5,148	53.8	46.2	17.7	28.5
82 Educational services.....	47,989	64.9	35.1	13.0	22.1
83 Social services.....	14,283	60.6	39.4	8.4	31.0
84 Museums/botanical zoological gardens.....	267	79.3	20.7	4.6	16.2
86 Membership organizations.....	17,683	53.0	47.0	28.0	19.0
87 Engineering/mgmt. services.....	36,287	40.7	59.3	37.8	21.5
88 Household.....	2,152	4.4	95.6	5.2	90.4
89 Services, not elsewhere classified.....	693	48.1	51.9	18.7	33.2
90 Public administration.....	183,462	61.6	38.4	7.6	30.9
97 Unknown (previously out of business).....	106	39.5	60.5	47.3	13.2
99 Nonclassifiable industry.....	2	100.0

¹ For workers reported on magnetic media based on pilot study sample estimates.

mainder of the States only show an average of 19 percent of the workers staying under the same State. Of those that moved to another State, an average of 54 percent moved to contiguous States and an average of 46 percent moved to other States. Three States under eating and drinking establishments show greater than 73 percent of

the workers staying under the same State—California, Georgia, and Texas. The remainder of the States kept an average of 27 percent of the workers under resident coding. The majority that moved went to a noncontiguous State—an average of 78 percent for all States. This is probably due to the home office effect.

Analysis of Actual State Distributions

Table 5 shows the actual State distribution of worker counts and FICA wages under both the old and the new geographic coding systems. Excluded from the actual State table are the unknown category and the military reserves, which were not included in published State tables.⁶ This table

shows the net change in State totals that would have occurred had the new system been in place for 1991. Those States with net gains and losses are further illustrated in chart 1. The chart shows the percentage differences between the old and new worker counts for each State. Some of the

heaviest net losses of workers are shown for Michigan (-789,700), New York (-2,192,800), and Ohio (-765,200). Since these States contain large cities, the losses can most likely be attributed to workers being recoded from their employer's home office address to their residence address,

which may or may not be the same State as their work site. Under the new coding system, State worker counts in the 1991 EE-ER File are closer to the 1991 worker estimates published by the Bureau of Labor Statistics,⁷ even though the geographic concepts and universe of coverage for

Table 4.—Examples of industry major groups with large shifts of workers to other States under new geographic code

Worker location under new geographic code	State location under old geographic code						
	Food manufacturing (major group 20)						
	California	Idaho	Illinois	New York	Ohio	Pennsylvania	Texas
Same State.....	71.8	78.8	39.5	34.0	22.1	80.1	37.0
Moved to another State....	28.2	21.2	60.5	66.0	77.9	19.9	63.0
Contiguous State.....	.7	13.6	14.5	23.6	25.0	13.1	.5
Noncontiguous State....	27.5	7.7	46.0	42.4	52.9	6.8	62.5
Moved to contiguous State →	Arizona Nevada Oregon	Oregon Washington Wyoming	Iowa Kentucky Michigan Minnesota Missouri	Connecticut Massachusetts New Jersey Pennsylvania Rhode Island	Iowa Kentucky Michigan Minnesota Missouri	Maryland New Jersey New York Ohio Virginia	Louisiana Oklahoma New Mexico
	Automobile manufacturing (major group 37)						
	Georgia	Illinois	Indiana	Iowa	Michigan	Ohio	Virginia
Same State.....	66.1	63.0	51.4	93.0	84.4	73.3	90.6
Moved to another State....	34.0	37.0	48.6	7.0	15.6	26.7	9.4
Contiguous State.....	5.2	17.8	33.1	7.0	6.1	15.4	6.6
Noncontiguous State....	28.8	19.8	15.5	.0	9.5	11.3	2.9
Moved to contiguous State →	Alabama Florida	Iowa Michigan Minnesota Wisconsin	Illinois Michigan Ohio Wisconsin	Minnesota	Illinois Indiana	Indiana Kentucky Michigan	West Virginia
	Retail general merchandise (major group 53)						
	Arizona	Connecticut	Missouri	New York	North Carolina	Oklahoma	Virginia
Same State.....	7.0	22.0	26.8	24.6	93.4	14.5	51.1
Moved to another State....	93.0	78.1	73.2	75.4	6.6	85.5	48.9
Contiguous State.....	59.5	38.0	3.5	4.2	6.6	53.8	17.4
Noncontiguous State....	33.5	40.0	69.7	71.2	.0	31.7	31.5
Moved to contiguous State →	Mississippi Missouri Oklahoma Tennessee Texas	Massachusetts New Jersey New York Rhode Island	Illinois	Connecticut Massachusetts New Jersey	South Carolina	Arkansas Kansas Louisiana Texas	Kentucky Maryland North Carolina Tennessee

Table 4.—Examples of industry major groups with large shifts of workers to other States under new geographic code—*Cont.*

Worker location under new geographic code	State location under old geographic code						
	Eating and drinking establishments (major group 58)						
	California	Florida	Georgia	Kansas	Tennessee	Texas	Wisconsin
Same State.....	73.7	37.6	76.9	19.6	24.6	82.6	27.9
Moved to another State....	26.3	62.4	23.1	80.5	75.4	17.5	72.1
Contiguous State.....	3.8	1.7	18.6	13.9	12.1	2.6	10.6
Noncontiguous State..	22.5	60.7	4.5	66.5	63.3	14.9	61.5
Moved to contiguous State →	Arizona Nevada Oregon	Alabama Georgia	Florida Tennessee	Arizona Colorado Missouri Nebraska Oklahoma	Alabama Arizona Georgia Kentucky Mississippi Missouri North Carolina Virginia	Arkansas Oklahoma	Iowa Minnesota

Chart 1.—Percent of increase or decrease in worker counts, based on 1-percent sample



Table 5.—Actual State distribution of workers and FICA taxable wages under old and new geographic coding systems, 1991

State and State code	Old worker count (in hundreds)	Old FICA (in millions)	New worker count (in hundreds)	New FICA (in millions)
All.....	1,472,073	\$22,820.3	1,472,067	\$22,821.5
64 Alabama.....	19,556	275.6	23,660	363.0
02 Alaska.....	2,644	40.7	2,834	38.7
86 Arizona.....	17,544	251.9	22,720	325.2
71 Arkansas.....	17,905	198.7	14,003	152.0
93 California.....	148,358	2,566.2	154,956	2,695.2
84 Colorado.....	20,475	307.0	20,305	297.0
16 Connecticut.....	25,237	444.2	18,708	330.7
90 District of Columbia.....	10,767	213.3	5,723	101.9
51 Delaware.....	7,944	129.6	4,813	69.6
50 Florida.....	63,426	863.9	71,095	957.8
58 Georgia.....	37,950	516.8	42,032	618.1
26 Hawaii.....	5,981	103.3	7,419	118.2
82 Idaho.....	6,154	71.9	6,650	88.0
33 Illinois.....	74,892	1,189.4	69,759	1,189.2
32 Indiana.....	28,853	434.1	35,400	577.2
42 Iowa.....	15,151	201.4	15,415	204.8
48 Kansas.....	16,556	202.9	14,903	200.9
61 Kentucky.....	17,601	216.6	19,206	249.8
72 Louisiana.....	20,403	309.0	19,780	258.4
11 Maine.....	5,442	66.6	5,337	71.1
52 Maryland.....	28,065	450.2	29,647	511.3
14 Massachusetts.....	38,377	630.0	36,915	677.6
35 Michigan.....	59,774	997.9	51,877	760.2
41 Minnesota.....	30,242	472.7	28,898	498.5
65 Mississippi.....	10,670	124.4	13,093	164.0
43 Missouri.....	30,719	414.9	28,341	361.0
81 Montana.....	4,308	50.3	3,999	44.2
47 Nebraska.....	9,961	129.8	10,028	141.6
88 Nevada.....	7,157	93.8	7,597	97.7
12 New Hampshire.....	5,661	87.0	6,873	120.3
22 New Jersey.....	47,596	887.4	51,445	994.8
85 New Mexico.....	7,277	87.8	9,129	103.9
21 New York.....	125,629	2,463.6	103,701	1,930.8
56 North Carolina.....	36,845	484.1	42,209	609.6
45 North Dakota.....	3,105	37.0	3,923	47.4
31 Ohio.....	65,955	1,006.1	58,303	888.8
73 Oklahoma.....	16,009	220.3	17,638	213.4
92 Oregon.....	15,911	235.2	17,708	257.8
23 Pennsylvania.....	64,893	1,085.6	68,356	1,044.6
15 Rhode Island.....	5,810	92.2	5,800	90.2
57 South Carolina.....	17,700	233.3	21,325	276.3
46 South Dakota.....	3,496	39.1	3,790	42.3
63 Tennessee.....	32,375	404.1	29,824	373.9
74 Texas.....	88,376	1,285.5	94,261	1,326.4
87 Utah.....	8,857	108.7	9,986	134.9
13 Vermont.....	3,138	43.0	3,444	53.7
53 Virginia.....	32,574	498.5	39,254	639.2
91 Washington.....	26,178	413.1	33,738	619.7
55 West Virginia.....	7,594	100.6	7,621	96.7
36 Wisconsin.....	31,214	462.2	31,256	494.9
83 Wyoming.....	2,475	29.0	3,084	36.4
24 American Samoa.....	63	.4	22	.2
03 Armed Forces.....	28,393	408.7	14,343	176.5
98 Foreign.....	2,201	30.9	1,341	17.5
29 Guam.....	478	5.5	347	3.4
30 Northern Mariana Islands.....	48	.3	53	.3
27 Puerto Rico.....	9,563	94.4	7,587	60.9
99 Ships at Sea.....	99	2.2	15	.1
28 Virgin Islands.....	448	7.1	294	3.7

the two employment series differ. (The Bureau of Labor Statistics assigns geographic codes to workers covered under State Unemployment Insurance and Unemployment Compensation for Federal Employees based on where their jobs are physically located.) For example, the Bureau of Labor Statistics reported that for 1991 total annual average employment in New York was 3 million jobs less than the total jobs that ORS reported in the EE-ER File for the same year, compared with 5 million less jobs in 1981.⁸

Other locations that showed heavy outflows of workers were the District of Columbia (-504,400) and Delaware (-313,100). The loss of workers in the District of Columbia can be attributed primarily to commuters. In Delaware, the decrease in number of workers was due to employers being classified under the State they incorporated in, not under their actual work location. In most cases, it was difficult to determine the primary reason for the increase or decrease in workers for any one State. A main exception would be the huge decrease of workers in New York State, which was definitely caused by workers shifting from their employers home office, since the decrease is far greater than what the surrounding States are absorbing.

Conclusion

This article describes the new geographic coding system that will be used to identify workers in the Continuous Work History Sample by geographic location. The approach to assigning geographic codes to workers based on the locations of their employers is no longer viable because too few multiestablishment employers are reporting their workers by establishment. The result is that workers are shown as being employed at the employer's home office location and not their actual work location. SSA, therefore, is undertaking a changeover in geographic coding from location of employer to location of employee residence.

Initially, the new system will be a hybrid system: Workers of large firms—those with 250 or more employees—and others submitting their W-2 reports on magnetic media being assigned geographic codes by W-2 address (generally the individual's residence), and all other workers being assigned a geographic code based on employer address. This hybrid system should provide more accurate geographic distributions of workers within a residence concept than the old system could provide within an employer location concept. Over time, as more W-2's are reported to SSA by magnetic media, there will be corresponding improvement in the accuracy of the residence distributions. In addition, basing published geographic distributions of workers on residence brings these data into line with existing publications of the geographic distributions of beneficiaries that are residence based.^{9,10}

The new geographic coding system is expected to be in place for the 1993 data year. The Office of Research and Statistics plans to produce State and county data under both coding systems: Employer based and residence based. The two tabulations will allow us to assess whether changes in worker and wage estimates between 1992 and 1993 are due to the change in the coding system or part of an existing time trend.

Notes

¹ The geographic classification coding scheme for residence coding will remain the same.

² *Employment and Wages Annual Averages, 1981*, U.S. Department of Labor, Bureau of Labor Statistics, June 1983.

³ During updates to the Single Unit Code File with 1991/1992 annual wage reporting information, Type of Employment codes were used to accrete household employers to the Single Unit Code File. This was after the single unit file was used to code the 1991 annual Employer-Employee File.

⁴ The algorithm was developed over time by SSA's Office of Research and Statistics and the Office of Information Management. The original source of input is unknown.

⁵ *Earnings and Employment Data for Wage and Salary Workers Covered Under Social Security, by State and County, 1991*, U.S. Department of Health and Human Services, Social Security Administration, Office of Research and Statistics, 1994.

⁶ *Ibid.*

⁷ *Employment and Wages Annual Averages, 1991*, U.S. Department of Labor, Bureau of Labor Statistics, January 1993.

⁸ *Employment and Wages Annual Averages, 1981*, U.S. Department of Labor, Bureau of Labor Statistics, June 1983.

⁹ *OASDI Beneficiaries by State and County, December 1993*, U.S. Department of Health and Human Services, Social Security Administration, Office of Research and Statistics, 1994.

¹⁰ *SSI Recipients by State and County, December 1993*, U.S. Department of Health and Human Services, Social Security Administration, Office of Research and Statistics, 1994.

References

Dill, Linda M. 1992. "The Development and Use of Industry Data by the Social Security Administration." *Social Security Bulletin*, Vol. 55, No. 4 (Winter), pp. 43-53.

Dill Linda M., Adah D. Enis, and Cheryl I. Williams. 1991. "The Decline in Establishment Reporting: Impact on CWSHS Industrial and Geographic Data." *Social Security Bulletin*, Vol. 54, No. 1 (January), pp. 2-20.

Smith, Creston M. 1989. "The Social Security Administration's Continuous Work History Sample," *Social Security Bulletin*, Vol. 52, No. 10 (October), pp. 20-28.

U.S. Department of Health and Human Services, Social Security Administration. 1994. *Earnings and Employment Data for Wage and Salary Workers Covered Under Social Security, by State and County—1991*. Washington, DC: Office of Research and Statistics.

U.S. Department of Health and Human Services, Social Security Administration. 1994. *SSI Recipients by State and County, December 1993*. Washington, DC: Office of Research and Statistics.

U.S. Department of Health and Human Services, Social Security Administration.

1994. State Statistics. Washington, DC: Office of Research and Statistics.

U.S. Department of Health and Human Services, Social Security Administration. 1993. *OASDI Beneficiaries by State and County, December 1992*. Washington, DC: Office of Research and Statistics.

U.S. Department of Health and Human Services, Social Security Administration. 1989. *Annual Statistical Supplement to the Social Security Bulletin, 1989*. Washington, DC: U.S. Government Printing Office. table 4.B10.

U.S. Department of Health and Human Services, Social Security Administration. 1978. *Policy Analysis with Social Security Research Files—Proceedings of a Workshop held March 1978 at Williamsburg, Virginia* (Research Report No. 52), Washington, DC: U.S. Government Printing Office.

U.S. Department of Labor. 1993. *Employment and Wages Annual Averages, 1991*. Washington, DC: Bureau of Labor Statistics. January.

U.S. Department of Labor. 1983. *Employment and Wages Annual Averages, 1981*. Washington, DC: Bureau of Labor Statistics. June.

Appendix: Sample Design and Selection

A pilot study was designed to test the new geographic coding operation as if it had been implemented for 1991, the latest year for which published data are available (see *Earnings and Employment Data for Wage and Salary Workers Covered Under Social Security, by State and County*). The pilot study cases came from the 1-percent sample Employee-Employer (EE-ER) File from which the published data were derived. The pilot study sample consisted of about a 0.25 percent subsample of all magnetic media W-2's (that is, about one-fourth of the magnetic media cases in the 1-percent sample) and the full 1-percent sample of the paper W-2 reports.

Before the 1993 reporting year, W-2 address data were not maintained

Table I.—Count and percentage distribution of wage and salary workers, by new State for States coded under old geographic

New State	Old State									
	Connecticut		Delaware		District of Columbia		Illinois		Maryland	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Total.....	22,722	100.0	6,852	100.0	13,931	100.0	66,391	100.0	23,456	100.0
Alabama.....	1,057	4.7	0		3	.0	205	.3	2	.0
Alaska.....	0	.0	42	.6	35	.3	13	.0	127	.5
Arizona.....	32	.1	18	.3	153	1.1	521	.8	72	.3
Arkansas.....	789	3.5	8	.1	21	.2	228	.3	28	.1
California.....	1,096	4.8	240	3.5	798	5.7	3,279	4.9	716	3.1
Colorado.....	71	.3	51	.7	101	.7	555	.8	54	.2
Connecticut.....	9,167	40.3	31	.5	72	.5	255	.4	55	.2
District of Columbia.....	19	.1	2	.0	3,600	25.8	12	.0	354	1.5
Delaware.....	50	.2	2,153	31.4	50	.4	28	.0	180	.8
Florida.....	226	1.0	140	2.0	386	2.8	1,895	2.9	443	1.9
Georgia.....	163	.7	243	3.5	218	1.6	942	1.4	354	1.5
Hawaii.....	64	.3	2	.0	23	.2	122	.2	49	.2
Idaho.....	12	.1	45	.7	10	.1	56	.1	6	.0
Illinois.....	274	1.2	621	9.1	185	1.3	40,623	61.2	446	1.9
Indiana.....	141	.6	136	2.0	76	.5	1,152	1.7	181	.8
Iowa.....	7	.0	39	.6	53	.4	278	.4	20	.1
Kansas.....	66	.3	11	.2	62	.4	447	.7	55	.2
Kentucky.....	100	.4	28	.4	58	.4	487	.7	103	.4
Louisiana.....	139	.6	120	1.8	88	.6	999	1.5	93	.4
Maine.....	180	.8	1	.0	14	.1	7	.0	19	.1
Maryland.....	355	1.6	155	2.3	2,684	19.3	384	.6	15,350	65.4
Massachusetts.....	784	3.5	80	1.2	172	1.2	230	.3	191	.8
Michigan.....	243	1.1	144	2.1	82	.6	596	.9	122	.5
Minnesota.....	86	.4	41	.6	89	.6	774	1.2	39	.2
Mississippi.....	90	.4	27	.4	25	.2	323	.5	14	.1
Missouri.....	90	.4	7	.1	157	1.1	1,370	2.1	131	.6
Montana.....	1	.0	0	.0	7	.0	95	.1	0	.0
Nebraska.....	5	.0	3	.0	88	.6	131	.2	12	.1
Nevada.....	11	.0	0	.0	32	.2	122	.2	27	.1
New Hampshire.....	164	.7	8	.1	9	.1	12	.0	18	.1
New Jersey.....	406	1.8	244	3.6	387	2.8	689	1.0	260	1.1
New Mexico.....	740	3.3	7	.1	29	.2	136	.2	26	.1
New York.....	1,198	5.3	110	1.6	371	2.7	1,005	1.5	385	1.6
North Carolina.....	90	.4	102	1.5	59	.4	934	1.4	159	.7
North Dakota.....	53	.2	4	.1	15	.1	6	.0	0	.0
Ohio.....	398	1.8	179	2.6	211	1.5	1,763	2.7	317	1.4
Oklahoma.....	807	3.6	45	.7	59	.4	220	.3	51	.2
Oregon.....	752	3.3	30	.4	88	.6	147	.2	2	.0
Pennsylvania.....	638	2.8	330	4.8	316	2.3	940	1.4	682	2.9
Rhode Island.....	137	.6	7	.1	7	.0	41	.1	19	.1
South Carolina.....	55	.2	405	5.9	54	.4	206	.3	61	.3
South Dakota.....	0	.0	0	.0	32	.2	53	.1	0	.0
Tennessee.....	200	.9	20	.3	101	.7	512	.8	136	.6
Texas.....	231	1.0	203	3.0	694	5.0	2,050	3.1	699	3.0
Utah.....	29	.1	0	.0	47	.3	91	.1	34	.1
Vermont.....	68	.3	4	.1	21	.2	25	.0	5	.0
Virginia.....	263	1.2	61	.9	1,863	13.4	313	.5	1,157	4.9
Washington.....	69	3.4	339	5.0	86	.6	255	.4	41	.2
West Virginia.....	68	.3	3	.0	46	.3	16	.0	144	.6
Wisconsin.....	246	1.1	348	5.1	59	.4	805	1.2	18	.1
Wyoming.....	78	.3	13	.2	15	.1	37	.1	0	.0
Foreign.....	3	.0	0	.0	0	.0	0	.0	2	.0
Guam.....	0	.0	0	.0	0	.0	0	.0	0	.0
Puerto Rico.....	10	.0	0	.0	18	.1	6	.0	0	.0

coding system that showed large shifts of workers to other States under new geographic system

hundreds]

		Old State								New State
Michigan		New York		Ohio		Tennessee		Texas		
Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	
54,853	100.0	121,295	100.0	64,166	100.0	27,820	100.0	81,602	100.0	Total
200	.4	363	.3	380	.6	0	.0	23	.0	Alabama
30	.1	2	.0	24	.0	552	2.0	532	.7	Alaska
209	.4	2,318	1.9	244	.4	161	.6	700	.9	Arizona
152	.3	156	.1	303	.5	170	.6	393	.5	Arkansas
1,357	2.5	5,171	4.3	1,320	2.1	1,161	4.2	2,325	2.8	California
333	.6	518	.4	143	.2	135	.5	670	.8	Colorado
342	.6	1,564	1.3	180	.3	41	.1	203	.2	Connecticut
48	.1	193	.2	29	.0	334	1.2	87	.1	District of Columbia
79	.1	180	.1	19	.0	45	.2	32	.0	Delaware
1,216	2.2	2,357	1.9	842	1.3	1,284	4.6	1,704	2.1	Florida
607	1.1	1,583	1.3	1,267	2.0	581	2.1	820	1.0	Georgia
28	.1	133	.1	26	.0	0	.0	22	.0	Hawaii
92	.2	97	.1	12	.0	17	.1	75	.1	Idaho
1,635	3.0	1,793	1.5	1,369	2.1	489	1.8	996	1.2	Illinois
911	1.7	1,219	1.0	2,223	3.5	853	3.1	242	.3	Indiana
473	.9	157	.1	161	.3	92	.3	203	.2	Iowa
208	.4	529	.4	89	.1	133	.5	253	.3	Kansas
267	.5	809	.7	1,671	2.6	268	1.0	304	.4	Kentucky
280	.5	569	.5	236	.4	52	.2	1,445	1.8	Louisiana
126	.2	182	.1	56	.1	9	.0	100	.1	Maine
355	.6	1,039	.9	383	.6	126	.5	548	.7	Maryland
528	1.0	1,459	1.2	624	1.0	28	.1	439	.5	Massachusetts
34,424	62.8	1,078	.9	1,294	2.0	358	1.3	507	.6	Michigan
264	.5	540	.4	224	.3	147	.5	286	.4	Minnesota
180	.3	223	.2	262	.4	489	1.8	503	.6	Mississippi
509	.9	541	.4	387	.6	314	1.1	635	.8	Missouri
16	.0	17	.0	25	.0	0	0	36	.0	Montana
192	.4	158	.1	139	.2	28	.1	124	.2	Nebraska
73	.1	169	.1	45	.1	108	.4	105	.1	Nevada
46	.1	296	.2	101	.2	0	.0	66	.1	New Hampshire
318	.6	6,200	5.1	806	1.3	277	1.0	530	.6	New Jersey
65	.1	141	.1	47	.1	24	.1	519	.6	New Mexico
1,231	2.2	75,672	62.4	1,494	2.3	210	.8	1,144	1.4	New York
655	1.2	1,761	1.5	500	.8	532	1.9	452	.6	North Carolina
68	.1	19	.0	34	.1	25	.1	21	.0	North Dakota
1,630	3.0	2,206	1.8	40,261	62.7	692	2.5	709	.9	Ohio
149	.3	453	.4	207	.3	89	.3	576	.7	Oklahoma
93	.2	278	.2	162	.3	76	.3	141	.2	Oregon
992	1.8	2,593	2.1	1,757	2.7	301	1.1	777	1.0	Pennsylvania
55	.1	122	.1	50	.1	15	.1	108	.1	Rhode Island
217	.4	625	.5	258	.4	148	.5	217	.3	South Carolina
31	.1	38	.0	10	.0	4	.0	72	.1	South Dakota
547	1.0	496	.4	983	1.5	15,223	54.7	813	1.0	Tennessee
1,815	3.3	2,254	1.9	1,665	2.6	1,053	3.8	58,396	71.6	Texas
70	.1	274	.2	134	.2	25	.1	241	.3	Utah
34	.1	255	.2	65	.1	0	.0	12	.0	Vermont
528	1.0	1,281	1.1	611	1.0	657	2.4	1,399	1.7	Virginia
217	.4	639	.5	181	.3	133	.5	546	.7	Washington
200	.4	239	.2	366	.6	120	.4	162	.2	West Virginia
708	1.3	267	.2	466	.7	187	.7	262	.3	Wisconsin
22	.0	12	.0	6	.0	0	.0	87	.1	Wyoming
0	.0	4	.0	0	.0	0	.0	0	.0	Foreign
0	.0	0	.0	3	.0	0	.0	0	.0	Guam
28	.1	47	.0	19	.0	0	.0	43	.1	Puerto Rico

as part of the 1-percent EE-ER statistical system. In the past, address data were maintained only as part of the 100 percent file used to post wage items to the Social Security Administration's Master Earnings File. For 1991, the file that contained the address data consisted of about 1,500 cartridge tapes. To obtain the 1-percent sample of all magnetic media cases for the pilot study (or for that matter, to fully implement the new coding system for years earlier than 1993), all of the cartridges would have had to have been retrieved and the 1-percent sample cases extracted. Because this was not feasible, a sample of cartridges was selected from the full set.

The 1,500 cartridges were produced in about 1,250 batches as outputs from a daily computer operation. The volume of W-2 items varied greatly from one batch to another, with a range of approximately 200 items to 900,000 items per batch. A paper listing of the cartridge numbers and the number of W-2 items in each batch provided a sampling frame from which to draw a cluster sample of W-2 items. In the first stage of selection, a sample of 125 batches were drawn from the full population with probability proportionate to the number of W-2 items in the batch. This yielded 133 cartridges. These cartridges were then processed obtaining the 1-percent sample cases. The sample selection process yielded 311,453 cases, about one-fourth the size of the original 1-percent sample magnetic media records.

Coding

The subsample of the magnetic media cases selected were put through the new geographic coding operation. About 90 percent of the sample cases received valid geographic codes based on the city name and ZIP Code. For those magnetic media cases for which a valid code could not be obtained, the geographic location of the employer, taken from the Single Unit Code File or Multiunit Coding File, was used. For all EE-ER records for which only

paper W-2's existed, the employer-based coding files were also used to provide a State and county code.

Establishing Case Weights

The magnetic media cases that were coded under the new system were matched back to the original 1991 EE-ER File to provide a data set from which changes in geographic codes at the employee-employer level could be assessed. The resulting sample that included the pilot subsample of magnetic media cases and the full 1-percent sample of nonmagnetic media cases was not a self-weighting sample. The selection probability for the nonmagnetic media cases was 1/100; but the selection probabilities for the magnetic media subsample varied. Although the second stage selection probability for these cases was constant, also 1/100, the first stage selection probability varied from batch to batch and was proportional to the number of W-2's in the batch. Thus large batches had a larger probability of selection than smaller batches. (This, of course, was the intention since the purpose of the cluster sample was to get a large number of 1-percent sample cases for the pilot study from a small number of cartridges. Roughly 10 percent of the cartridges yielded 25 percent of the 1-percent sample cases.)

Case weights for the pilot study were constructed in two stages. First, each sample case was assigned a weight equal to the reciprocal of its probability of selection. For the nonmagnetic media cases this weight was 100. For the magnetic media subsample, the weight was 100* (the reciprocal of the first stage selection probability). In the second stage, a post-stratification adjustment was made to ensure that the pilot study estimated totals by original State code and coverage code were the same as the original 1-percent EE-ER sample. The coverage code defines the type of employment, such as agriculture, household, State and local government, Federal civilian, and the military. The

second stage adjustment factors were obtained by (1) tabulating the full 1-percent sample and the pilot study subsample (using the first stage case weights) by original State and coverage group, (2) forming ratios by dividing the first table by the second, and (3) for each pilot study case, multiplying the first stage case weight by the appropriate ratio depending on the original State code and coverage group of the case.

The reason for post-stratifying by original State was to improve the estimates of the number of changes in State code induced by the change in the coding system. Coverage code was included in the post-stratification scheme because certain undesirable effects of the cluster sample in the first stage regarding the possibility of selecting the entire roster of 1-percent sample cases from an individual employer.

The possibility of selecting the entire roster of workers from a single employer might add to the variability of estimates from the pilot study sample to the extent that the change in geographic code for an individual worker is largely the function of the employer-based code assigned in the original EE-ER File. In particular, when the full roster of workers of an employer with special characteristics such as a military employer or a Federal civilian employer, is selected into the sample with probability less than 1, the weighted estimates for workers with that class of employer may be substantially overstated. To the extent that those same workers are likely to have differences in geographic codes under the old and the new system, estimates of geographic change from the pilot sample may be overstated. Controlling for coverage group of the employer in the post-stratification scheme can help reduce the variability of the estimates due to the selection of entire rosters of workers for employers with special characteristics.