(Reprinted from October 1978, page 14)

The 1973 CPS-IRS-SSA Exact Match Study

by Beth Kilss and Frederick J. Scheuren*

The 1973 CPS-IRS-SSA Exact Match Study—a joint undertaking of the Social Security Administration and the Bureau of the Census—links survey records for persons in the March 1973 Current Population Survey to their respective earnings and benefit information in SSA administrative records and to selected items from their 1972 Internal Revenue Service individual income tax returns. The resulting set of files provides a very broad base for cross-section and longitudinal analyses of income-distribution questions. This article attempts to provide an overview of the techniques employed in the study. Among the topics discussed are the confidentiality requirements in force during the project. The original study goals are also described and a list of some of the completed research is provided.

An overview and summary of the 1973 Exact Match Study is given here: How it was carried out, where it stands today, and what the future might hold. Much of the material has already appeared in one form or another in an Office of Research and Statistics series of reports, **Studies From Interagency Data Linkages,** or in papers delivered before the American Statistical Association and elsewhere over the past 4 years (see list, page 20). What is attempted in this retelling is to construct a complete story of the study.

Organizationally, the article is divided into three sections: (1) A description of the background of the study, its research objectives, and the confidentiality requirements under which it has to be conducted; (2) discussion of the procedural steps taken in the project; and (3) some brief speculations on the future usefulness of the data bases being created.

Study Background

The 1973 Exact Match Study has been a joint undertaking of the Bureau of the Census and the Social Security Administration (SSA). Its starting point was the March 1973 Current Population Survey (CPS). A match was then made between the CPS sampled individuals and their social security benefit and earnings records. A limited set of tax items from 1972 Federal income tax returns was also furnished to the Bureau of the Census by the Internal Revenue Service (IRS) for matching to the CPS as part of this project.

The 1973 effort represents a continuation by the Social Security Administration, the Bureau of the Census, and IRS of a long line of interagency data linkages for statistical purposes. Matching studies have been conducted to evaluate the last three decennial censuses, for example. In surveys conducted by the Bureau of the Census for the Social Security Administration[1] interview schedules are combined routinely with administrative information on SSA earnings and benefits. Two-way matches involving IRS and SSA statistical samples have also been fairly common.[2]

Administrative data linkages to the Current Popula-

*Division of Economic and Long-Range Studies, Office of Research and Statistics, Social Security Administration. This article is a brief version of a paper presented in March 1978 at a Social Security Administration Workshop on Policy Analysis With Social Security Research Files held in Williamsburg, Va.

The authors gratefully acknowledge the extensive assistance of Faye Aziz, Linda DelBene, and in particular, Roger Herriot and Benjamin Bridges in the paper's preparation. A longer version of this paper and the other workshop papers cited here will appear in the published proceedings of the workshop. **Policy Analysis With Social Security Research Files,** early in 1979.

[1] See, for example, Thomas Tissue, "The Survey of the Low-Income Aged and Disabled: An Introduction," **Social Security Bulletin,** February 1977, and Lenore E. Bixby et al., **Demographic and Economic Characteristics of the Aged: The 1968 Survey of the Aged** (Research Report No. 45), Social Security Administration, 1975.

[2] Warren Buckler and Creston Smith, **The Continuous Work History Sample (CWHS): Description and Contents,** and Peter K. Cook, **Estimating the Failure Rate of a Unique Identifier Match Procedure Using Social Security Numbers** (papers presented at the 1978 SSA Workshop on Policy Analysis With Social Security Research Files).

tion Survey began with the March 1964 CPS,[3] and have been carried out periodically ever since.[4] Although the 1973 Exact Match Study is generally similar, it departs in several respects from the earlier CPS projects, with some major differences in scope, methods, and objectives:

(1) The sample is many times larger than that used in any previous joint effort, consisting of more than 100,000 individuals aged 14 or older;

(2) the process used to link the data from the various sources is more automated than that used in past efforts (one reason a larger sample of cases could be matched); and

(3) the goals of the 1973 work have been much more ambitious (in keeping with the study's larger sample and improved data-processing methods).

Research Objectives

Serious gaps exist in our general knowledge of the overall income distribution of persons and families in the United States, despite the ongoing statistical programs of IRS, SSA, and the Bureau of the Census—agencies that collect and publish detailed income information. The strengths and weaknesses of each of these efforts are well-known[5] and are discussed here only to the extent needed to put the research objectives of the 1973 Exact Match Study in proper perspective.

Chart 1 may provide the summary needed for this purpose. The chart looks at the content and coverage differences among the Continuous Work History Sample, the Statistics of Income sample for individual income taxfilers, and the March **Income Supplement** to the Current Population Survey. Two aspects are examined. Conceptual or definitional strengths and weaknesses are described first and are then qualified by the operational strengths or weaknesses of each sample. The CPS, for example, is incomparably richer in demographic information than either the IRS or SSA samples. It also may have the broadest coverage of money income sources. On the operational side, however, income reporting in the CPS is known to be deficient in relation to either of the administrative sources.

A natural consequence of comparisons such as those in chart 1 is to propose that the three statistical programs be merged in some way. Many attempts have been made to do this synthetically.[6] The 1973 Exact Match Study is the first large-scale attempt to build a dataset that actually brings together the income information of all three agencies.

The 1973 study was designed with a great number of specific goals in mind. These can be grouped under the general headings of "evaluation" and "augmentation," as chart 2 shows. The primary interest of the Bureau of the Census, for example, was to evaluate and, potentially, to find ways of improving upon the procedures it employs in carrying out the Current Population Survey. The SSA's primary interest was in creating an improved data base by augmenting the CPS with information from its own administrative record systems and from IRS income tax returns to address such policy issues as the redistributive effects of changes in income and payroll taxes and alternative social security benefit structures.

It is important to add that the Exact Match Study was also looked upon as an intermediate step in the construction of "corrected" personal income-size distributions of the U.S. population. Major follow-up projects to achieve this objective have been underway for some time in the Social Security Administration and at the Bureau of Economic Analysis.[7] Finally, to the extent that confidentiality requirements permitted, the Exact Match dataset was to be "published" in the form of public-use tapes to meet the research community's needs for improved information in the area of income distribution and redistribution.

Confidentiality Requirements

In the match project, as in earlier linkage efforts, great care has been taken to ensure the confidentiality of the shared information. The laws and regulations under which the IRS, the Bureau of the Census, and SSA operate impose very definite restrictions on such exchanges. To adhere to these provisions (chart 3), special operating procedures were instituted by the Social Security Administration and the Bureau of the Census to guarantee that the linked data would be used only for statistical purposes and not administrative ones. (No processing of linked data was carried out at IRS.) Some of the steps taken include the following:

[3] J. Steinberg and L. Pritzker, "Some Experiences with and Reflections on Data Linkage in the United States," **Bulletin of the International Statistical Institute,** vol. 42, 1967, pages 786–805.

[4] See Gayle Thompson, "Work Experience and Income of the Population 60 and Older, 1971," **Social Security Bulletin,** May 1975, and Susan Grad, **Income of the Population Aged 60 and Older, 1971** (Staff Paper No. 26), Social Security Administration, 1977.

[5] See E. Budd, D. Radner, and J. Hinrichs, **Size Distribution of Family Personal Income: Methodology and Estimates for 1964** (Bureau of Economic Analysis Staff Paper No. 21), Department of Commerce, 1973.

[6] See Benjamin A. Okner, "Constructing a New Data Base from Existing Microdata Sets: The 1966 Merge File," **Annals of Economic and Social Measurement,** 1972, pages 325–42; J. Scott Turner and Gary B. Gilliam, **Reducing and Merging Microdata Files** (OTA Paper 7), Office of Tax Analysis, Department of the Treasury, 1975; and Daniel Radner, **The Statistical Matching of Microdata Sets: The Bureau of Economic Analysis 1964 Current Population Survey—Tax Model Match** (Ph.D. dissertation), Department of Economics, Yale University, 1974.

[7] For a discussion of some results from one of the projects, see Daniel Radner, **Age and Family Income** (paper presented at the 1978 SSA Workshop on Policy Analysis With Social Security Research Files).

When competing confidentiality regulations existed, the strictest provisions were followed.

Project computer tapes, whether they contained linked data or not, were stored in locked facilities when not in use.

All confidential information from the Bureau of the Census was in the custody of regular or special Census employees at all times.

As an added precaution, only limited extracts of the basic CPS and SSA data were used in most of the searching and matching work done at the Social Security Administration.

General Study Procedures

The 1973 study has produced a number of related computerized data bases that can be employed either alone or in conjunction with each other; six are being documented and distributed for public use, as shown in chart 4. Only the procedural steps taken to develop the most important of these datasets—the **1973 Current Population Survey—Administrative Record Exact Match File**—are discussed here. Confining attention to only this one file considerably simplifies the exposition of the way the project was conducted without leaving out any of its essential features. For the sake of completeness, footnotes (and an occasional aside) indicate which developments resulted in the five other files.

1973 Current Population Survey—Administrative Record Exact Match File

The major steps in the preparation of the **1973 Current Population Survey—Administrative Record Exact Match File** are set forth in chart 5, which is a highly abbreviated representation of a complex project. It does, however, provide the needed structure for the general description that follows. The study can be divided into four major parts:

Matching together the three CPS sources and obtaining potentially usable social security numbers (SSN's)—the three CPS sources are those numbered (1) to (3) in the chart;

extracting and editing two sources of social security earnings data, (4) and (5);

extracting and merging the remaining three administrative sources, (6) to (8); and

completing the data linkage and preparing the necessary weights and match codes.

CPS Data

The project was started by constructing a CPS data base, as the upper left portion of chart 5 indicates. The March 1973 CPS is the basic source to which all other information obtained was linked. For this study, the entire CPS sample was included (approximately 50,000 households).[8] Extensive information was obtained from—

the basic March demographic and labor-force items, as well as the **Supplement** questions on income and work experience;

the June CPS **Supplement** questions on monthly income and food-stamp recipiency[9] (two of the eight rotation groups interviewed in March were also eligible for interview in June); and

the March 1973 Control Card file, which is important because it contains the personal identification information (name, SSN, date of birth, etc.) needed for matching to administrative records.

To match the CPS to IRS and SSA data, the personal identifying information had to be perfected. This objective was accomplished by taking SSN's and names, etc., from the control-card tape and validating them against the corresponding items in SSA files. Where the SSN's were judged invalid or were missing, an attempt was made to replace or supply them through manual search procedures. Despite resource and other operational constraints, the machine-validation and manual-search procedures resulted in about 90,000 potentially usable numbers from the 100,000 persons aged 14 or older involved here.

SSA Earnings Data

Once the set of potentially usable SSN's had been obtained, the earnings part of the Exact Match data base was completed. The two social security earnings sources included were—

(1) an extract from SSA's computerized longitudinal summary earnings record (SER) files for all those for whom a potentially usable SSN could be determined (annual data were shown for 1972 and 1971 with a historical summary for earlier years); and

(2) for a subsample (approximately 25 percent) of persons with earnings at the 1972 taxable maximum, quarterly wage amounts, obtained by manually transcribing information from SSA microfilm records.

As chart 5 shows, these SSA sources were first merged with each other and then the combined file was matched to the CPS.

Two other files, not shown in the chart, were also created. One was the first public-use file from the project, the 1973 Current Population Survey—Summary

[8] All households were matched whether or not an interview was obtained in March. Only the interviewed households are discussed here, since the noninterviews have not been included on any of the public-use files from the study.

[9] Most June items were not retained, but they have been provided on a separate file—the **June 1973 Current Population Survey Exact Match File** (see chart 4).

Earnings Record Exact Match File (provided in 1975). The second was constructed by re-extracting from the basic longitudinal SER in 1977 a different and more up-to-date version of the SSA earnings data. This dataset is available for public use as the Longitudinal Social Security Earnings Exact Match File.

Chart 1.—Some overall strengths and weaknesses in the income-distribution information from selected statistical samples

Social Security Administration 1-Percent Continuous Work-History Sample[1]	Internal Revenue Service Statistics of Income Sample[2]	Bureau of the Census Current Population Survey[3]
Conceptual Issues	**Conceptual Issues**	**Conceptual Issues**
Income and other variables available. Annual data are available on covered wages and taxable self-employment. Quarterly data on covered taxable wages are available for nonagricultural wage workers. Data on monthly benefit entitlements are provided for social security claimants. Other items include industry of employer and place of employment (for workers) and detailed information about the program characteristics of each OASDI claimant or beneficiary. Data are basically longitudinal.	**Income and other variables available.** Annual data on money income subject to tax are available in great detail. Notable omissions include social security benefits and other transfer incomes (public assistance, unemployment compensation, etc.) and most noncash sources and fringe benefits (employer contributions to pensions and health plans for example). Longitudinal data (and most demographic information) are not usually provided. Definitions are subject to change by legislation; some historical comparisons are thus difficult.[5]	**Income and other variables available.** Annual data are available on virtually all money income types. Nonmoney income data are not usually provided. The demographic variables are extensive, unlike those in IRS or SSA samples; they include age, race, and sex; education; family status; and industry and occupation. Some longitudinal analysis is possible but not for periods longer than 16 months. With a few minor exceptions, CPS income concepts have remained virtually unchanged since 1947, when income data first began to be collected regularly in the survey.
Groups included. Coverage of social security claimants is complete. Coverage of wage earners and the self-employed is extensive. Some major groups are not covered, including Federal civil-service employees, railroad workers, and some employees of State and local governments or nonprofit institutions. (Despite the exclusions, 9 out of 10 U.S. workers engage in covered employment.) Important legislative changes in coverage have occurred over the life of the sample (some data go back to 1937), but for the most part these can be disentangled if desired.	**Coverage.** Restricted in that not everyone is required to file a tax return. Data are available on a tax-unit basis—not for each person or family.	**Groups included.** Coverage includes most of the resident U.S. population. The CPS is a sample of the entire civilian noninstitutional population residing in the 50 States and the District of Columbia. Armed Forces members are included but only if they are living off post or on post with their families. Income information is generally only obtained for persons aged 14 years or older.
Operational Issues	**Operational Issues**	**Operational Issues**
Quality of reporting. Reporting problems vary among the data items in the sample: Wages and benefit data are well reported; self-employment earnings data are less adequate; data on industry and place of employment have important deficiencies for some purposes.[4] Historically, wages above the taxable maximum have had to be estimated.	**Quality of reporting.** Reporting problems are relatively small overall, but the quality does vary from source to source. Wages are perhaps the best reported, and farm income (in part for conceptual reasons) may be the most unsatisfactory.	**Quality of reporting.** Reporting problems exist for all income types and are particularly severe for property and transfer incomes (other than social security benefits). Missing income information is also an important factor. The nonincome variables suffer from problems of reporting and nonreporting but are generally thought to have much smaller errors associated with them. Several major changes in collection techniques and data-processing methods make it difficult to use the CPS for time-series analyses.
Extent of compliance (with earnings reports). Thought to be extremely high for most workers. Some lack of complete coverage may exist for certain groups: Farm laborers, private household workers, and, perhaps, illegal aliens working at covered jobs.	**Extent of compliance.** The extent of compliance with the statutory filing requirements is unknown but believed to be quite high.	**Extent of coverage.** Unlike the IRS or SSA samples, the CPS is known to suffer from a small but not insignificant amount of undercoverage. This undercoverage is of two types: Missed households (where all members are omitted) and missed persons in enumerated households. A partial adjustment for undercoverage has always been made in the CPS, but its adequacy is questionable.

[1] For a more complete discussion of this sample, see Warren Buckler and Creston Smith, **op. cit.**

[2] For a more extended discussion of how this sample is presently constituted, see **Statistics of Income—1974, Individual Income Tax Returns.**

[3] For an up-to-date general discussion of the CPS sample, see Robert Hanson, **The Current Population Survey: Design and Methodology,** U.S. Bureau of the Census, Technical Paper No. 40 (January 1978). For an extended recent discussion of CPS income concepts and reporting issues, see **Consumer Income,** Series P-60, No. 105.

[4] See, for example, David W. Cartwright, **Major Limitations of CWHS Files and Prospects for Improvement** (paper presented at the 1978 SSA Workshop on Policy Analysis with Social Security Research Files).

[5] This series began in 1916 and is the oldest of the three. See **Statistics of Income—1965, Individual Income Tax Returns** for a complete historical summary.

Note: This chart is obviously a highly summarized evaluation that is likely to be changed significantly in the near future because of changes in all three data systems. In particular, at the Social Security Administration, the wage reports for 1978 and future years will be annual (not quarterly); the wage information (on Form W-2) will be for both covered and presently noncovered employment; the total, not just the taxable wage, will be provided. The Bureau of the Census is now introducing improvements in the CPS that may eliminate virtually all coverage errors arising from missed households. Experiments are also underway to try to reduce the amount of underreporting of income.

Chart 2.—Bureau of the Census and Social Security Administration research goals in the 1973 CPS-IRS-SSA Exact Match Project

Evaluation Goals

The study's evaluation goals relate mainly to the CPS. These goals and some of the others that ascertain the value of SSA estimation procedures are stated below.

1. To evaluate the CPS procedures used to impute for missing income information;

2. to evaluate the CPS procedures used to adjust for noninterview nonresponse;

3. to evaluate the CPS procedures used to adjust for survey coverage errors;

4. to evaluate (and "correct") CPS wage and property income reporting by comparing it with the corresponding information provided to the IRS or to SSA;

5. to evaluate (and "correct") CPS social security reporting by comparing it with SSA recorded amounts;

6. to evaluate the efficacy of SSA procedures for estimating covered wages above the taxable maximum; and

7. to evaluate payroll tax and other program simulations that SSA engages in when carrying out analyses of the implications of alternative policy decisions.

Augmentation Goals

The study's augmentation goals relate mainly to SSA policy research issues. They also have the important (but secondary) purpose of increasing general knowledge of distribution questions. Some particular goals were:

1. To augment SSA and IRS coverage so that the combined effect of income and payroll taxes can be looked at in the context of the total population;

2. to augment our understanding of the differences in the basic units of analysis in each statistical setting: tax returns versus persons versus social security claimants versus Census families;

3. to augment the SSA earnings and benefit information so that policy researchers can better examine the equity and cost implications of alternatives to the current social security benefit structure; and

4. to augment SSA's longitudinal earnings information with variables such as noncovered wages, education, and family status; this approach may lead to a better understanding of the lifetime covered employment patterns of persons who have contributed to the social security program.

Chart 3.—Basic confidentiality requirements for 1973 Exact Match Study

Bureau of the Census Requirements

Information derived from the Bureau of the Census is governed by policies and procedures established under title 13 of the U.S. Code. This title requires that information about identifiable individuals remain under the direct control of employees of the Bureau at all times. On rare occasions, to better achieve its statistical goals (such as in this linkage project), the Bureau swears in, as its own temporary employees, a small group of employees of other agencies. In this instance, those Social Security Administration employees directly involved in the linkage (about 15) were hired and sworn in as Census employees without compensation. These few individuals—technically employees of both agencies at once—have been given legal access to both Census and SSA data, so that the linkage could be performed. Both regular and "special" Census employees are sworn to uphold the confidentiality of all Census information and are subject to criminal penalties if they fail to do so.[1]

Social Security Administration Requirements

Information derived from Social Security Administration files is governed by title II of the Social Security Act and the regulations as established under that Act (specifically, Regulation No. 1, Section 401 and 422).[2]

To release SSA earnings and benefit information for identifiable individuals to the Bureau of the Census, a special Commissioner's decision had to be obtained. This decision, dated June 28, 1973, was made subject to the following conditions:

1. No identified SSA information was to be given to the Bureau of the Census for any CPS respondent who refused to give his social security number to the Census interviewer.

2. All SSA data given to the Bureau of the Census were to continue to have the protected treatment required by the social security laws and regulations. Furthermore, the data were to be subjected to that Bureau's own confidentiality restrictions as imposed under title 13.

3. After linkage, all individual identification must be removed or scrambled in the resultant file. (This requirement has since been modified to allow for a longitudinal continuation of the project. The extension is not indefinite and will be reviewed every 2 or 3 years.)

Internal Revenue Service Requirements

The Internal Revenue Service (IRS), under an executive order (promulgated under IRS Code, Section 6103) provided a magnetic tape file of abstracts of 1972 individual income tax returns to the Bureau of the Census for statistical purposes. Subsequently, IRS agreed to permit the Bureau to match a very limited amount of this data to CPS and SSA information, subject to these provisions:

(1) That individually identifiable IRS data continue to be subject to title 13 and the various IRS confidentiality restrictions (specifically, IRS Code Section 7213); and

(2) that after matching and removal of individual identifiers, IRS is to have veto power over any data item on subsequent match files to be prepared for SSA, if IRS believes that the inclusion of the data item could possibly result in disclosure.

As with Census data, unauthorized disclosure of SSA or IRS information is a punishable offense that can result in fines or imprisonment or both.

[1] Vincent P. Barabba, "The Right of Privacy and the Need to Know," **American Statistical Association Proceedings, Social Statistics Section, 1974,** pp. 33–38, and Robert Davis, "Confidentiality and the Census, 1790–1929," in **Records, Computers and the Rights of Citizens,** Report of the Secretary's Advisory Committee on Automated Personal Data Systems (Department of Health, Education, and Welfare), 1973, pp. 178–201.

[2] Joseph Steinberg and Heyman Cooper, "Social Security Statistical Data, Social Science Research and Confidentiality," **Social Security Bulletin,** 1967, pp. 2–14, and Lois Alexander and Thomas Jabine, "Access to Social Security Microdata Files for Research and Statistical Purposes: An Overview," **Social Security Bulletin,** August 1978. (Also presented at the 1978 SSA Workshop on Policy Analysis With Social Security Research Files.)

IRS Income and SSA Beneficiary Data

The last three sources brought together in the study were—

an extract from IRS's individual income tax return master file (IMF);

an extract from SSA's master beneficiary record (MBR) files, for all persons who were or had been claimants in or before February 1974; and

another (earlier) extract from SSA beneficiary files, for all persons who were active as claimants in or before December 1972.

The basic matching and extracting of the IRS data was carried out at the Bureau of the Census, in contrast to most of the other steps in the study, which were carried out by special agents of the Bureau of the Census in the Social Security Administration, using either SSA files or CPS files provided by the Census staff who were also working on the project.

Because of the way SSA's record-keeping system operates, two sets of old-age, survivors, and disability insurance beneficiary data were obtained. Benefit information is kept for administrative purposes on an entitlement basis; the survey concept requires reporting on a cash-payment basis. To try to assess this difference, therefore, information was obtained on 1972 benefits at two different times, December 1972 and February 1974, but even the December 1972 file departs from a cash concept in a number of ways. Most important perhaps, retroactive payments made in 1972 for the previous year are related in the records to the period of entitlement rather than the time of actual payment.

Creating the "Final" Dataset

The 1973 Match Project had to be conducted by using data systems developed and used principally for other purposes. In the early stages of the study the task was to try to minimize the impact of these operational restrictions on the quality of the data linkage. In creating the "final" dataset, various statistical adjustment techniques were also used to mitigate the effect of these restrictions on subsequent analyses.

The linkage of CPS, IRS, and SSA information was examined in two ways. First came an extended process of trying to distinguish between "good" and "bad" matches. These matching issues were addressed with the user in mind. The outcome was the creation of alternative rules about what is to be considered a "match."

Second, the matched data were reweighted to take account of the fact that some persons in the sample should have had administrative data, but, because of faulty or missing SSN's or other anomalies in the administrative data base, none could be found. Ten weights or "estimators" are provided on the file. These include three weights created in the standard processing of the CPS,[10] plus three other weights based on population totals that have been corrected for the Census undercount—an undercount of about 5 million individuals.[11] The remaining four "estimators" were developed solely for this project and incorporate both a Census undercount adjustment and "corrections" made necessary because the entire Census sample could not be matched properly.[12]

Future Prospects and Completed Results

It may be too early to decide whether the preparation of the exact match data tapes really will fully meet the goals for which they are intended. The research done so far does indicate, however, that most of the project's objectives will be achieved, at least in part. The expectation is that the data will prove useful in addressing not only the original goals but many others not thought of when the project was planned. It must be added, however, that although the exact match data base can be used to address many questions, the number of questions it can answer will be much smaller.

At the end of the article is a list (that is as nearly complete as possible) of the reports and papers that have been produced from the study. Included are 43 in all: Fifteen of these are concerned primarily with the methods used; the remaining reports provide research results. Published reports from the series, **Studies from Interagency Data Linkages,** are excluded from the list since most of these document the data bases and do not provide results (see chart 4).

Perhaps the most important observation that might be made about the datasets produced is that they remain essentially unfinished. In a sense, of course, all data are incomplete and await the introduction of the researcher's experience and prior subject-matter knowledge. More than this is meant, though, by that observation. In particular, resource constraints were severe in the study, with the consequence that important limitations, such as the incompleteness in the matching, must be faced. An attempt has been made to "correct" for these lim-

[10] Robert Hanson, **The Current Population Survey: Design and Methodology** (Technical Paper 40), Bureau of the Census, 1978.

[11] J.S. Siegel, **Estimates of Coverage of Population by Sex, Race, and Age: Demographic Analysis, 1970 Census of Population and Housing: Evaluation and Research Program** (PHE(E)-4), Bureau of the Census, 1974.

[12] Still other weights will be added to those mentioned here. These are being created in an attempt to further examine the sensitivity of the study's results to the problems of matching and survey undercoverage. See H. Lock Oh and Frederick J. Scheuren, "Multivariate Raking Ratio Estimation in the 1973 Exact Match Study," **American Statistical Association Proceedings, Section on Survey Research Methods, 1978.** See also Report No. 10 in the data linkage series (in preparation).

Chart 4.—Public-use datasets available from the 1973 CPS-IRS-SSA Exact Match Study

1. **The 1973 Current Population Survey—Summary Earnings Record Exact Match File**—the first public-use file available from the Exact Match Study. The content is limited to basic demographic information, plus CPS and SSA earnings items. (It is documented in Reports Nos. 5 and 6 in the series, **Studies From Interagency Data Linkages.**)

2. **The 1964 Current Population Survey—Administrative Record Pilot Link File**—the only publicly available file from the Pilot Link Study. It was prepared to give researchers a means of introducing a historical dimension to their analyses of the 1973 Match Study Files. To the extent possible, therefore, the item content on the 1964 file has been made identical with that of the 1973 files (Nos. 1 or 3 in this list). See Report No. 7 in the data-linkage series for more details.

3. **The 1973 Current Population Survey—Administrative Record Exact Match File**—the most important of the datasets prepared in the study. It contains virtually all the same items as the first 1973 public-use file, plus more CPS data, and all the information obtained in the study from SSA beneficiary records and IRS tax returns. Documentation will be found in Report No. 8 of the data-linkage series.

4. **The June 1973 Current Population Survey Exact Match File**—an extract of selected income and food-stamp information from the one-quarter of the June CPS that overlaps with the March 1973 CPS Study sample. Documentation will be found in Report No. 9 in the data-linkage series, still in preparation.

5. **The Longitudinal Social Security Earnings Exact Match File**—a longitudinal extract from the social security earnings records for matched adults in the 1973 study sample. Annual earnings amounts are available on the file from 1951 through 1976. Updates of this file are contemplated in the future. See Report No. 9 in the data-linkage series.

6. **The 1972 Augmented Individual Income Tax Model Exact Match File**—an independent sample containing the public-use data made available by IRS in its 1972 Statistics of Income Tax Model file; it has been augmented by the addition of selected SSA demographic information. See Report No. 9 in the data-linkage series.

Note: All the files listed above have already been released and are available either from the Social Security Administration or the National Archives Record Service (Machine Readable Archives Division (NNR), Washington, D.C. 20408). Readers should also note that many more datasets are available from the study than originally contemplated (see Frederick J. Scheuren and Barbara Tyler, "Matched Current Population Survey and Social Security Data Bases," **Review of Public Data Use, 1975,** pp. 7–10).

itations. The adjustments carried out, although grounded in statistical theory, undeniably have a subjective element that cannot be ignored. Researchers—depending on the questions they address—may therefore need to revise certain aspects of the exact match datasets before carrying out their analyses. To the greatest extent possible, the files and documentation have been designed to facilitate any reworking that might be deemed necessary.

Exact Match Study Reports*

Alvey, Wendy, and Cobleigh, Cynthia. "Exploration of Differences Between Linked Social Security and Current Population Survey Earnings Data for 1972," **American Statistical Association Proceedings, Social Statistics Section, 1975,** pp. 121–128.

Burkhauser, Richard. **An Economic Model of Early Social Security Acceptance** (paper presented at the 1978 SSA Workshop on Policy Analysis With Social Security Research Files).

Burkhauser, R.V., and Warlick, J. **Disentangling the Annuity and Redistributive Aspects of Social Security** (paper presented at the Annual Meeting of the American Economic Association, August 1978).

Cobleigh, Cynthia, and Alvey, Wendy. "Validating Reported Social Security Numbers," **American Statistical Association Proceedings, Social Statistics Section, 1974,** pp. 145–154.

Feldstein, Martin, and Pellechio, Anthony. **Social Security Wealth: The Impact of Alternative Inflation Adjustments** (paper presented at the 1978 SSA Workshop on Policy Analysis with Social Security Research Files.

Hendricks, Gary, and Peters, Elizabeth. **Social Security Coverage of Government Employees** (paper presented at the 1978 SSA Workshop on Policy Analysis With Social Security Research Files).

Herriot, Roger, and Spiers, Emmett. "Measuring the Impact on Income Statistics of Reporting Differences Between the Current Population Survey and Administrative Sources," **American Statistical Association Proceedings, Social Statistics Section, 1975,** pp.147–157.

Herzog, Thomas, and Scheuren, Frederick. "Dallying with Some CPS Design Effects for Proportions," **American Statistical Association Proceedings, Social Statistics Section, 1976,** pp. 396–401.

Herzog, Thomas. "More Dallying with CPS Design Effects," **American Statistical Association Proceedings, Social Statistics Section, 1977,** pp. 326–332.

Ireland, C. Terrence, and Scheuren, Frederick. **The Rake's Progress** (paper presented at the 1974 American Statistical Association meetings in St. Louis).

Johnston, Mary. "Evaluation of Current Population Survey Simulations of Payroll Tax Changes," **American Statistical Association Proceedings, Social Statistics Section, 1975,** pp. 495–500.

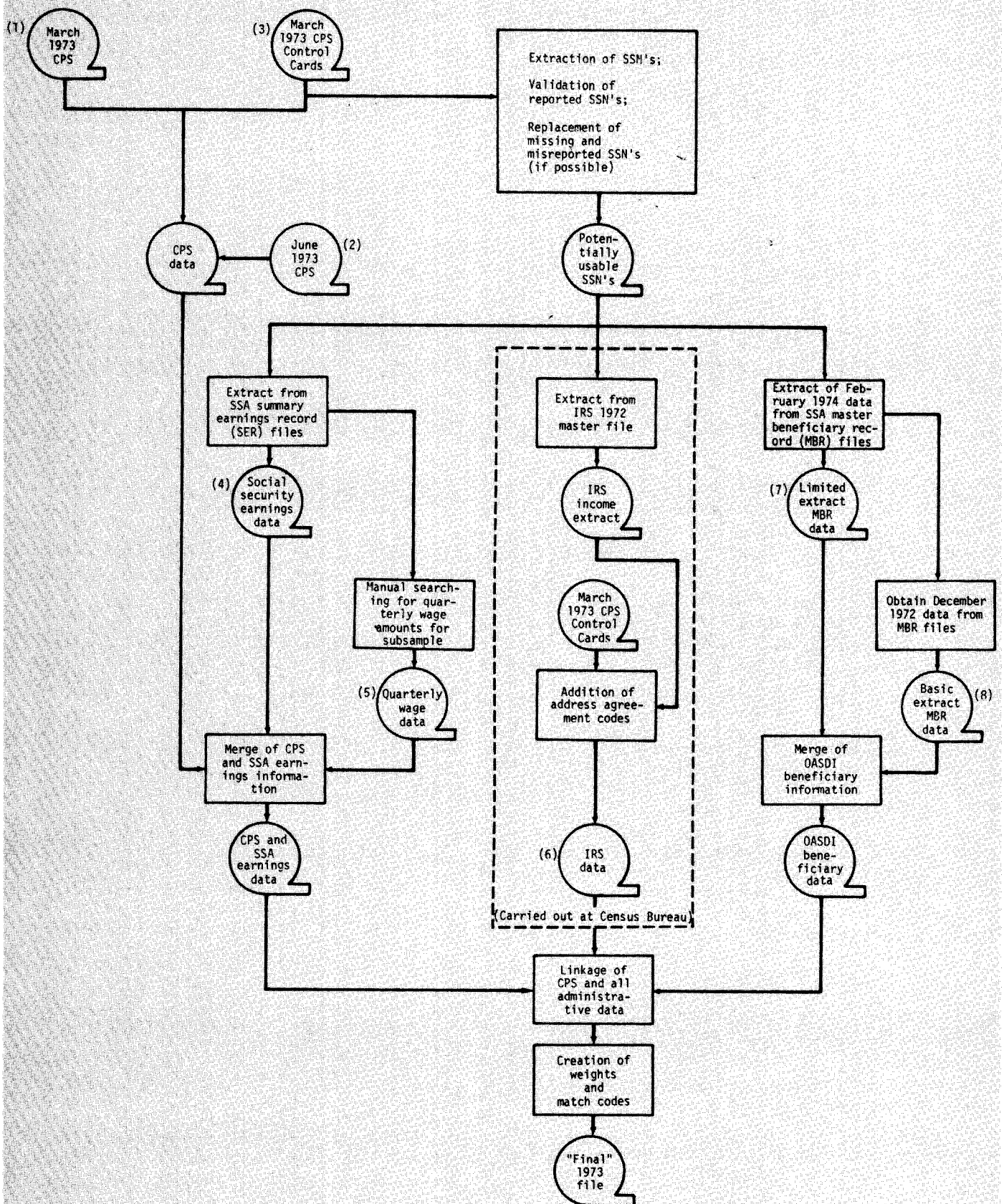
Kestenbaum, Bertram. "Evaluating SSA's Current Procedures for Estimating Untaxed Wages," **American Statistical Association Proceedings, Social Statistics Section, 1976,** pp. 461–466.

Kestenbaum, Bertram. "Men with Low OASDHI-Covered Earnings Not Counted as Poor in the CPS," **Social Security Bulletin,** January 1978, pp. 15-17.

Kestenbaum, Bertram. **Some Findings from the Exact Match Data Linkage** (Research and Statistics

*The list does not include the published reports in the series, **Studies From Interagency Data Linkages** (seven reports have already been issued, No. 8 is in press, and several others are in preparation).

Chart 5.—Development of the 1973 Current Population Survey—Administrative Record Match File



Note No. 4), Social Security Administration, May 1978.

Kestenbaum, Bertram, and Prero, Aaron. **Retirement Benefits Based on a Married Couple's Combined Earnings** (Research and Statistics Note No. 9), Social Security Administration, July 1978.

Kilss, Beth, and Alvey, Wendy. "Further Exploration of CPS-IRS-SSA Wage Reporting Differences for 1972," **American Statistical Association Proceedings, Social Statistics Section, 1976,** pp. 471–476.

Kilss, Beth, and Tyler, Barbara. "Searching for Missing Social Security Numbers," **American Statistical Association Proceedings, Social Statistics Section, 1974,** pp. 137–144.

Kleiner, Morris. **The Use of the CWHS for Labor Market Information: A Comparative Analysis** (paper presented at the 1978 SSA Workshop on Policy Analysis With Social Security Research Files).

Lancaster, Clarise, and Scheuren, Frederick. "Counting the Uncountable Illegals: Some Initial Statistical Speculations Employing Capture-Recapture Techniques," **American Statistical Association Proceedings, Social Statistics Section, 1977,** pp. 530–536.

Leimer, Dean. **Projected Rates of Return to Future Social Security Retirees Under Alternative Benefit Structures** (paper presented at the 1978 SSA Workshop on Policy Analysis With Social Security Research Files).

Millea, Mary, and Kilss, Beth. "Exploration of Differences Between Linked Social Security and Internal Revenue Service Wage Data for 1972," **American Statistical Association Proceedings, Social Statistics Section, 1975,** pp. 138–146.

Oh, H. Lock. "Osculatory Interpolation with a Monotonicity Constraint," **American Statistical Association Proceedings, Statistical Computing Section, 1977,** pp. 332–338.

Oh, H. Lock, and Scheuren, Frederick. "Multivariate Raking Ratio Estimation in the Exact Match Study," **American Statistical Association Proceedings, Section on Survey Research Methods, 1978.**

Oh, H. Lock, and Scheuren, Frederick. **New Methods for Comparing Income Data from Survey and Administrative Sources** (paper presented at Washington Statistical Society meeting), 1974.

Oh, H. Lock and Scheuren, Frederick. "Working Notes on the Derivation of 1972 Social Security Earner, Tax Return, and OASDI Beneficiary Counts for CPS-Eligible U.S. Civilians on April 1, 1973" (unpublished paper).

Pellechio, Anthony. **Old Age and Survivors Insurance (OASI)** (paper presented at the Annual Meeting of the American Economic Association, August 1978).

Pellechio, Anthony. The Effect of Social Security on Labor Force Participation of Potential Retirees (unpublished paper), Harvard University, 1977.

Pugh, Robert, and Silberman, Tom. **Simulation Model of Women Under Social Security: Initial Model File,** SSA Staff Paper No. 31, 1978.

Radner, Daniel. **Age and Family Income** (paper presented at the 1978 SSA Workshop on Policy Analysis With Social Security Research Files).

Radner, Daniel. **Federal Income Taxes, Social Security Taxes, and the U.S. Distribution of Income, 1972** (paper presented at the 15th General Conference of the International Association for Research in Income and Wealth, University of York, August 1977).

Sailer, Peter, and Vogel, Linda. "Exploration of Differences Between Linked Current Population Survey and Internal Revenue Service Income Data for 1972," **American Statistical Association Proceedings, Social Statistics Section, 1975,** pp. 129–137.

Scheuren, Frederick. **Methods of Estimation for the 1973 Exact Match Study** (paper presented at the 1977 Annual Meeting of the American Statistical Association).

Scheuren, Frederick, et al. Working Notes on Social Security Applications of Multivariate Raking (MULTIRAKE), 1976.

Scheuren, Frederick, and Oh, H. Lock. "A Data Analysis Approach to Fitting Square Tables," **Communications in Statistics,** 1975, pp. 595–615.

Scheuren, Frederick, and Oh, H. Lock. "Fiddling Around with Mismatches and Nonmatches," **American Statistical Association Proceedings, Social Statistics Section, 1975,** pp. 627–633.

Scheuren, Frederick, and Tyler, Barbara. "Matched Current Population Survey and Social Security Data Bases," **Review of Public Data Use** (Data Use and Access Laboratories), 1975, pp. 7–10.

Schultz, James. Liberalizing the Social Security Test—Who Benefits and What Would It Cost? (Unpublished paper), Brandeis University, 1976.

Stevens, Joyce, and Herriot, Roger. "Current Earnings Differentials of Men and Women: Some Exploratory Regression Analyses," **American Statistical Association Proceedings, Social Statistics Section, 1975,** pp. 673–678.

Vaughan, Denton, and Ireland, C. Terrence. "Adjusting for Coverage Errors in the March 1973 Current Population Survey," **American Statistical Association Proceedings, Social Statistics Section, 1975,** pp. 688–693.

Vaughan, Denton, and Yuskavage, Robert. "Investigating Discrepancies Between Social Security Administration and Current Population Survey Benefit Data for 1972," **American Statistical Association Proceedings, Social Statistics Section, 1976,** pp. 824–829.

Vogel, Linda, and Coble, Terry. "Current Population Survey Reporting of Social Security Numbers," **American Statistical Association Proceedings, Social Statistics Section, 1974,** pp. 130–136.

Yuskavage, Robert; Hirschberg, David; and Scheuren, Frederick. "The Impact on Personal and Family Income of Adjusting the Current Population Survey for Undercoverage," **American Statistical Association Proceedings, Social Statistics Section, 1977,** pp. 70–80.

Yuskavage, Robert, and Oh, H. Lock. **Four Alternative Estimates of CPS Income Data for 1972** (paper presented at the 1977 Annual Meeting of the American Statistical Association).