VALIDATING REPORTED SOCIAL SECURITY NUMBERS

Cynthia Cobleigh and Wendy Alvey
Social Security Administration

In the first two papers at this session, the discussion has focussed on the problem of incomplete reporting of social security numbers in the March 1973 Current Population Survey and what we have done about it so far. This paper will examine the degree to which the SSN's reported in the survey are "correct." Comparisons are made between name, race, sex, and date of birth, as shown on the CPS, and the SSA values for these items. In some cases, it is fairly obvious that the account number reported in the survey was incorrect. However, the confirmatory information is, itself, subject to response error, so some tolerance must be allowed. This paper describes some of the considerations that need to be taken into account in deciding what tolerances will be necessary. 1/

Section 1 provides an overview of the validation process and a brief examination of certain obvious errors that occurred. After eliminating from further analyses definable cases containing such errors, there remain persons for whom it is reasonable to construct CPS-SSA comparisons of demograhpic characteristics. For these individuals, sex and race differences are examined in section 2. In section 3, surname comparisons are made. Sections 4 and 5 deal with date of birth and age, respectively.

1. AN OVERVIEW

Every available SSN "belonging" to a survey person had to be checked. 2/ The purpose of this procedure was to determine whether the same number on the SSA administrative record and the CPS record was referring to the same individual. The search for a matching SSN was conducted on the Social Security Administration's Summary Earnings (Magnetic Tape) File, which contains a record for each of the 210 million account numbers ever issued. The SSN's sent through this search came from various sources.

When the survey data was placed on tape in Jeffersonville, it was keyed twice. The first time a person's record was constructed, it went onto what is termed the "A" file. In the second round, done by more experienced operators, these "A" records were verified, rekeyed when differences arose, and put on the "B" file. We were not necessarily convinced that the "B" file contained the more accurate data. There are subjective elements involved in the process, such as the ability to read the handwritten survey forms, which made this judgment inadvisable without further investigation. So, initially, we processed both. After validation procedures were completed, we determined that the "B" file actually did contain much more accurate data; therefore, in all following discussion, "reported SSN's" will refer to the "B" file only.

For these reported SSN's, some preliminary comparisons on sex, race, name, and date of birth were examined. Based on gross dissimiliarities of names and dates of birth, a subset of these SSN's were processed through the manual search. In addition, for persons without an SSN in the survey, the manual search described in the preceding paper was done.

Figure 1.—ACCOUNT NUMBERS BY SOURCE

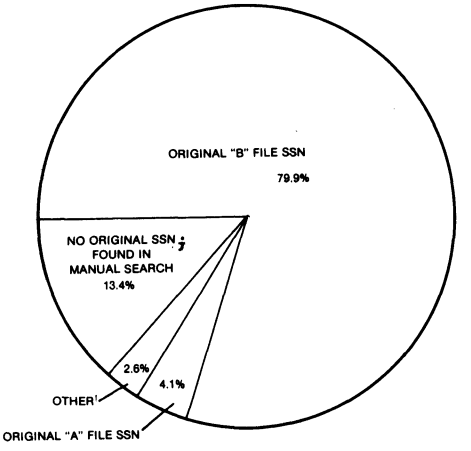


Figure 1 illustrates the various sources of all the account numbers that we now have for the 1973 Match Project. As can be seen, about four percent came from the "A" file; 80 percent came from identical "A" and "B" records or from a "B" record only. Another 13 percent of the total SSN's were obtained for persons who were missing a number in the survey. The additional manual search, carried out for SSN's judged to be incorrect, yielded 2.6 percent more numbers.

Originally Reported SSN's.--This paper will mainly concern itself with the subset of reported SSN's (i.e., "B" file cases) from which obvious errors have been excluded. Before limiting ourselves to this universe, though, some brief remarks about the types of errors encountered seem appropriate. About three percent of the SSN's are not now considered viable possibilities. Over half of them (2.06 percent) were found to be transposition or single digit errors. Furthermore, some persons reported an SSN belonging to

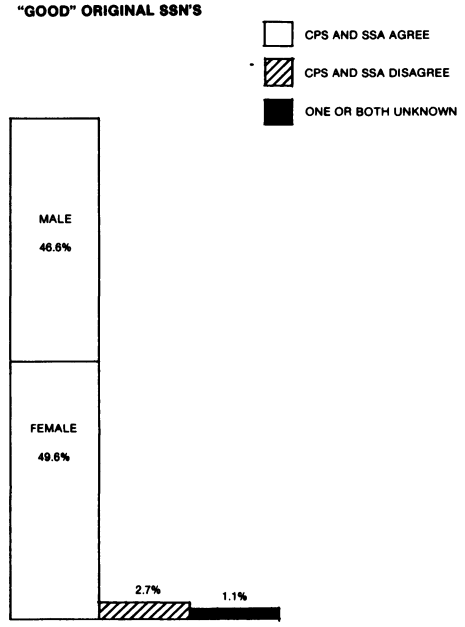
another household member (.53 percent). This could occur if, for example, a woman reported the number under which she collected benefits, rather than her own. The remaining SSN's (.45 percent) could not be located in the administrative files.

The causes of these errors are varied. Keypunching and copying mistakes are only two of the several possibilities. In any event, it is not meaningful to conduct a CPS-SSA comparison of demographic characteristics in such cases. Eliminating these errors from further discussion, we will now focus our attention on what can be considered the potentially "good" segment of reported SSN's. The first two characteristics to be considered are sex and race.

2. SEX AND RACE AGREEMENT

Agreement on Sex.--When considering sex alone, figure 2 illustrates that for 96

Figure 2.—AGREEMENT ON SEX FOR PERSONS WITH POTENTIALLY "GOOD" ORIGINAL SSN'S

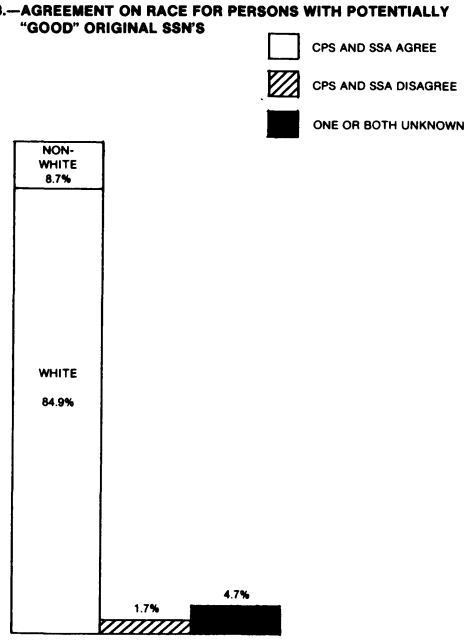


percent of the cases with potentially "good" original SSN's, CPS and SSA sex agree. In this group, females and males are represented in about the same proportions as in the population. Of the remaining persons, 1.1 percent reported sex as unknown on either one or both files, while 2.7 percent disagreed on sex.

An interesting pattern emerges among the cases whose CPS-SSA sex is in disagreement; almost two-thirds are SSA male/CPS female combinations. It can be ventured that a majority of these account numbers were reported by widows who gave their husband's SSN, under which they are receiving benefits. Since these women have their own number, the situation will be remedied when the benefit portion of the project is matched in. This step will substitute the correct SSN for the present one. 3/

Agreement on Race.--Considering race alone, figure 3 illustrates that, about 94

Figure 3.—AGREEMENT ON RACE FOR PERSONS WITH POTENTIALLY "GOOD" ORIGINAL SSN'S



percent of the time, the Census and SSA records agree on race, with whites and nonwhites being represented in the same ratio as in the population. The proportions of persons in the unknown and disagreement categories have reversed from what was noted for sex. Unknown race was reported on one or both files 4.7 percent of the time; only 1.7 percent had a CPS-SSA race disagreement.

Before examining agreement on sex and race combined, let us suggest some possible explanations for the different patterns between race and sex when they are unknown or disagree. As just mentioned, there is a much higher incidence of unknown race as opposed to unknown sex. The proportion of unknowns in the CPS and SSA files also differ, as can be seen below:

Table 1.-- Percentage of cases with unknown race or sex by source

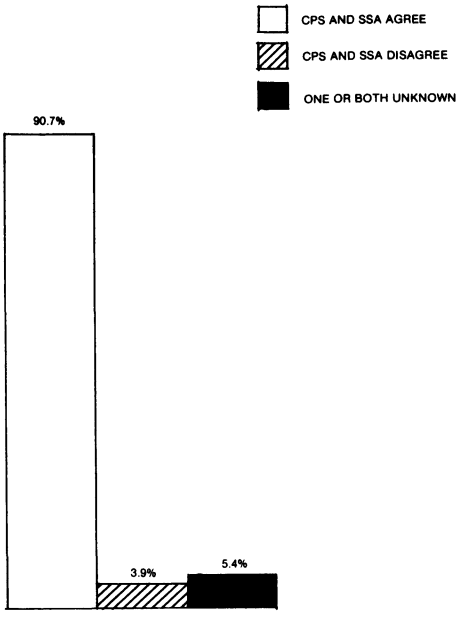
	Sex	Race
SSA unknown	0.14	2.90
CPS unknown	0.92	1.78
CPS and SSA unknown	0.00	0.04

The manner in which the SSA data was collected may be the source of some of the problem. Social Security information on race and sex is generally taken from the original application for a social security number, Form SS5. For a short period of time, IRS Form 3227 was used in cooperation with the Internal Revenue Service to provide SSN's for taxpayers who needed such identification. This form did not request race information. Although it was later changed to include this additional data, no effort was made to obtain a race for those who received account numbers in the interim; their race remained unknown.

It is likely that a substantial portion of the disagreement on race arises from the way it is collected in the CPS. Census interviewers do not ask the race question; the control card is filled in based on observation. Unless all household members are present to make another conclusion possible, the interviewer generally assumes that all related individuals are the same race. The respondent's race may be difficult to determine or other household members may be of a different race, and, as a result, race disagreements may occur.

Agreement on Sex and Race Combined.--Combining sex and race produces the results displayed in figure 4. Here, 91

Figure 4.—AGREEMENT ON SEX AND RACE COMBINED FOR PERSONS WITH POTENTIALLY "GOOD" ORIGINAL SSN'S

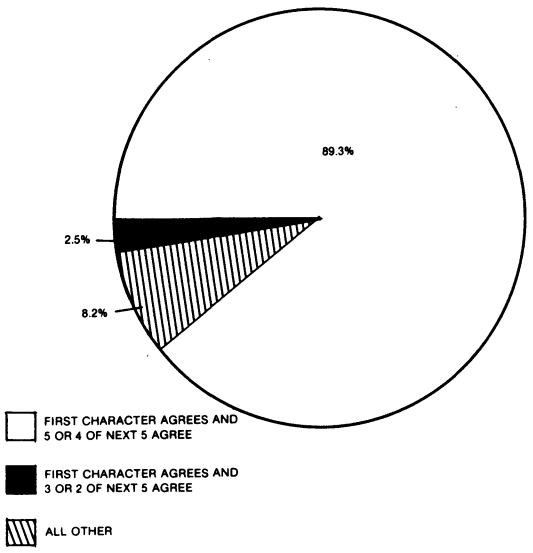


percent of the persons with potentially "good" original SSN's reported the same sex and race to both Census and Social Security. Unlike the distributions when sex and race are considered separately, the proportion of unknowns and disagreements are relatively equal. Sex-race combinations that disagree comprise 3.9 percent of the cases, while unknowns constitute the remaining 5.4 percent.

3. SURNAME AGREEMENT

In this discussion, the complete surname is not being compared; all references to surname pertain to no more than the first six characters. 4/ Figure 5 illustrates the results when names are examined character-by-character. Regardless of the other characteristics, we can say with relative certainty that the 89 percent who have surname agreement on the first character and four or five of the remaining characters are the same individual on both files.

Figure 5.—NAME AGREEMENT FOR PERSONS WITH POTENTIALLY "GOOD" ORIGINAL SSN'S



At first glance, it would appear as though all persons in the striped area (eight percent) should be discarded since the character-by-character surname disagreement is substantial. However, this particular group points out the problems that arise when only a single variable is used to evaluate a match. CPS women comprise 69 percent of this pie section; of these females, 59 percent are individuals whose other characteristics (sex, race, and date of birth) agree exactly. What has occurred for these women is that the name in SSA's administrative records is most probably a former one, and the SSN does refer to the same person on both files.

It is much more difficult to say anything definite about the 2.5 percent who lie between the two extremes. It will prob-

ably prove true that no single decision will resolve all the cases in this fuzzy in-between group. Those whose "demographics" strongly agree are likely to be victims of a spelling or keypunching error. However, obviously, as more disagreement on various characteristics appears, we become more inclined to conclude that the CPS and SSA persons are not the same.

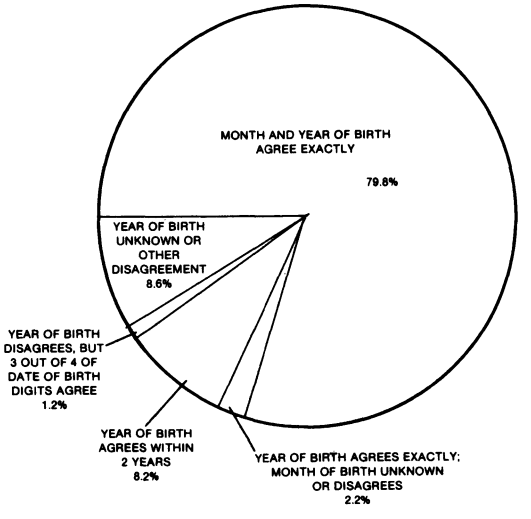
Instead of taking into consideration six characters of the surname, we could apply a number of other rules, one of which is to confine the evaluation to only the first four characters. Such a standard has been used by the Internal Revenue Service. 5/ According to the IRS rule, two surnames are judged the same if the first character agrees and, of the next three, there is no more than one difference or one transposition. Referring back to figure 5, if the first character agrees and, at most, one of the remaining five disagrees, this standard would be satisfied. Most of the cases in the remaining two pie sections would be rejected, however. The main exceptions are the previously-mentioned women whose former surname appears on the administrative record. IRS validates only the husband's SSN on a joint return, and, in actuality, these women would not even be examined by them.

4. DATE OF BIRTH AGREEMENT

Date of birth has been of use in evaluating surname. It merits some remarks as a variable in itself.

Figure 6 provides an illustration of the distribution of date of birth agreement. In 80 percent of the cases, persons reported exactly the same month and year of birth to both Census and Social Security. Even though this exact agreement is somewhat less than that cited earlier for sex and race, it still indicates a very respectable rate. In about two percent more of the cases, year of birth agrees exactly, but month of birth either disagrees or is unknown. The rest of the records contain disagreeing years of birth. However, at least half of them are similar enough to be acceptable for matching if other characteristics agree, such as sex, race, and surname (i. e., the eight percent which agree within two years and the one percent for which three out of four of the date of birth digits agree). The remaining nine percent of these cases either have an unknown year of birth or some other disagreement on year of birth.

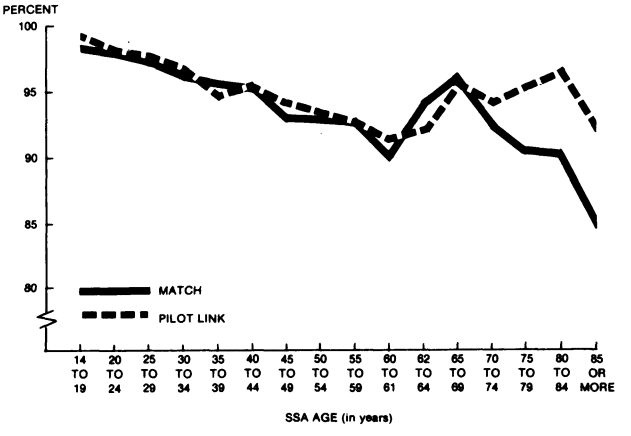
Figure 6.—BIRTH AGREEMENT FOR PERSONS WITH POTENTIALLY "GOOD" ORIGINAL SSN'S



5. AGE AGREEMENT

Overall Agreement by Age Class.--In this section, we will confine our examination of age agreement to those persons who reported the same race and sex to both Census and Social Security. Figure 7 presents the total distribution of such cases by age class agreement. 6/ Here the 1973 Match information (the solid line) is compared to an earlier Pilot Linkage Project (the broken line) which was also conducted jointly by SSA and the Census Bureau. SSA age group is plotted on the horizontal axis; the vertical axis shows the percentage of individuals in each SSA age group who were in the same age class in the CPS.

Figure 7.—AGE CLASS AGREEMENT IN MATCH AND PILOT LINK STUDIES FOR PERSONS WITH SSA AGE 14 OR OLDER HAVING POTENTIALLY "GOOD" ORIGINAL SSN'S AND AGREEMENT ON RACE AND SEX



It is quite evident from looking at the Match data that there is a marked downward trend in age class agreement as age in-

creases. One possible explanation for this is that, as time elapses since the date of birth was originally reported to Social Security, awareness of SSA age and, hence, consistency in CPS-SSA age reporting decline. Three points, however, seem to vary from the trend. At ages 60 to 61, the amount of age class agreement drops below the expected level (to 90 percent). This may be partially due to the fact that this age class is smaller than the rest (and, hence, specifications for falling in the cell are more stringent). The next two age groups show agreement which is decidedly higher than the trend. This occurs at the time in these individuals' lives when they are applying for or, at least, inquiring about Social Security benefits. 7/ It is, therefore, reasonable to suggest that not only they, but also members of their households, would be more aware of their SSA age. Consequently, CPS-SSA age reporting would tend to be more consistent.

Up to the 65 to 69 year age group, the trends we have commented on in the Match data also occur for the Pilot Link Project. The two studies do not, however, behave in the same way for older persons: the agreement rates for persons 70 years or older in the Link study tend to be a good bit better than those in the Match. This may be due to the fact that SSA benefit data were used to improve the Pilot Link age information; the 1973 Match Project has not yet made use of this additional data.

Age Agreement by Race and Sex.--Figure 8 looks at those who agree on age class by race and sex. Here again, the Match and Pilot Link studies are compared. It is immediately evident that the age class agreement distributions by race and sex are practically identical for the two samples. In both studies, about 96 percent of the whites, both female and male, reported ages to SSA and CPS which fall in the same age group; about 90 percent of the nonwhites, again regardless of sex, agree on age.

Since there is so little difference in age agreement between the sexes, it seemed unnecessary to look at males and females separately by age group. We do, however, consider it worthwhile to look at age class agreement by race (see figure 9). As can be seen, the curve for whites (the solid line) is quite flat, with only a slight downward trend. The dip at ages 60 to 61 and the subsequent rise in agreement rates for 62 to 64 and 65 to 69 year-olds are still apparent, but they have become noticeably dampened as compared to what they were in figure 7. Nonwhites (the broken line), on the other hand, show a steeper decline in age class agreement as age increases. Furthermore, the three departures from the trend which occur in the 60 to 69 year-old age period are much sharper. This obvious difference in age reporting by race is, no doubt, due to multiple causes, and no one good explanation can be ventured at this time. However, it should be mentioned that the patterns shown here are similar to those found in other studies which have looked at differences in age reporting by race [e.g., 36, 101, 139, 147].

Figure 8.—AGE CLASS AGREEMENT IN THE MATCH AND PILOT LINK STUDIES FOR PERSONS WITH POTENTIALLY "GOOD" ORIGINAL SSN'S BY RACE AND SEX

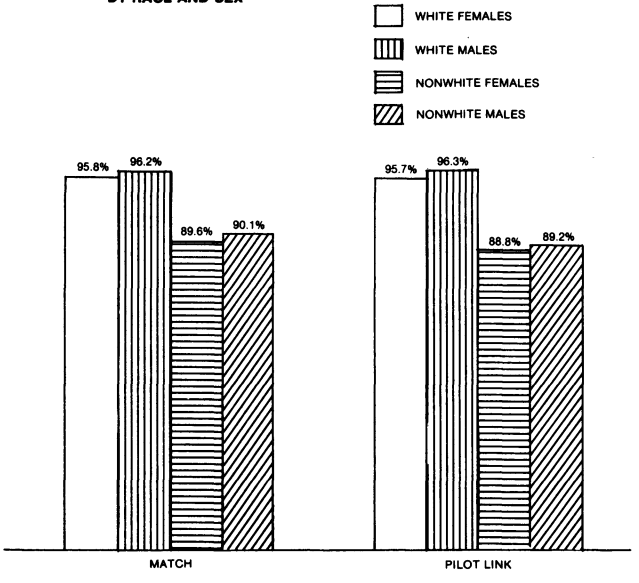
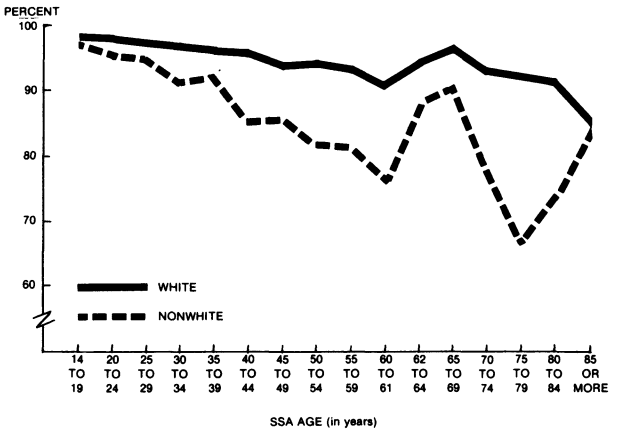


Figure 9.—AGE CLASS AGREEMENT FOR PERSONS IN THE MATCH STUDY WITH SSA AGE 14 OR OLDER HAVING POTENTIALLY "GOOD" ORIGINAL SSN'S AND AGREEMENT ON RACE AND SEX



Conclusion.--This paper's analysis of the "correctness" of originally reported SSN's

is a preliminary one. Even so, several interesting patterns have emerged. Certain common demographic characteristics have been examined, including sex, race, surname, date of birth, and age. Overall, once obvious errors are eliminated from consideration, the extent of agreement between the CPS sample information and SSA's records is quite high:

1. SSA-CPS sex and race are the same 96 percent and 94 percent of the time, respectively.

2. For surname, about 89 percent of the persons whose SSN's were matched agreed closely enough to be considered the same individual. Because of the use of a former name, the amount of agreement for males and females differed significantly (92 percent versus 87 percent).

3. Exact agreement on month and year of birth was about ten percent less than that for name. When disagreements on date of birth were examined, it became evident that as age increases, agreement decreases. One major exception was observed: agreement improves substantially at the age of retirement. Furthermore, age agreement for nonwhites was considerably less than for whites, a pattern which exists in other studies.

The next step in the Project will be to look at all the common CPS-SSA characteristics simultaneously and decide exactly what tolerances should be established in the final matching.

FOOTNOTES

1/ Striking the balance between nonmatches and mismatches not only involves employing some of the theory on "optimum" matching rules [23, 27, 70, 125, 151], but also requires decisions as to what adjustment it will be possible to make for nonmatches and mismatches in the subsequent analysis. For dealing with nonmatches, one of the techniques we expect to employ is a "raking" procedure, presented in a 1974 SSA paper entitled "The Rake's Progress." For dealing with mismatches (which cannot be entirely eliminated no matter what matching rule is chosen), we hope to use adaptations of standard methods that are robust against mismatches [73]. One such adaptation that may have some merit is contained in a series of 1974 SSA papers entitled "Fitting Square Tables with Nonsquare Procedures." All of these papers are available upon request. (See footnote 8 in the session introduction for the address.)

2/ The validation included a few cases of persons less than 14 years of age with a reported account number.

3/ As these women generally seem to be using their husband's SSN rather than their own, it is likely that this is the number they reported to IRS. Thus, in order to obtain tax information for these individuals, it will probably be necessary to use their reported number rather than their correct SSN for matching to IRS' files.

4/ While the complete last name is available from the CPS, only the first six characters of the surname were available on the Summary Earnings File. When benefit data is matched in, we will be able to make a more extensive comparison for persons who are beneficiaries.

5/ As the Internal Revenue Service is the third source of data to be utilized in the Match project, it seemed beneficial to mention what effect its guidelines would have had. The purpose of the IRS standard for matching surnames is that the posting of social security numbers to the tax files can be checked during data processing of the returns. Surnames have also been compared elsewhere [e.g., 91, 95] by using procedures other than character-by-character checking.

6/ Both the CPS and SSA ages were calculated as of the end of 1972, using year of birth. (For the Pilot Link, age is as of the end of 1963.) Unfortunately, only two digits of the year of birth are available on the Summary Earnings File. Because of this, it was not always possible to distinguish between persons under 14 and those 100 years of age or older. Thus, we cannot be sure that the 85+ age class reflects the behavior of all sample persons with an SSA age over 84 years.

7/ Eligibility for retirement benefits begins at age 62, although most people do not retire until they reach the age of 65.

SELECTED BIBLIOGRAPHY ON THE MATCHING OF PERSON RECORDS FROM DIFFERENT SOURCES

This bibliography was selected principally from the following three sources:

U. S. Bureau of the Census. Indexes to survey methodology literature, Technical Paper No. 34, 1974.

Forsythe, J. List of References on Results and Methodology of Matching Studies, 1966 (Unpublished Census Bureau Memorandum).

U. S. Public Health Service. Use of vital and health records in epidemiologic research, Vital and Health Statistics, series 4, no.7, 1968.

Scope

The references listed here are restricted essentially to published information on the results and methodology of matching person records. No material relating to establishment matching is included. Several compilations of articles exist on the confidentiality issues raised by record linkages. As a rule, therefore, we have excluded these citations from the bibliography. Two recent such compilations are:

Report of the President's Commission on Federal Statistics, vol. 1, 1971, pp. 246-254.

U. S. Department of Health, Education, and Welfare. Records, Computers, and the Rights of Citizens: Report of the Secretary's Advisory Committee on Automated Personal Data Systems, 1973, pp. 298-330.

Some other limitations should also be mentioned:

1. The listing does not contain citations to studies which began with an administrative record and then drew a sample of cases to be interviewed. (Excluded from the bibliography, therefore, are basically all reverse record check studies of financial characteristics, as well as prospective epidemiological studies.)

2. Only references to "exact" matching are included. Synthetic or "statistical" matching studies are not shown. (For citations to the literature on synthetic matches, see Radner, D. B. The Statistical Matching of Microdata Sets: the Bureau of Economic Analysis 1964 Current Population Survey--Tax Model Match, Yale Univ., New Haven, 1974. See also the Annals of Economic and Social Measurement, vol. 3, 1974, where several articles on synthetic matching are presented.)

Completeness of coverage

The bibliography's coverage of major U. S. studies involving linkages between survey (or census) schedules and administrative records is believed to be reasonably complete for the period 1950-1973. However, only a few references are given to work done outside the United States and to research engaged in before 1950.

FOOTNOTES

1/ This bibliography was compiled by Fritz Scheuren and Wendy Alvey with the assistance of Irene Jacobs, Office of Research and Statistics Library, Social Security Administration, and Patricia Fuellhart and her staff, Statistical Research Division, Bureau of the Census.

[1] Acheson, E. D. The Oxford Record Linkage Study, Report on the Second Year's Operations, Oxford Regional Hospital Board, Oxford, 1963.

[2] Acheson, E. D. Oxford Record Linkage Study, Brit. Jour. Prev. and Soc. Med., vol. 18, 1964, pp. 8-13.

[3] Acheson, E. D. The Oxford Record Linkage Study, a review of the method with some preliminary results, Proc. Roy. Soc. Med., vol. 57, 1964, pp. 11 ff.

[4] Acheson, E. D., and Evans, J. G. The Oxford Record Linkage Study, Biometrics, vol. 2, 1963, pp. 367 ff.

[5] Anderson, O. W., and Feldman, J. J. Family Medical Costs and Voluntary Health Insurance: a Nationwide Survey, McGraw-Hill, New York, 1956.

[6] Anderson, O. W., and Sheatsley, P. B. Comprehensive Medical Insurance, Health Information Foundation Research Series, No. 9.

[7] Anderson, R., and Anderson, O. W. A Decade of Health Services, Univ. of Chicago, Chicago, 1967.

[8] Bachi, R., Baron, R., and Nathan, G. Methods of record-linkage and applications in Israel, Bulletin of the International Statistical Institute, vol. 42, Proc. of 36th Session, Sydney, 1967, pp. 751-765.

[9] Bahn, A. K. Methodological study of population of out-patient psychiatric clinics, Maryland, 1958-59, Public Health Monograph No. 65., PHS Pub. no. 821, 1961.

[10] Balamuth, E. Health interview responses compared with medical records, U. S. Public Health Service, Vital and Health Statistics, series 2, no. 7, 1965.

[11] Bancroft, G. The American Labor Force; Its Growth and Changing Composition, Wiley, New York, 1958, pp. 151-175.

[12] Belloc, N. B. Validation of morbidity survey data by comparison with hospital records, J. Amer. Stat. Assn., vol. 49, 1954, pp. 832-846.

[13] Binder, S. Present possibilities and future potentialities, The Use of Vital and Health Records for Genetic and Radiation Studies, United Nations, New Ysrk, 1962, pp. 161-170.

[14] Bixby, L. E. Income of people aged 65 or older: overview from the 1968 survey of the aged, Social Security Bulletin, report no. 1, 1970, pp. 28-34.

[15] Brounstein, S. H. Data Record Linkage Under Conditions of Uncertainty, 7th Annual Conference of the Urban and Regional Information Systems Association, Los Angeles, California, 1969.

[16] Chase, H. C. A study of infant mortality from linked records: method of study and registration aspects, U. S. Public Health Service, Vital and Health Statistics, series 20, no. 7, 1970.

[17] Christenson, H. T., Cultural relativism and premarital sex norms, Amer. Soc. Rev., vol. 15, 1960.

[18] David, M., Gates, W., and Miller, R. Linkage and retrieval of micro-economic data. Heath, Lexington, Mass., 1974.

[19] Davidson, L. Retrieval of misspelled names in an airline passenger records system, Communications of the ACM, vol. 5, 1962, pp. 169-171.

[20] Densen, P. M., and Shapiro, S. Research needs for record matching, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1963, pp. 20-24.

[21] DuBois, Jr., N. S. D. On the problem of matching documents with missing and inaccurately recorded items (preliminary report), Annals of Mathematical Statistics, vol. 35, 1964, p. 1404.

[22] DuBois, Jr., N. S. D. A document linkage program for digital computers, Behavioral Science, vol. 10, 1965, pp. 312-319.

[23] DuBois, Jr., N. S. D. A solution to the problem of linking multivariate documents, Jour. Amer. Stat. Assn., vol. 64, 1969, pp. 163-174.

[24] Dunn, H. L. Record linkage, American Journal of Public Health, vol. 36, 1946.

[25] Dunn, H. L., and Grove, R. D. Completeness of birth registration in the United States, Estadistica, vol. 1, 1943, pp. 3-17.

[26] Elinson, J., and Trussell, R. E. Some factors relating to degree of correspondence of diagnostic information obtained by household interviews and clinical examinations, American Journal of Public Health, vol. 47, 1957, pp. 311-321.

[27] Fellegi, I. P., and Sunter, A. B. A theory for record linkage, Jour. Amer. Stat. Assn., vol. 64, 1969, pp. 1183-1210.

[28] Ferber, R. The Reliability of Consumer Reports of Financial Assets and Debts, University of Illinois, Urbana, 1966.

[29] General Register Office. 1951 Census of England and Wales, General Report, London, 1958, pp. 41-45, 5056.

[30] Goldberg, I. D., Goldstein, H., Quade, D., and Rogot, E. The use of vital records for blindness research, Amer. Jour. Pub. Health, vol. 54, 1964, pp. 278-285.

[31] Goldberg, S. Discussion of data storage and linkage, Bulletin of the International Statistical Institute, Proc. of 36th Session, Sydney, vol. 42, 1967, pp. 806-808.

[32] Grove, R. D. Studies in the completeness of birth registration, U. S. Bureau of the Census, Vital Statistics -- Special Reports, vol. 27, no. 18, 1943.

[33] Guralnick, L., and Nam, C. B. Census-NOVS study of death certificates matched to census records, The Milbank Memorial Fund Quarterly, vol. 37, 1959, pp. 144-153.

[34] Haber, L. D. Evaluating response error in the reporting of the income of the aged: benefit income, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1966.

[35] Haenszel, W., Loveland, D. B., and Sirken, M. G. Lung-cancer mortality as related to residence and smoking histories. I. white males, Journal of the National Cancer Institute, vol. 28, no. 4, 1962, pp. 947-1001.

[36] Hambright, T. Z. Comparability of age on the death certificate and matching census record, U. S., May-Aug. 1960, U. S. Public Health Service, Vital and Health Statistics, series 2, no. 29, 1968.

[37] Hansen, M. H. The role and feasibility of a national data bank, based on matched records and interviews, Report of the President's Commission on Federal Statistics, vol. 2, pp. 1-63.

[38] Hauser, P. M., and Kitagawa, E. M. Sscial and economic mortality differentials in the United States, 1960: outline of a research project. Proc. Amer. Assn. Soc. Stat. Sec., 1960, pp. 116-120.

[39] Hauser, P. M., and Lauriat, P. Record matching--theory and practice (abstract), Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1963, p. 25.

[40] Hedrich, A. W., Collison, J., and Rhoads, F. D. Comparison of birth tests by several methods in Georgia and Maryland, U. S. Bureau of the Census, Vital Statistics -- Special Reports, vol. 7, no. 60, 1939, pp. 681-695.

[41] Hoel, P. G., and Peterson, R. P. A solution to the problem of optimum classification, Annals of Mathematical Statistics, vol. 20, 1949, pp. 433-438.

[42] Hogben, L., Johnstone, M. M., and Cross, K. L. Identification of medical documents, British Med. J., 1948, pp. 632-635.

[43] Horn, W. Reliability Survey: a survey on the reliability of responses to an interview survey, Het PTT-bedrijf, vol. 10, 1960, pp. 105-156.

[44] Horn, W. Nonresponse in an interview survey: some specific phenomena (a case-study), Het PTT-bedrijf, vol. 12, 1963, pp. 11-19.

[45] Jaro, M. Unimatch--a computer systen for generalized record linkage under conditions of uncertainty, Spring Joint Computer Conference, 1972, AFIPS--Conference Proceedings, vol. 40, 1972, pp. 523-530.

[46] Johnson, Jr., C. E. Consistency of reporting of ethnic origin in the Current Population Survey, Technical Paper No. 31, 1974, pp. 23-27.

[47] Kaplan, D. L., Parkhurst, E., and Whelpton, P. K. The comparability of reports on occupation from vital records and the 1950 census, Vital Statistics -- Special Reports, vol. 53, no. 1, 1961, pp. 1-27.

[48] Kelly, T. F. Factors affecting poverty: a gross flow analysis, The President's Commission on Income Maintenance Programs: Technical Studies, 1970, pp. 1-82.

[49] Kelly, T. F. The creation of longitudinal data from cross-section surveys: an illustration from the Current Population Survey, Annals of Economic and Social Measurement, vol. 2, 1973, pp. 209-214.

[50] Kelly, W. H. Methods and Resources for the Construction and Maintenance of a Navaio Population Register, a report prepared for the National Cancer Institute by the Bureau of Ethnic Research, Department of Anthropology, University of Arizona, Tucson, 1964.

[51] Kennedy, J. M. Linkage of Birth and Marriage Record Using a Digital Computer, Atomic Energy of Canada, Ltd., Chalk River, Ontario, 1961.

[52] Kennedy, J. M. The use of a digital computer for record linkages, The Use of Vital and Health Statistics for Genetic and Radiation Studies, United Nations, New York, 1962, pp. 155-160.

[53] Kitagawa, E. M., and Hauser, P. M. Methods used in a current study of social and economic differences in mortality, Emerging Techniques in Population Research, Milbank Memorial Fund, New York, 1963, pp. 250-266.

[54] Kitagawa, E. M., and Hauser, P. M. Differential Mortality in the United States: a Study in Socioeconomic Epidemiology, Harvard Univ., Cambridge, 1973, pp. 183-227.

[55] Klebba, A. J., Maurer, J. D., and Glass, E. J. Mortality trends: age, color, and sex, U. S. 1950-69, U. S. Public Health Service, Vital and Health Statistics, series 20, no. 15, 1973.

[56] Livingston, R. Evaluation of the reporting of public assistance income in the special census of Dane County, Wisconsin (May 15, 1968), Proc. Ninth Workshop on Public Welfare Research and Statistics, New Orleans, 1969.

[57] Livingston, R. Evaluating the reporting of public assistance income in the 1966 Survey of Economic Opportunity, Proc. Tenth Workshop on Public Welfare Research and Statistics, 1970.

[58] Lloyd, J. W. Discussion of matching of census and vital records in social and health research: problems and results, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1965, p. 141.

[59] Loeb, J. Weight at birth and survival of the newborn by age of mother and total-birth order, U. S., early 1950, U. S. Public Health Service, Vital and Health Statistics, series 21, no. 5, 1965.

[60] Mandel, B. J., Wolkstein, I., and Delaney, M. M. Coordination of old-age and survivors insurance wage records and the post -enumeration survey, Studies in Income and Wealth: an Appraisal of the 1950 Census Income Data, Princeton, vol. 23, 1958, pp. 169-178.

[61] Marks, E. S., Mauldin, W. P., and Nisselson, H. The post-enumeration survey of the 1950 censuses: a case history in survey design, Jour. Amer. Stat. Assn., 1953, pp. 220-243.

[62] Marks, E. S., and Waksberg, J. Evaluation of coverage in the 1960 Census of Population through case-by-case checking, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1966, pp. 62-70.

[63] Masi, A. T., Sartwell, P. E., and Shulman, L. E. The use of record linkage to determine familial occurrence of disease from hospital records (Hashimoto's disease), Am. J. Pub. Health, vol. 54, 1964, pp. 1887-1894.

[64] McCarthy, M. A. Comparison of the classification of place of residence on death certificates and matching census records, U. S. Public Health Service, Vital and Health Statistics, series 2, no. 30, 1969.

[65] Meltzer, J. W., and Hochstim, J. R. Reliability and validity of survey data on physical health, Public Health Reports, vol. 85, 1970, pp. 1075-1086.

[66] Miller, H. P., and Paley, L. R. Income reported in the 1950 census and on income tax returns, Studies in Income and Wealth: an Appraisal of the 1950 Census Income Data, Princeton, vol. 23, 1958, pp. 179-201.

[67] Moriyama, I. M. Uses of vital recsrds for epidemiological research, J. Chron. Dis., vol. 17, 1964, pp. 889-897.

[68] Morrison, F. S., and Blen, A. L. Method of testing in the 1950 census the completeness of birth registration, Estadistica, vol. 7, 1949, pp. 185-193.

[69] Murray, J. H., and Haber, L. D. Methodology and validation, Appendix A of the Aged Population of the United States: the 1963 Social Security Survey of the Aged, by L. A. Epstein and J. H. Murray. SSA/ORS research report no. 19, 1967, pp. 193-219.

[70] Nathan, G. On Optimal Matching Processes, Case Institute of Technology, Cleveland, 1964.

[71] Nathan, G. Outcome probabilities for a record matching process with complete invariant information, J. Amer. Stat. Assn., vol. 62, 1967, pp. 454-469.

[72] Neter, J., Maynes, E. S., and Ramanathan, R. The effect by mismatching on the measurement of response errors, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1964, pp. 2-8.

[73] Neter, J., Maynes, E. S., and Ramanathan, R. The effect of mismatching on the measurement of response errors, J. Amer. Stat. Assn., vol. 60, 1965, pp. 1005-1027.

[74] Newcombe, H. B. Detection of genetic trends in public health, Effect of Radiation on Human Heredity, World Health Organization, Geneva, 1957, pp. 157-168.

[75] Newcombe, H. B. Environmental versus genetic interpretations of birth-order effects, Eugenics Quarterly, vol. 11, 1964, p. 36ff.

[76] Newcombe, H. B. Panel discussion, session on epidemiological studies, Second International Conference on Congenital Malformations, The International Medical Congress, Ltd., New York, 1964, pp. 345-349.

[77] Newcombe, H. B. Pedigrees for population studies, a progress report, Cold Spring Harbor Symposia on Quantitative Biology, vol. 29, 1964, p. 21ff.

[78] Newcombe, H. B. Population genetics, population records, Methodology in Human Genetics, Holden-Day, New York, 1962, pp. 92-113.

[79] Newcombe, H. B. Risk of fetal death to mothers of different ABO and RH blood type, Am. J. Human Genet., vol. 15, 1963, pp. 449-464.

[80] Newcombe, H. B. Screening for efects of maternal age and birth order in a register of handicapped children, Ann. Human Genet., vol. 27, 1964, pp. 367-382.

[81] Newcombe, H. B. Untapped knowledge of human populations, Transaction of the Royal Society of Canada, vol. 56, series 3, section 3, 1962, pp. 173-180.

[82] Newcombe, H. B., Axford, S. J., and James, A. P. A plan for the study of fertility of relatives of children suffering from hereditary and other defects, Atomic Energy of Canada, Ltd., report no. 511, Chalk River, Ontario, 1957, p. 50.

[83] Newcombe, H. B., James, A. P., and Axford, S. J. Family linkage of vital and health records, Atomic Energy of Canada, Ltd., report no. 470, Chalk River, Ontario, 1957.

[84] Newcombe, H. B., and Kennedy, J. M. Record linkage making maximum use of the discrimination power of identifying information, Communications of the ACM, vol. 5, 1962, pp. 563-566.

[85] Newcombe, H. B., Kennedy, J. M., Axford, S. J., and James, A. P. Automatic linkage of vital records, Science, vol. 130, 1959, pp. 954-959.

[86] Newcombe, H. B., and Rhynas, P. O. Child spacing following stillbirth and infant death, Eugenics Quarterly, vol. 9, 1962, pp. 25-35.

[87] Newcombe, H. B., and Rhynas, P. O. Family linkage of population records, The Use of Vital and Health Statistics for Genetic and Radiation Studies, United Nations, New York, 1962, pp. 135-154.

[88] Newcombe, H. B., and Tavendale, O. G. Effects of father's age on the risk of child handicap or death, Am. J. Human Genet., vol. 17, 1965, pp. 163-178.

[89] Newman, S. M. Problem in mechanizing the search in examining patent applications, Patent Office Research and Development Report No. 3, 1956.

[90] Nicholson, W. The income data series in the graduated work incentive experiment: an analysis of their differences, Chapter 6, Part C, The New Jersey Graduated Work Incentive Experiment, 1974.

[91] Nitzberg, D. M., and Sardy, H. The methodology of computer linkage of health and vital records, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1965, pp. 100-106.

[92] Ohlsson, I. Merging of data for statistical use, Bulletin of the International Statistical Institute, vol. 42, Proc. of 36th Session, Sydney, 1967, pp. 751-765.

[93] Ono, M., Patterson, G. F., and Weitzman, M. S. The quality of reporting social security numbers in two surveys, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1968, pp. 197-205.

[94] Perkins, W. M., and Jones, O. D. Matching for census coverage checks, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1965, pp. 122-139.

[95] Phillips, Jr., W., and Bahn, A. K. Experience with computer matching of names, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1963, pp. 26-29.

[96] Phillips, Jr., W., Bahn, A. K., and Miyasaki, M. Person-matching by electronic methods, Communications of the ACM, vol. 5, 1962, pp. 404-407.

[97] Phillips, Jr., W., Gorwitz, K., and Bahn, A. K. Electronic maintenance of case registers, Public Health Reports, vol. 77, 1962, pp. 503-510.

[98] Pollack, E. S. Use of census matching for study of psychiatric admission rates, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1965, pp. 107-115.

[99] President's Committee to Appraise Employment and Unemployment Statistics. Measuring Employment and Unemployment, 1962, pp. 386-394.

[100] Rogers, P. B., Council, C. R., and Abernathy, J. R. Testing death registration completeness in a group of premature infants, Public Health Reports, vol. 76, no. 8, 1961, pp. 717-724.

[101] Schachter, J. Matched record comparison of birth certificate and census information: United States, 1950, U. S. Public Health Service, Vital Statistics -- Special Reports, vol. 47, no. 12, 1962, pp. 365-399.

[102] Scheuren, F. J., Bridges, B., and Kilss, B. Report no. 1: subsampling the Current Population Survey: 1963 Pilot Link Study, Studies from Interagency Data Linkages, Social Security Administration, 1973.

[103] Scheuren, F. J., and West, G. 1966 and 1967 Survey of Economic Opportunity: Computer Consistency Checks, Office of Economic Opportunity, 1971, pp. 113-124 (Processed).

[104] Schneider, P., and Knott, J. Accuracy of census data as measured by the 1970 CPS-Census-IRS Matching Study, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1973, pp. 152-159.

[105] Shapiro, S. Estimating birth registration completeness, J. Amer. Stat. Assn., vol. 45, 1950, pp. 261-264.

[106] Shapiro, S. Recent testing of birth registration completeness in the United States, Population Studies, vol. 8, 1954, pp. 3-21.

[107] Shapiro, S., and Densen, P. Research needs for record matching, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1963, pp. 20-24.

[108] Shapiro, S., and Schachter, J. Birth registration completeness, 1950, Public Health Reports, vol. 67, 1952, pp. 513-524.

[109] Shapiro, S., and Schachter, J. Methodology and summary results of the 1950 birth registration test in the United States, Estadistica, vol. 10, no. 37, 1952, pp. 688-699.

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

[111] Siegel, J. S., and Zelnik, M. An evaluation of coverage in the 1960 census of population by techniques of demographic analysis and by composite methods, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1966, pp. 71-85.

[112] Silver, J. Discussion of matching of census and vital records in social and health research: problems and results, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1965, p. 140.

[113] Simpson, J. E., and Von Arsdol, Jr., M.D. The matching of census and probation department record systems, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1965, pp. 116-121.

[114] Sirken, M. G. Hospital utilization in the last year of life, U. S. Public Health Service, Vital and Health Statistics, series 2, no. 10, 1965.

[115] Sirken, M. G. Research uses of vital records in vital statistics surveys, Research Methods in Health Care, ed. by J. B. McKinlay, Prodist, New York, 1973, pp. 39-46.

[116] Sirken, M. G., Maynes, E. S., and Frechtling, J. A. The survey of consumer finances and the census quality check, Studies in Income and Wealth: an Appraisal of the 1950 Census Income Data, vol. 23, Princeton, 1958, pp. 127-168.

[117] Sirken, M. G., Pifer, J. W., and Brown, M. L. Design of Surveys Linked to Death Records, 1962.

[118] Steinberg, J. Some aspects of statistical data linkage for individuals, Data-Bases, Computers, and the Social Sciences, Wiley, New York, 1970, pp. 238-251.

[119] Steinberg, J. Some observations on linkage of survey and administrative record data, Studies from Interagency Data Linkages, Social Security Administration, 1973, pp. 1-14.

[120] Steinberg, J., Hearn, S., and Deutch, J. Social Security Administration's evaluation and measurement system, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1966, pp. 262-268.

[121] Steinberg, J., and Pritzker, L. Some experiences with and reflections on data linkage in the United States, Bulletin of the International Statistical Institute, vol. 42, 1967, pp. 786-805.

[122] Sunter, A. B. A statistical approach to record linkage, record linkage in medicine, Proc. of the International Symposium in Oxford, E. & S. Livingstone, Ltd., London, 1968.

[123] Tepping, B. J. Study of Matching Techniques for Subscriptions Fulfillment, Philadelphia, 1955.

[124] Tepping, B. J. Progress Report on the 1959 Matching Study, National Analysts Inc., Philadelphia, 1960.

[125] Tepping, B. J. A model for optimum linkage of record, J. Amer. Stat. Assn., vol. 63, 1968, pp. 1321-1332.

[126] Tepping, B. J. The application of a linkage model to the Chandrasekar-Deming technique for estimating vital events, Technical Notes 4, U. S. Bureau of the Census, Washington, D. C., 1971.

127] Tepping, B. J., and Bailar, B. A. Enumerator variance in the 1970 census, Proc. Amer. Stat. Assn. Soc. Stat. Sec., 1973, pp. 160-169.

[128] Tepping, B. J., and Chu, J. T. A Report on Matching Rules Applied to Reader's Digest Data, National Analysts, Inc., Philadelphia, 1958.

[129] Turner, Jr., M. L. A new technique measuring household change, Demography, vol. 4, 1967, pp. 341-351.

[130] U. S. Bureau of the Census. Infant enumeration study: 1950 completeness of enumeration of infants related to: residence, race, birth month, age and education of mother, occupation of father, Procedural studies of the 1950 Census, no. 1, 1953.

[131] U. S. Bureau of the Census The post-enumeration survey: 1950, Technical Paper, No. 4, 1960.

[132] U. S. Bureau of the Census. Evaluation and Research Program of the U. S. Censuses of Population and Housing, 1960: Record Check Studies of Population Coverage, series ER60, no. 2, 1964.

[133] U. S. Bureau of the Census. Evaluation and Research Program of the U. S. Censuses of Population and Housing, 1960: Accuracy of Data on Population Characteristics as Measured by CPS-Census Match, series ER60, no. 5, 1965.

[134] U. S. Bureau of the Census. Evaluation and Research Program of the U. S. Censuses of Population and Housing, 1960: the Employer Record Check, series ER60, no. 6, 1965.

[135] U. S. Bureau of the Census. Evaluation and Research Program of the U. S. Censuses on Population and Housing, 1960: Effects of Interviewers and Crew Leaders, series ER60, no. 7, 1968.

[136] U. S. Bureau of the Census. Evaluation and Research Program of the U. S. Censuses of Population and Housing, 1960: Record Check of Accuracy of Income Reporting, series ER60, no. 8, 1970.

[137] U. S. Bureau of the Census. 1970 Census of Population and Housing Evaluation and Research Program: Test of Birth Registration Completeness 1964 to 1968, PHC(E)-2, 1973.

[138] U. S. Bureau of the Census. 1970 Census of Population and Housing Evaluation and Research Program: the Coverage of Housing in the 1970 Census, PHC(E)-5, 1973.

[139] U. S. Bureau of the Census. 1970 Census of Population and Housing Evaluation and Research Program: the Medicare Record Check: an Evaluation of the Coverage of Persons 65 Years of Age and Over in the 1970 Census, PHC(E)-7, 1973.

[140] U. S. Bureau of the Census. 1970 Census of Population and Housing Evaluation and Research Program: the CPS-Census Match, PHC(E)-11, 1974.

[141] U. S. Public Health Service. Statistics of the United States, 1950, vol. 1, 1954, pp. 108-112.

[142] U. S. Public Health Service. Weight at birth and its effect on survival of the newborn in the United States, early 1950, Vital Statistics -- Special Reports, vol. 39, no. 1, 1954.

[143] U. S. Public Health Service. Reporting of hospitalization in the Health Interview Survey, Public Health Service Publication no. 584-D4. Vital and Health Statistics, series 2, no. 6, 1965.

[144] U. S. Public Health Service. Comparison of hospitalization reporting in three survey procedures, Vital and Health Statistics, series 2, No. 8, 1965.

[145] U. S. Public Health Service. Interview response on health insurance compared with insurance records, Vital and Health Statistics, series 2, no. 18, 1966.

[146] U. S. Public Health Service. A study of infant mortality from linked records, Vital and Health Statistics, series 20, nos. 12, 13, and 14, 1972-1973.

[147] U. S. Social Security Administration. Workers covered under Social Security: 1963 administrative information and pilot link results compared, Studies from Interagency Data Linkages, report no. 3 (In preparation).

[148] U. S. Social Security Administration. Individual income taxpayers: 1963 administrative information and pilot link results compared, Studies from Interagency Data Linkages, report no. 5 (In preparation).

[149] U. S. Social Security Administration. Report on Policies and Procedures for Establishing Initial Entitlement to RSDI Benefits, series 1, nos. 21-24, 1974.

[150] U. S. Social Security Administration. The 1% Sample Longitudinal Employee-Employer Data File, 1971 (Processed).

[151] Wells, B. Optimum Matching Rules, Univ. of North Carolina, 1974.