
Commentary: Interagency Data Matching Projects for Research Purposes

By Daniel B. Radner*

For many years the matching of income and other data from various data sources has been viewed as a very useful way of improving the quality and comprehensiveness of the information available for the analysis of government programs. Matching of various data files has been used to assess the accuracy of research data and to augment data files with a few selected items from another data file for limited purposes.

Matching has also been used to construct more comprehensive data bases intended for many different research purposes. These comprehensive data bases will be discussed here. The 1973 CPS-IRS-SSA Exact Match Study described in the article reprinted on the following pages is an important example of the construction of a comprehensive research data base at the Social Security Administration (SSA).

When attempting to analyze government programs, it often is found that any single household survey or administrative record system lacks important variables that are needed in the analysis. One solution to the problem is the carrying out of a household survey specifically designed for the analysis. But this solution is very expensive, it is impossible to obtain some crucial data items—

such as earnings histories—accurately in a household survey, and respondent burden is increased. An alternative solution is the matching of several data sets, usually from different government agencies, to provide a more comprehensive and more accurate data base for the analysis of programs. Such matches usually link information from household surveys with information from administrative record systems. Because of confidentiality and resource constraints, these matches often are joint projects between two or more agencies, although the relative roles of the different agencies can differ substantially.

Preserving the confidentiality of linked (and other) records has always been of paramount importance at SSA.¹ In each of the matching projects described here, SSA (and the other agencies involved) went to great lengths to preserve confidentiality. Detailed safeguard procedures were followed strictly. The reprinted historical article shows in detail the basic confidentiality requirements that applied to the 1973 Exact Match.

¹For example, see Joseph Steinberg and Heyman C. Cooper, "Social Security Statistical Data, Social Science Research, and Confidentiality," *Social Security Bulletin*, October 1967, pages 3-15 and Lois A. Alexander and Thomas B. Jabine, "Access to Social Security Microdata Files for Research and Statistical Purposes," *Social Security Bulletin*, August 1978, pages 3-17.

These requirements led to the establishment of such procedures in the 1973 Exact Match as strict limitations on access to data tapes and the use of tape abstracts that contained only the information needed for linking. Of course, any files released for public use contained no identifying information.

Three major interagency projects that emphasized the matching of data files to provide better data for the analysis of programs will be discussed here. The first was the 1963 Pilot Link Study, which was an important forerunner of the 1973 Exact Match. Three agencies—SSA, the Bureau of the Census, and the Internal Revenue Service (IRS)—were involved in this study. Data from the Census Bureau's Current Population Survey (CPS), IRS income tax returns, and SSA earnings and benefit records were linked by means of an exact match. In an **exact** match, information from different data files for the same person is merged, using a unique identifier (the Social Security Number, for example) to make the link. The primary purpose of the 1963 Pilot Link project was the construction of a more accurate data file for use in the analysis of income distribution issues related to SSA programs. This match was documented in

Continued on page 56

Table M-25.—SSI: Number of persons receiving State-administered supplementation, total amount, and average payment, by reason for eligibility and State, January 1988¹

State	Number				Total amount (in thousands)				Average payment			
	Total	Aged	Blind	Disabled	Total	Aged	Blind	Disabled	Total	Aged	Blind	Disabled
Total	² 269,935	110,330	3,105	146,071	² \$29,393	\$11,699	\$385	\$16,677	² \$108.89	\$106.04	\$124.12	\$114.17
Alabama	15,583	9,983	125	5,475	854	532	7	315	54.82	53.32	54.53	57.56
Alaska ³	4,726	1,744	62	2,920	1,081	397	14	670	228.71	227.52	233.65	229.31
Arizona	3,918	878	2	3,038	281	77	(4)	203	71.63	88.08	(5)	66.88
Colorado ³	19,623	14,555	92	4,976	2,011	1,655	4	352	102.48	113.69	42.57	70.80
Connecticut	17,874	6,869	113	10,892	4,569	1,594	25	2,950	255.62	232.09	217.50	270.86
Florida	10,399	4,981	(6)	⁷ 5,418	879	397	(6)	⁷ 482	84.50	79.72	(6)	⁷ 88.88
Idaho ³	2,985	970	21	1,994	350	96	2	253	117.40	99.06	78.24	126.73
Illinois	52,110	5,460	259	46,391	5,014	316	20	4,679	96.22	57.79	76.03	100.85
Indiana	782	372	6	404	357	133	3	221	456.04	358.49	452.83	545.91
Kentucky	6,835	3,461	89	3,285	891	447	6	438	130.33	129.08	68.79	133.32
Maryland	² 1,794	(6)	(6)	(6)	² 537	(6)	(6)	(6)	² 299.17	(6)	(6)	(6)
Minnesota ³	11,879	2,667	160	9,052	2,059	311	25	1,723	173.35	116.64	156.81	190.35
Missouri	8,416	6,365	338	1,713	328	197	63	68	38.95	30.89	186.51	39.78
Nebraska	7,055	2,161	98	4,796	414	94	6	324	58.62	38.74	56.53	67.62
New Hampshire	4,224	1,273	166	2,785	500	99	25	375	118.32	78.15	150.76	134.76
New Mexico	² 280	(6)	(6)	(6)	² 21	(6)	(6)	(6)	² 75.00	(6)	(6)	(6)
North Carolina	14,052	8,388	264	5,400	4,334	2,554	91	1,689	308.42	304.48	343.34	312.82
North Dakota	6	5	...	1	(4)	(4)	...	(4)	(5)	(5)	...	(5)
Oklahoma	55,367	32,039	544	22,784	2,744	1,495	30	1,219	49.57	46.67	54.61	53.52
Oregon	14,169	3,947	684	9,538	982	538	55	390	69.29	136.20	79.83	40.84
South Carolina	2,960	1,196	18	1,746	402	160	3	240	135.84	133.52	142.11	137.37
South Dakota	316	205	3	108	45	31	(4)	13	142.84	153.25	(5)	123.19
Utah	² 8,355	(6)	(6)	(6)	² 74	(6)	(6)	(6)	² 8.80	(6)	(6)	(6)
Virginia	5,274	2,695	36	2,543	649	584	8	57	123.06	216.57	226.97	22.49
Wyoming	953	116	25	812	19	2	1	16	20.00	20.00	20.00	20.00

¹ Data reported to the Social Security Administration by individual States. All data subject to revision. Excludes optional supplementation data for Missouri and North Dakota.

² Includes data not distributed by reason for eligibility.

³ Estimated data.

⁴ Less than \$500.

⁵ Not computed on base of less than \$500.

⁶ Data not available.

⁷ Includes data for the blind.

CONTACT: Joyce Jordan (301) 965-9852 for further information.

Commentary

Continued from page 22

several reports in the SSA series, **Studies from Interagency Data Linkages.**²

The 1973 Exact Match described in the reprinted article was a joint project between SSA and the Bureau of the Census. This match differed from the Pilot Link in several important ways: (1) the 1973 Exact Match

¹ Joseph Steinberg, **Some Observations on Linkage of Survey and Administrative Record Data**, Social Security Administration, Office of Research and Statistics, August 1973, and **Studies from Interagency Data Linkages** (Report Nos. 1-3, and 7), Office of Research and Statistics, Social Security Administration, various years.

sample was much larger, permitting more reliable and more detailed analyses; (2) computer matching was used much more extensively in the 1973 Exact Match, allowing a larger sample to be matched; (3) the 1973 Exact Match was documented in much more detail, permitting better assessments of the quality of the match; and (4) the release of public use files was emphasized more in the 1973 Exact Match, allowing the data to be used extensively by outside researchers (although a Pilot Link public use file was released). As shown in the

lengthy list of papers in the reprinted article, the data files from the 1973 Exact Match have been used by many researchers in the analysis of government programs and in assessments of the accuracy of data. The 1973 Exact Match was also documented in detail in the SSA **Studies from Interagency Data Linkages** series.³

The 1973 Exact Match file served as the starting point for

³ Social Security Administration, Office of Research and Statistics, **Studies from Interagency Data Linkages** (Report Nos. 4-6 and 8-11), various years.

Table M-26.—SSI: Number of persons receiving State-administered supplementation only, total amount, and average payment, by reason for eligibility and State, January 1988¹

State	Number				Total amount (in thousands)				Average payment			
	Total	Aged	Blind	Disabled	Total	Aged	Blind	Disabled	Total	Aged	Blind	Disabled
Total	² 72,676	27,739	455	42,688	² \$12,170	\$3,944	\$91	\$7,599	² \$167.46	\$142.18	\$199.31	\$178.01
Alabama	1,966	1,412	6	548	69	50	(3)	19	34.93	35.16	(4)	34.39
Alaska ⁵
Arizona	316	132	184	31	19	12	97.83	140.71	67.07
Colorado ⁶	5,959	5,538	5	416	915	826	1	88	153.53	149.09	245.00	211.47
Connecticut	15,852	5,824	92	9,936	4,068	1,336	20	2,712	256.60	229.39	217.79	272.91
Florida ⁷
Idaho ⁶	538	261	277	55	25	30	101.36	94.57	107.75
Illinois	25,761	1,193	56	24,512	3,683	116	6	3,562	142.99	97.02	109.36	145.30
Indiana ⁵
Kentucky	1,589	1,151	4	434	233	170	(3)	63	146.94	147.83	(4)	144.92
Maryland	² 1,794	(5)	(5)	(5)	² 537	(5)	(5)	(5)	² 299.17	(5)	(5)	(5)
Minnesota ⁶	2,397	643	20	1,734	703	153	8	542	293.24	237.58	396.25	312.70
Missouri	2,043	1,446	86	511	109	57	25	26	53.29	39.66	295.14	51.16
Nebraska	1,199	448	15	736	138	31	2	105	114.71	69.32	102.93	142.58
New Mexico ⁷
North Carolina	4,246	3,012	44	1,190	1,056	749	12	295	248.80	248.81	282.70	247.52
North Dakota ⁷
Oklahoma	6,702	5,085	18	1,599	223	167	1	55	33.23	32.90	41.72	34.18
Oregon	2,314	1,594	109	611	351	246	15	91	151.85	154.04	134.22	149.30
South Carolina ⁷
Utah ⁷
Wyoming ⁷

¹ Data reported to the Social Security Administration by individual States. All data subject to revision. Excludes data for optional and mandatory programs in New Hampshire, South Dakota, and Virginia; for optional programs in Missouri and North Dakota.

² Includes data not distributed by reason for eligibility.

³ Less than \$500.

⁴ Not computed on base of less than \$500.

⁵ Data not available.

⁶ Estimated data.

⁷ No persons receiving State supplementation only.

CONTACT: Joyce Jordan (301) 965-9852 for further information.

the third major SSA matching project—the Statistical Match project. In part, this was a joint project between SSA and the Bureau of Economic Analysis of the Department of Commerce. The Statistical Match project had as its goal the construction of a microdata file that would contain more accurate estimates of income, as well as information on tax liabilities. In the 1973 Exact Match, it had been possible to include only selected items from tax returns. In one step in the Statistical Match project, many more income tax return items were added to the Exact Match

file using a statistical match. In a **statistical match**, information is merged for similar persons in the files being matched, using similar characteristics to make the link. Statistical matching, which is less reliable than exact matching, was used because the files being matched were samples that in general did not contain the same persons. After the statistical match, the income amounts were adjusted for underreporting by making them consistent with independent income control totals from the National Income and Product Accounts and with other outside information. The results

of the Statistical Match project were described in an article in the **Social Security Bulletin** and the statistical matching steps were described in a detailed SSA report.⁴

⁴ Daniel B. Radner, "Distribution of Family Income: Improved Estimates," **Social Security Bulletin**, July 1982, pages 13-21, and **The Construction of the 1972 Statistical Match File: Part I—Statistical Matching Steps** (Report No. 13, Studies in Income Distribution), Office of Research and Statistics, Social Security Administration, September 1982.