

ACTUARIAL NOTE

Number 134
June 1996

SOCIAL SECURITY ADMINISTRATION
Office of the Actuary

SUITABILITY OF BENEFICIARY RECORDS FOR DETERMINING THE PROGRAM EXPERIENCE OF COUPLES

by Bertram M. Kestenbaum, A.S.A.

Introduction

From the perspective of program administration, the identity of a beneficiary's spouse, and even the fact of marital status, are not particularly relevant if neither person is an auxiliary beneficiary of the other; hence this information is not always routinely entered into the Social Security Administration's (SSA's) Master Beneficiary Record (MBR), the principal administrative file for the Old-Age, Survivors, and Disability Insurance program. Accordingly, the MBR would appear to be poorly suited to support investigations into the program experience of couples or the estimation of the effects of legislative proposals such as "earnings sharing," which would alter the structure of benefits to couples.

Indeed, such investigations and estimates are typically based on administrative data linked to household surveys—ongoing surveys such as the Survey of Income and Program Participation and the Current Population Survey, or special surveys such as the Longitudinal Retirement History Study and the Survey of the Aged. These linkages to surveys provide a wealth of information about beneficiaries' income, health, education, and even marital and childbearing history. Nevertheless, sample sizes are modest and the linkage process is protracted. Linked files can quickly become dated; as of this writing, for example, the most recent fully linked file is for May 1990.

The information necessary to associate the beneficiary records of husband and wife, though often absent from the MBR, is available in claims folders in SSA local offices. Indeed the December 1976 "Improved Family Benefits Data Project" (Lingg 1982) achieved the "coupling" of beneficiary records without resorting to a household survey, but rather by canvassing SSA offices across the country for claims folders for a sample of beneficiaries, and then launching a significant manual effort to examine and code paper documents. The drawbacks in this approach are obvious. The ideal approach would be to employ a method or combination of methods to couple beneficiary records using only the data elements available in machine-readable administrative files.

Actually, one means for identifying beneficiary couples in the MBR, though neither member is an auxiliary of the other, already exists. The cross-reference section of an MBR record is available for recording the Social Security number (SSN) of a spouse. In fact, our Office has used cross-reference data in estimating the effects of certain legislative proposals for altering the structure of couples' benefits. We know, however, that too often spouse information is not recorded in the cross-reference section.

This *Note* describes the implementation and results of a pilot study investigating the efficacy of certain unconventional methods for coupling MBRs. These new methods can be used with large samples, are relatively quick, and miss only a small fraction of couples. However, steps must be taken to limit the incidence of incorrect couplings.

One new method, which is our primary focus, is surname/address matching. Our assumption is that a male retiree and a female retiree with the same surname and address are likely to be husband and wife. Of course we are aware of the possibilities that they may be parent and child or brother and sister, and that they may be unrelated persons living in either a multi-unit structure or group quarters.

Another new method is commonness of a bank account number. This method takes advantage of the fact that most beneficiaries use direct deposit, and assumes that a male retiree and a female retiree whose benefit checks are direct-deposited to the same bank account are likely to be husband and wife. Here, too, there is the possibility of a different familial relationship.

The pilot study

The nature of an investigation of the commonness of surname and address or of a bank account number ruled out the sample design used most often in our work—selection for the sample by patterns in the Social Security number. Instead, we used an area sample and selection with probability proportional to size, to select 25 zip-code areas. The 25 zip-code areas

selected are listed in the Appendix, together with a measure of size.

We obtained MBR records for all primary beneficiaries with benefits in force in these 25 areas as of mid-March 1994. Because address information is maintained in the MBR for about 4 years after the death of a beneficiary or the termination of a benefit for any other reason, we also included records of primaries who were terminated after January 1990. Additionally, we included records of lump-sum payments made to widow(er)s after that date.

Among the records, 22,091 were for females with retirement benefits in force. This represents about a 1-in-575 sample of the approximately 12.6 million in-force female primary retirement beneficiaries at the time. Our focus in the pilot study is on uncovering the identity of spouses of these 22,091 females.

A total of 8,690, or 39 percent, of these in-force female retirees are now, or have been, dually entitled as an auxiliary of their husband in addition to having primary entitlement. In an additional 75 cases, the husband is or was dually entitled, with auxiliary entitlement on the wife's record.

Cross-reference data in records of female primaries permit us to make an additional 4,062 couplings. In most of these cases the female has never been the male's auxiliary. In some cases, however, the female was his auxiliary at some time but is no longer; for example, she may have been a widow beneficiary at ages 60-61 and then became entitled to a (larger) primary benefit at age 62. Cross-reference data in the records of male primaries identify 750 more couplings.

In total, the conventional methods of identifying husbands—using the dual entitlement and the cross-reference sections of MBR records—yield 13,577 successes. Many of the identified husbands had died more than 4 years ago, and some of those alive or recently deceased did not reside in the same zip-code area as their spouse. (Possible reasons for zip-code differences include divorce, institutionalization of one member of the couple, and the relocation of a widow after her husband's death.) The number of husbands identified by conventional methods who furthermore had an MBR record in the sample of 25 areas is 6,990, of whom 5,800 were in-force and 1,190 died in the past 4 years.

New coupling methods

The two new coupling methods are matching on surname/address and matching on bank account number. For records of lump-sum payments to widows, where the payee is named, matching is on given name, surname, and address.

For name matching and bank account number matching we required an exact match. We recognized, however, that variations in address occur so frequently that the same requirement for address matching is too restrictive. (For example, there are quite a few variations on a street name like 'General George Patton!')

The Geography Division of the Bureau of the Census kindly shared with us an address standardizer program, which divides an address into its component parts and applies standardized abbreviations to street names (Bureau of the Census 1994). Processing addresses through this standardizer program, with some modifications, and requiring a match on only selected address components, substantially enhanced our address matching capability.

Additionally, we take advantage of the fact that more than 90 percent of addresses in the MBR are coded to the "ZIP+4" level—smaller geographic units corresponding to the area of a mail delivery route—and recognize as matches agreement on surname and ZIP+4, although this method is more prone to incorrect matching.

Matches in which the female retiree was more than 15 years older than the male retiree were excluded because of the significant possibility that the matching male was a son, not a spouse. We did not, however, exclude similar cases of a female retiree much younger than the male, because a woman and her father would not have a surname in common unless she had not married.

In the event that a retiree is incapable of managing his or her finances and an individual or organization is selected to be a representative payee, the MBR address would be that of the payee. Hence, several unrelated beneficiaries may have the same MBR address because they have the same payee, and may easily have the same surname, as well, purely by chance, if the payee is a large organization serving many beneficiaries. Indeed, we excluded any matches based on the address of the San Francisco Public Guardian.

The same potential for incorrect matches of unrelated persons exists for very large group quarters such as large nursing homes and for houses with many apartments. However, the two large group quarters in the 25-zip-code area presented no problem. One was a home for men only and the other, the Association of the Sisters of Mercy, had only female members. We also excluded bank account number matches when there was an indication on either record of representative payment to someone other than a spouse.

The new methods combined produced 8,325 pairings of female primaries: 7,146 to in-force husbands and

1,179 to husbands deceased in the past 4 years. This compares favorably with the 6,990 successes of the conventional methods in the 25-zip-code area. The performance of each of the new methods is as follows:

Method	Number of matches
Surname and exact address	5,225
Surname and standardized address	7,520
Surname and ZIP+4	7,562
Full name and address: lump-sum cases	233
Bank account number	3,493

The conventional methods and new methods combined identify 8,715 pairings, including 390 identified only by the conventional methods and 1,725—1,582 spouses and 143 widows—identified only by the new methods. An estimate of the number of couples missed by both sets of methods is possible under the assumption that the two sets of methods are independent of each other. (Let X be the number missed; then $X / (8715 + X) = (390 / 6990) * (1725 / 8325)$.) That estimate is: only 102 couples.

Surname/address matching fails for non-coresidents. Accordingly, we find that surname/address matching is less successful for widows than for married women, and less successful for women in the fourth year of widowhood than for women in the first year of widowhood, as widows tend to relocate from the home they shared with the decedent. Some of the couples missed by surname/address matching techniques do share a residence but have inconsistencies in their recorded addresses too pronounced to be resolved by the standardizer program; some additional editing or Soundex (phonetic) coding would be worthwhile.

Though impressive in terms of power, surname/address matching techniques suffer from a susceptibility to incorrect matching. In particular, in a random sample of 80 pairings produced by the new methods but not the conventional ones, we found, by checking in the NUMIDENT file, that 9 were not husband-wife pairs but instead either parent-child or brother-sister pairs.

The NUMIDENT file contains applications for an SSN card, a replacement for a lost card, or a revised card due to a change in information such as a change in surname upon marriage. (The NUMIDENT also contains information for decedents on fact and date of death; for persons issued more than one SSN, on the several numbers they hold; and for claimants to program benefits up through the mid-1970's, on the nature and date of the claim.) Most of the NUMIDENT is in machine-readable form; however, for accounts on which claims were filed up through the

mid-1970's, the original application exists only on microfilm (Social Security Administration n.d.).

The application form for an original, replacement, or corrected Social Security card asks for, in addition to the applicant's current name, the applicant's name at birth, and the names of the applicant's parents. The maiden name on the application is the key to determining the relationship of the two members of a pair: it is almost certainly not a spousal relationship if the common surname is the female's maiden name.

From this sample of 80 we infer that about 200 of the 1,725 pairings accomplished by the new methods and not the conventional ones are spurious.

Conclusion

The new methods applied in this pilot study are powerful means for pairing with a wife retiree the husband retiree who is alive or recently deceased. The address standardizer software substantially enhances the matching capabilities, and further expenditures of effort to edit addresses would be profitable.

However, the problem of spurious matches needs to be addressed, especially since, generally speaking, the negative impact of a spurious match is greater than that of a missed correct match. Although some incorrect matches can be identified through internal checks, for example, if one female matches to two males, or if the Primary Insurance Amount (PIA) of a female not dually entitled is less than half of the matched male's PIA, a systematic weeding out of the false positives requires an interface with the NUMIDENT.

Thus, by employing the MBR either alone, or, better, in conjunction with the NUMIDENT, using an area-sample design and a combination of the conventional and new methods discussed here, investigations of the program experience of couples can be performed in a productive and timely fashion.

REFERENCES

Bureau of the Census. 1994. *Address Standardizer Software release 5.5*.

Lingg, Barbara. 1982. "Social Security Benefits of Female Retired Workers and Two-Worker Couples." *Social Security Bulletin* Vol. 45, No. 2, pp. 3-24.

Social Security Administration. n.d. *Users' Manual: Studying Employment and Program Experience of Immigrants and Hispanics with Social Security Administrative Records*.

Appendix: The 25 Zip Codes

Zip	Location	Number of female retirees in-force, March 1994
01609	Worcester, MA	991
04256	Mechanic Falls, ME	228
18337	Milford, PA	563
25704	Huntington, WV	967
29483	Summerville, SC	1,168
32413	Panama City, FL	442
33139	Miami Beach, FL	4,110
33455	Hobe Sound, FL	1,697
37221	Nashville, TN	900
45869	New Bremen, OH	177
47952	Kingman, IN	117
50125	Indianola, IA	898
53590	Sun Prairie, WI	843
58341	Harvey, ND	190
60651	Chicago, IL	1,265
66101	Kansas City, KS	1,075
67103	Mayfield, KS	9
72205	Little Rock, AR	2,044
78660	Pflugerville, TX	203
79035	Friona, TX	162
79905	El Paso, TX	1,109
81401	Montrose, CO	1,122
93210	Coalinga, CA	328
94102	San Francisco, CA	778
95814	Sacramento, CA	705
	Total	22,091