PROPERTIES OF THE SOCIAL SECURITY NUMBER RELEVANT TO ITS USE IN RECORD LINKAGES

Thomas B. Jabine, Consultant, Committee on National Statistics

Linkage of records from two data systems is aided greatly by the presence in both systems of the same numeric identifier, for example, the social security number (SSN) for persons or the employer identification number (EIN) for businesses. When matching variables for two records are compared, agreement on such numeric identifiers is usually given a large weight in deciding whether a true match exists.

Because of their importance for record linkage, it is important to have complete and current information on the relevant properties of each of these numeric identifiers. Such properties include: coverage, general structure and method of issuance, information content, and appropriate methods of validation. Properties relevant to sample selection using numeric identifiers are also of interest, since many record-linkage studies are based on a sample from one of the data systems.

This paper provides a description of the properties of the social security number (SSN) that are relevant to its use in record linkages. The description should be regarded as a first draft and readers are urged to suggest corrections and additions.

If this description of the SSN proves useful, it is suggested that the Administrative Records Subcommittee of the Federal Committee on Statistical Methodology make arrangements to: (1) prepare and disseminate descriptions, using the same format, of other commonly used numeric identifiers, such as the EIN and the unemployment insurance number, and (2) update the descriptions periodically and whenever significant changes occur.

Special thanks are due to Richard Wehrly of the Social Security Administration for providing information used in developing the SSN description. However, any errors are the sole responsibility of the author and readers are cautioned that the description of the SSN has not been officially reviewed by the Social Security Administration.

NUMERIC IDENTIFIER DESCRIPTION

1. Name of identifier
   The social security number (SSN).
2. Administrative uses
   SSNs were issued initially so that earnings of persons in jobs covered by the social security retirement program could be reported, by their employers, to the Social Security Administration (SSA) and credited to the persons accounts for subsequent use in determining benefit eligibility and payment amounts.

An early decision was made to use SSNs as identifiers in the State-operated unemployment insurance programs. No other significant uses developed until 1961 when the Internal Revenue Service, after discussions with SSA, decided to use the SSN as a taxpayer identification number. After implementation of this decision, other

uses by Federal and State governments followed rapidly, and the SSN is now widely used as an identifier for workers, taxpayers, drivers, students, welfare beneficiaries, civil servants, servicemen, veterans, pensioners and others (HEW Secretary's Advisory Committee, 1973).

Legal justification for use of the SSN as an identifier by Federal agencies comes from Executive Order 9397, issued in 1943, which directed Federal agencies to use the SSN when establishing a new system of permanent account numbers. The Privacy Act of 1974 placed some restrictions on use of SSNs by Federal, State and local government agencies, but uses formally established prior to January 1, 1975 were not affected and these restrictions have had only a minor effect on widespread administrative use of the SSN by governments and private organizations (Privacy Protection Study Commission, 1977).

3. Coverage
   a. Units.--SSNs are issued to persons.
   b. Legal coverage provisions.--An SSN will be issued to any United States citizen upon application and presentation of acceptable evidence of identity. Foreign nationals legally present in the United States will be issued SSNs if legally entitled to work or if they have an acceptable "nonwork reason" for needing an SSN, e.g., the need for a taxpayer identification number.

All persons with Federally taxable income and their spouses are required to obtain SSNs for use as taxpayer identification numbers. SSNs are also required for many types of benefits and for other purposes: social security, driver's license, welfare benefits, voter registration, participation in scholastic aptitude testing programs, etc. For some of these, requirements vary by State.

   c. Volume and characteristics of issuance to date.--SSNs were first issued in November 1936. By the end of 1975, over 235 million SSNs had been issued and there were an estimated 180 million living SSN holders (Social Security Administration, 1981b). As of the close of 1983, approximately 287,083,000 SSNs had been issued. It is estimated by SSA that there were 204,760,000 living SSN holders at the end of 1981. When SSN holders die, their SSNs are not reissued to other applicants.

The table in Attachment A shows the number of SSNs issued annually, by sex of applicant, through the end of 1979. Following the large number of issuances in the first 14 months (November 1936 to December 1937), the volume of annual issuances has fluctuated for a variety of reasons, with a tendency to increase in recent years as coverage of SSA benefit programs and the use of SSNs for non-SSA programs has expanded. Today most of the SSNs are issued to applicants under 20 years of age. In 1979, 62.8 percent of the SSNs were issued to persons under 15 and another 26.2 percent to

persons between 15 and 19 (Social Security Administration, 1981b).

From time to time, surname counts based on the first six characters of the surname are made from SSA's account number files. Kilss and Tyler (1974) show the rankings of common surnames based on 1964 counts. Based on a 1974 tabulation, the ten most common surnames were:

       Smith
       Johnso(n)
       Willia(ms)(mson)
       Brown
       Jones
       Miller
       Davis
       Martin(ez)(son)
       Anders(on)
       Wilson

The letters in parentheses following some names are intended to show the more common surnames that have these first six characters.

d. Uniqueness, stability.--Until 1972, applicants for SSNs were not asked if they had already been issued numbers, nor were they asked for proof of identity. As a result many persons now have more than one SSN (Privacy Protection Study Commission, 1977). As of 1973, it was estimated that 4.2 million persons had two or more SSNs (HEW Secretary's Advisory Committee, 1973). More recent estimates are not available. Today, intentional issuance of multiple numbers to the same person is permitted only in exceptional circumstances, generally involving national security or the protection of the person in question.

In most cases where a person is known to have more than one SSN, SSA's computerized SSN files contain a record for each of his or her SSNs and cross references linking all of the SSNs.

Sometimes more than one person uses the same SSN. Some reasons why this happens are discussed in item 8b. Estimates of the frequency with which this occurs are not readily available, but it is believed to be much less prevalent than issuance of multiple numbers to the same person (HEW Secretary's Advisory Committee, 1973).

4. General structure and information content

The social security number has nine digits arranged as follows: 000-00-0000. The first three digits are called the area number, the next two are the group number, and the last four are the serial number. There are no check digits. The serial number provides no information about the person to whom an SSN has been assigned; however, the area and group numbers do contain a limited amount of information.

The area number, digits one to three of the SSN, carries some information either about the SSN holder's occupation or his or her place of residence at the time the number was issued. For the ranges of area numbers used to date, the information content is as follows:

(1) Area numbers 001 to 626. With a few exceptions, each of these area numbers has been assigned to a single State, one or more to a State. For most SSNs, the area number indicates only the SSN holder's State of residence at the time of issuance, as derived from the mailing address on the SSN application. For SSNs issued in the early days of social security, the area number indicated the specific SSA field office from which the number was issued, regardless of where the applicant lived.

(2) Area numbers 700-728. These numbers were assigned to railroad workers through 1963. Since then, railroad workers have been assigned SSNs with the same area numbers as other applicants.

The group number, digits four and five, in combination with the area number, provides a rough indication of when the SSN was issued. In particular, it is possible to tell whether an SSN was issued before or after another SSN having the same area number but a different group. Within an area number, the group numbers are always used in the following sequence:

  - Odd numbers from 01 to 09
  - Even numbers from 10 to 98
  - Even numbers from 02 to 08
  - Odd numbers from 11 to 99

The group number 00 has never been used. Only the first two sets of group numbers in the above sequence were used through 1965. Since then the third and fourth sets have been used with some area numbers. Current information on the last group number assigned for each area number can be obtained from SSA (see Section 9.a.).

5. Issuance procedures

All SSNs are issued by the Social Security Administration. Prior to July 1, 1963, the Railroad Retirement Board issued SSNs (in the 700 series) to all railroad employees.

A single application form, Form SS-5, Application for a Social Security Number Card, is used for initial applications, requests for replacements for lost cards and corrections, such as name changes. A copy of the application form is shown in Attachment B. Applications must be accompanied by evidence of age, identity and U.S. citizenship or lawful alien status. They may be submitted either in person or by mail, except that aliens and persons 18 or older making initial applications must apply in person.

Most SSN applications are submitted to SSA field offices. In 37 States, applications for new welfare applicants needing SSNs are developed by the State welfare agencies and submitted by the State directly to SSA's Office of Central Records Operations. SSA district offices sometimes make arrangements with schools for "mass enumerations" in which SSA and school officials collaborate in obtaining and reviewing applications from all students who wish to obtain SSNs.

The application forms (SS-5) and accompanying evidence submitted to district offices are screened for completeness and accuracy by district office personnel, who make further contacts with applicants when necessary. The SS-5 information is then keyed in the district office for direct transmission to SSA central operations.

The central processing of the applications consists of validation (which is essentially a matching operation) against existing SSN files, followed by appropriate actions. The exact

nature of the validation depends on the type of application. For example, if an initial applicant alleges that he or she has not been issued an SSN previously, the purpose of the validation is to confirm that allegation. Validation procedures are discussed further in item 9b.

The final step depends on the results of the validation. The main possibilities are: assigning an SSN and mailing a card to a new applicant, mailing a replacement card to an applicant, correcting information (such as name) about the applicant in the SSN computerized files, or asking the field office to supply additional information.

When a new SSN is assigned, the next available number for the State from which the application was submitted is used. The sequence of availability proceeds from the lowest area number used in a given State through the highest area number for that State, using the same group number. For example, in New Hampshire, which has been assigned area codes 001, 002, and 003, the last available number in group 001-52 would be followed by the first available number in group 002-52, and the last available number in that group would be followed by the first available number in group 003-52.

6. Sampling properties

In theory, a probability sample could be selected using digital patterns based on any of the nine digits of the SSN or combinations thereof. However, consideration of the information content of the first five digits, as described in item 4, makes it clear that use of any of those digits should be avoided. It would be most inconvenient to select a sample that turned out to include only persons who were railroad workers at the time their SSNs were issued and had all been issued their SSNs not later than 1963!

The serial number part of the SSN, however, does not have this kind of problem and consequently is frequently used for digital sampling from a file of records that includes SSNs. Assuming a uniform distribution of 9,999 possible serial numbers (SSNs ending in 0000 have never been issued), it is possible to choose a digital sampling pattern that will approximate any desired sampling fraction. There are usually several alternatives. For example, to select a sample of approximately 5 percent (1 in 20) of the records, one could use

(1) 5 of the 100 possible combinations of the 8th and 9th digits;

(2) 50 of the 1,000 possible combinations of digits 7, 8 and 9;

(3) 500 of the 9,999 combinations of digits 6, 7, 8 and 9;

(4) 5 of the 100 possible combinations of the 7th and 8th digits

and so forth. The combinations of digits selected may be chosen at random with or without replacement (the latter would be preferable) or systematically with a random start. In the latter case, for exmple, we might choose the pair 73 at random and include with it the pairs 93, 13, 33 and 53.

The use of selected digits or combinations of digits for sampling is actually a form of cluster sampling. In the illustration used above, we could describe a population of records as consisting of 100 clusters, each consisting of all records with SSNs having a particular pair of 8th and 9th digits. Five of these clusters are selected by an appropriate probability sampling mechanism.

In practice, samples of this kind, especially when only the 8th and 9th digits are used, behave pretty much like random samples, chosen without replacement. In particular, reasonably accurate estimates of sampling error can be calculated as though the data were from a simple random sample.

In selecting samples based on the serial number portion of the SSN, the following points should be considered:

(1) The serial number 0000 is not used. The effect of this, which is quite small, on the expected sample size can easily be calculated.

(2) The digital patterns used for any particular sample determine only the _expected_ sampling fraction or size. The sample size _realized_ by using a particular set of digits or combination of digits will, in general, differ somewhat from its expected value. If precise control of sample size is important, this can be achieved by oversampling initially and then subsampling units at random or systematically from the initial sample.

(3) As discussed in item 3d, some persons have been issued more than one SSN. Such persons may have multiple chances of selection in a sample of persons obtained by selecting SSNs, depending on what record sets are being used. If the number of SSNs that each sample person has can be determined, appropriate adjustments can be made in estimates based on the sample. Because the phenomenon is infrequent, however, it is usually ignored in practice.

(4) Various studies (Hawkes and Harris, 1969; Page and Wright, 1979) have shown that the distributions of SSNs by ending digit in selected record sets is essentially uniform. However, studies conducted with various record sets in the late 1960s and early 1970s (Hawkes and Harris, 1969; Internal Revenue Service, 1973) showed a negative linear relationship between the ascending sequence of digits in positions 6 and 7 and the number of SSNs in these record sets having those digits. This probably resulted from the fact that, until 1972, SSNs in each area-group combination were issued consecutively by serial number, from 0001 to 9999. Since then, they have been issued in a randomized order, largely to avoid issuing consecutive numbers to persons with the same surname. Because of the new issuance procedure, one would expect this relationship to disappear gradually. However, to be on the safe side, it is recommended that: (1) digital sampling patterns use only the 8th and 9th digits whenever requirements can be met in that way, and (2) whenever multiple combinations of two or more digits are used, they should be selected systematically rather than at random from the range of possible combinations.

7. Links with other numeric identifiers

At the Federal level, there are two kinds of links between SSNs and employer identifica-

tion numbers (EINs). For employees, the link occurs in the W-2/W-3 annual wage and tax reporting system (prior to 1978, reporting was quarterly). For many years SSA has used this link for statistical purposes, in the Continuous Work History Sample system, to add employer locations and industry data to records of earnings and demographic characteristics for sample persons. More recently, the Statistics of Income Division of IRS has used the same link to obtain employer industry codes to use as an aid in coding occupations reported by individual taxpayers on their returns.

The second link between SSNs and EINs applies to persons who operate businesses as sole proprietors. This link applies primarily to sole proprietors with employees; those with no employees are not, in general, required to obtain and use EINs. The link occurs in two ways: on income tax returns of sole proprietors, and on new applications for EINs. On income tax returns, the business schedules (C and F) call for entries of both the EIN (if the taxpayer has one) and the SSN. On EIN application forms (Form SS-4), applicants who are sole proprietors are asked to enter their SSNs.

There are undoubtedly several links between the SSN and other numeric identifiers at the State and local levels. One obvious one is the link between SSNs and employer unemployment insurance (UI) identification numbers, which is necessary for the operation of the UI program. The precise nature of the linkage varies by State and, for the minority of States which operate under the "wage request" system, it may not exist in any readily accessible sense.

8. Reporting formats and problems

a. Formats.--Many different administrative and statistical forms include spaces for recording SSNs, either by the holders or by someone else completing the form. There is no standard format for this purpose. The particular format used may have some effect on the accuracy with which SSNs are entered on the forms and read from the forms for purposes of manual transcription or data entry.

Format features that vary include: width and height of the space provided for the number; separators used for the area, group, and serial numbers; use of boxes for individual digits; and the label used to indicate what should be entered. Some examples of these features appear below. All of them show the actual size of the entry space on the form.

Example 1. Department of State, Passport Application, Form SDP-11 (7-79)

BIRTHPLACE (City, State or Province, Country)		BIRTH DATE Month Day Year
URE DATE	HEIGHT ___ Ft ___ In.	COLOR OF HAIR	COLOR OF EYES
IENT RESIDENCE (Street address, City, State, ZIP Code)		SOCIAL SECURITY NO. (Not mandatory)
	FATHER'S NAME	BIRTH

Of several formats examined, this one provided the narrowest space for entering the

SSN, with a width of 1 1/4 inches. Most others were in the range of 1 1/2 to 2 inches.

Example 2. Internal Revenue Service, Employee's Withholding Allowance Certificate, Form W-4 (10-79)

f the Treasury—Internal Revenue Service

:hholding Allowance Certificate

Your social security number ▶	i	i

old at higher Single rate
se is a nonresident alien, check the single block.

This format allowed the smallest vertical distance of those examined, 5/32 inch. It uses vertical dotted lines as separators for the three parts of the SSN.

Example 3. Internal Revenue Service, Application for Employer Identification Number, Form SS-4, (8-76).

n page 4)

im 1)	3 Social security number, if sole proprietor
	5 Ending month of accounting year

This format also uses the dotted vertical lines as separators. In this case, the spaces for the three portions of the SSN are all the same length, 5/8 inch. Other forms using separators make the lengths of the three spaces roughly proportional to the number of digits to be entered, i.e., 3, 2, and 4.

Example 4. Bureau of the Census/Department of Health and Human Services, Income Survey Development Program, 1978 Research Panel-July Questionnaire, Form ISDP-403.

Last		
First	Middle	(0048)
Social Security Number		
(0046) [ ][ ][ ] - [ ][ ] (0047) [ ][ ][ ][ ]		[
Last		

This format illustrates the use of separate boxes for each digit of the SSN. The three parts of the SSN are separated by horizontal dashes. The circled numbers are source codes for data entry.

Example 5. Social Security Number Card (Original, Replacement or Correction), Form SS-5 (5-84) (see Attachment B).

This item is completed only for persons who already have SSNs and are applying for a replacement or correction. This format uses a box for each digit, with intervening spaces, and horizontal dashes to separate the three parts of the SSN. The wording of the item label reflects the fact that the form is

sometimes completed by someone other than the "applicant."

Example 6. Internal Revenue Service, Form 1040 EZ Income Tax Return for Single Filers with no Dependents.

Please print your numbers like this.

1234567890

Social security number

This format is used for handwritten entries by taxpayers that will be read automatically by optical character reading equipment. On the actual form, the boxes for the individual digits are in light blue. The boxes for the area, group and serial parts of the SSN are separated.

Example 4 above comes from a questionnaire that is completed by trained Census Bureau interviewers. The other examples are all from forms that are filled by members of the general public. No experimental research on alternative formats for recording SSNs has been identified. Some other research has suggested that the use of individual character separators may actually reduce legibility of entries (Wright, 1980).

b. Reporting and processing errors.--Most errors in SSNs in data files occur for two reasons: (1) the person completing the form or answering the questions gave an SSN for the wrong person, or (2) the SSN is for the right person, but it was reported, recorded, transcribed or keyed incorrectly.

The first type of error can occur, for example, when a widow reports the number under which she is receiving benefits, rather than her own. Another example is what SSA calls the "pocketbook number." The number 078-05-1120 appeared on a sample account number card contained in wallets sold nationwide in 1938. Several thousand people mistakenly reported this number to their employers as their own! By the 1970s there were over 20 different pocketbook numbers (HEW Secretary's Advisory Committee, 1973, p. 112).

People who lose their social security cards can apply for replacement cards bearing the SSN already issued to them. In cases where they are not able to give their SSN on the application, SSA must determine the correct SSN based on other identifying information. Occasionally a mismatch occurs and the person will be issued a replacement card bearing someone else's SSN.

The second type of error is usually an error in a single digit or a transposition of digits, types of errors that could be easily corrected if a check digit were used.

Cobleigh and Alvey (1974) describe errors detected when SSNs reported in the Current Population Survey were validated against Social Security Administration files. About three percent of the reported SSNs were clearly in error. Roughly two-thirds of these were found to have transposition or single-digit errors. Another one-sixth were SSNs belonging to other members of the same household, and the remainder could not be located in SSA's files.

9. Validation procedures

a. Intra-record validation.--When undertaking record linkages based on SSNs, it is usually desirable to start by identifying SSNs that are clearly invalid. A first step might be to look at the SSN itself and determine whether it is within the range of numbers issued to date. SSA will make available, on request, up-do-date information on the area numbers that have been issued so far and, for each of those numbers, the "highest" group number issued. "Highest" must be interpreted in terms of the standard sequence for use of group numbers within an area number, as explained in item 4 above.

Attachment C provides this information as of January 2, 1985. As of that date, the only area numbers used were those in the ranges 001 to 587, 589 to 595, 600 and 601, and 700 to 728. Also, group number 00 and serial number 0000 are never used. Current information on highest group numbers may be obtained from the director of the OASDI Statistics Division; Office of Research, Statistics and International Policy; Social Security Administration.

If records to be linked have information on date of birth or age, the SSN can be checked for consistency with age. The operating rule is that a person whose SSN was issued x years ago must be at least x years old. Since virtually all numbers issued through 1961 were issued to employed persons, only a few errors would be made by requiring that persons with numbers issued in this period be at least x + 15 years old. For SSNs issued from 1951 onwards, the SSA can provide fairly precise information about the years in which numbers with specific area-group combinations were issued (contact the source given in the preceding paragraph). For numbers issued prior to 1951, only rough estimates of issuance periods for area-group combinations are possible.

b. Validation against SSA records.--Validation is defined broadly here as a process in which SSN information for individuals from sources external to SSA records is checked against those records to determine its validity. Specifically, if the external record includes an SSN, it is desired to know whether the SSN is the correct one for that person and, if it is not correct, what the correct SSN, if any, is for that person. If the external record for a person has no SSN, it is desired to know whether that person has an SSN and, if so, what it is. This kind of validation requires matching external records to SSA records and should be thought of in that context.

Validation of SSN information is done routinely by SSA for program purposes. Somewhat less frequently it is undertaken for statistical purposes. Some examples of the latter are:

(1) Validation of SSNs collected in pretests for the 1970 Census of Population (Ono et al., 1968).

(2) Validation of SSNs collected in the March 1973 Current Population Survey, as a preparatory step before adding SSA and IRS administrative data to the survey records (covered in several reports and articles, e.g., Cobleigh and Alvey, 1974; Social Security Administration, 1981a).

(3) Validation of SSNs collected in panel surveys as part of the Income Survey Development Program (Kasprzyk, 1983).

(4) In various mortality followup studies, as a preparatory step before determining which members of an externally identified study population have died, according to SSA records.

Attachment D provides a summary description of SSA's current validation procedures for program operations. A combination of computerized and manual procedures is used, and unresolved cases are returned to district offices with an instruction to seek additional information from the applicant or claimant. The SSN files maintained by SSA are now fully computerized and a more sophisticated computer validation system is being developed.

A variety of validation procedures have been used in statistical applications; some of them are described in the references cited above.

The circumstances under which SSA will validate SSN information for administrative or statistical purposes are limited by law and by SSA regulations and policies. Anyone wishing to validate SSN information for statistical or research purposes should contact SSA's Office of Research, Statistics and International Policy.

10. Use as a matching variable

Arellano (n.d.) discusses use of the SSN in record linkages based on the model proposed by Fellegi and Sunter (1969). He recommends that the SSN not be used for blocking, because of the possibility that some individuals in the files to be linked may not have been issued SSNs. To use the SSN as a component of the comparison vector, Arellano recommends that the 9 digits of the SSN be partitioned into four elements on a 2,2,2,3 basis. He identifies 17 possible configurations of the SSN component of the comparison vector, covering the possible realizations of agreements and disagreements in the four elements, plus the case in which no SSN is available for one or both members of the comparison pair. He then suggests procedures for assigning conditional probabilities to these configurations for the matched and unmatched sets. These probabilities are based on assumptions about the kinds of errors that can occur in the matched set and on observed frequencies of realizations of the first three elements of the partitioned SSNs in the files to be linked (realizations of the fourth element are assumed to be uniformly distributed).

Rogot et al. (1983) report on linkages of records from the Census Bureau's Current Population Survey with the National Death Index, using each person's name, SSN and date of birth as key matching variables. Based on the results of an evaluation study in which "truth" (match or non-match) was based on a consensus of three raters using all available information for a set of "possible matches,"

they concluded that whenever SSNs agreed, it was appropriate to classify the pair of records as a positive link, provided there was agreement on sex. The use of probabilistic matching procedures was restricted to cases for which the SSNs did not agree or were missing on one or both records.

REFERENCES

Arellano, M.
(n.d.) Optimum utilization of the social security number for matching purposes. No further identification available.
Cobleigh, C. and Alvey, W.
1974 Validating reported social security numbers. American Statistical Association Proceedings, Social Statistics Section, 145-150.
Fellegi, I. and Sunter, A.
1969 A theory for record linkage. Journal of the American Statistical Association, 64(328); 1183-1210.
Hawkes, T. and Harris, R.
1969 An analysis of social security numbers in the SMI actuarial sample. Actuarial Note No. 62. Social Security Administration, U.S. Department of Health, Education, and Welfare.
HEW Secretary's Advisory Committee on Automated Personal Data Systems
1973 Records, Computers and the Rights of Citizens. Washington, D.C.: U.S. Department of Health, Education, and Welfare.
    Chapt. VII. The Social Security Number as a Standard Universal Identifier
    Chapt. VIII. Recommendations Regarding Use of the Social Security Number
Internal Revenue Service
1973 Evaluation of the randomness of the four ending digits of the social security numbers on the master file, 1971. Unpublished report by Mathematical Statistics Branch, Statistics Division. U.S. Department of the Treasury.
Kasprzyk, D.
1983 Social security number reporting, the use of administrative records, and the multiple frame design in the Income Survey Development Program. Pp. 123-144 in Technical, Conceptual and Administrative Lessons of the ISDP (M. David, ed.) Washington, D.C.: Social Science Research Council.
Kilss, B. and Tyler, B.
1974 Searching for missing social security numbers. American Statistical Association Proceedings, Social Statistics Section, 137-144.
Ono, M., Patterson, G. and Weitzman, M.
1968 The quality of reporting social security numbers in two surveys. American Statistical Association Proceedings, Social Statistics Section, 197-205.
Page, W. and Wright, G.
1979 A Statistical Study of the VA Annual Patient Census Sampling Procedure, 1975-1977, Controller Monograph Technical Series, No. 1. Veterans Administration.

Privacy Protection Study Commission
1977 Personal Privacy in an Information Society. Washington, D.C.: U.S. Government Printing Office.
Chapt. 16. The Social Security Number.

Rogot, E., Schwartz, S., O'Conor, K. & Olsen, C.
1983 The use of probabilistic methods in matching census samples to the National Death Index. Pp. 75-80 in Statistics of Income and Related Administrative Record Research: 1983, Internal Revenue Service.

Social Security Administration
1981a Methods of Estimation for the 1973 Exact Match Study. Studies from interagency data linkages, Report No. 10. Washington, D.C.: Department of Health and Human Services.

Social Security Administration
1981b Social security numbers issued, 1937-79. Research and Statistics Notes, No. 7, by F. Bamberger. Washington, D.C.: U.S. Department of Health and Human Services.

Wright, P.
1980 Strategy and tactics in the design of forms. Visible Language 14(2): 151-193.

Table 1.--Social Security Numbers Issued, By Sex of Applicants, 1937-79

(In thousands)

Year	Total	Male	Female
1937[1]	37,139	26,981	10,158
1938	6,304	4,010	2,294
1939	5,555	3,291	2,264
1940	5,227	3,080	2,147
1941	6,678	3,702	2,976
1942	7,637	3,547	4,090
1943	7,426	2,905	4,521
1944	4,537	1,830	2,707
1945	3,321	1,506	1,815
1946	3,022	1,432	1,590
1947	2,728	1,299	1,429
1948	2,720	1,305	1,415
1949	2,340	1,113	1,227
1950	2,891	1,406	1,485
1951	4,927	2,420	2,507
1952	4,363	2,292	2,071
1953	3,464	1,664	1,800
1954	2,743	1,299	1,444
1955	4,323	2,304	2,019
1956	4,376	2,391	1,985
1957	3,639	1,793	1,846
1958	2,920	1,384	1,536
1959	3,388	1,645	1,743
1960	3,415	1,663	1,752
1961	3,370	1,665	1,705
1962	4,519	2,109	2,410
1963	8,617	3,739	4,878
1964	5,623	2,707	2,916
1965	6,131	2,746	3,385
1966	6,506	2,894	3,612
1967	5,920	2,855	3,065
1968	5,862	2,856	3,006
1969	6,289	3,105	3,184
1970	6,132	3,004	3,128
1971	6,401	3,122	3,279
1972	9,564	3,948	5,616
1973	10,038	4,849	5,189
1974	7,998	3,950	4,048
1975	8,164	3,992	4,172
1976	9,043	4,507	4,536
1977	7,724	3,872	3,852
1978	5,260	2,682	2,578
1979	5,213	2,649	2,564

[1]Includes issuances in November and December 1936.

Source: Social Security Administration, 1981b.

ATTACHMENT B

Form SS-5.--Application for a Social Security Number Card

DEPARTMENT OF HEALTH AND HUMAN SERVICES
SOCIAL SECURITY ADMINISTRATION

Form Approved
OMB No. 0960-0066

FORM SS-5 — APPLICATION FOR A SOCIAL SECURITY NUMBER CARD
(Original, Replacement or Correction)

MICROFILM REF. NO. (SSA USE ONLY)

Unless the requested information is provided, we may not be able to issue a Social Security Number (20 CFR 422-103(b))

INSTRUCTIONS TO APPLICANT ▶ Before completing this form, please read the instructions on the opposite page. You can type or print, using pen with dark blue or black ink. Do not use pencil.

1

NAA	NAME TO BE SHOWN ON CARD	First	Middle	Last
NAB	FULL NAME AT BIRTH (IF OTHER THAN ABOVE)	First	Middle	Last
ONA	OTHER NAME(S) USED			

2

| STT | MAILING ADDRESS | (Street/Apt. No., P.O. Box, Rural Route No.) | | | |
| CTY | CITY | STE | STATE | ZIP | ZIP CODE |

3 CSP CITIZENSHIP (Check one only)
- ☐ a. U.S. citizen
- ☐ b. Legal alien allowed to work
- ☐ c. Legal alien not allowed to work
- ☐ d. Other (See instructions on Page 2)

4 SEX
- ☐ MALE
- ☐ FEMALE

5 ETB RACE/ETHNIC DESCRIPTION (Check one only) (Voluntary)
- ☐ a. Asian, Asian-American or Pacific Islander (Includes persons of Chinese, Filipino, Japanese, Korean, Samoan, etc., ancestry or descent)
- ☐ b. Hispanic (Includes persons of Chicano, Cuban, Mexican or Mexican-American, Puerto Rican, South or Central American, or other Spanish ancestry or descent)
- ☐ c. Negro or Black (not Hispanic)
- ☐ d. Northern American Indian or Alaskan Native
- ☐ e. White (not Hispanic)

6 DOB DATE OF BIRTH ▶ MONTH | DAY | YEAR | **7** AGE PRESENT AGE | **8** PLB PLACE OF BIRTH ▶ CITY | STATE OR FOREIGN COUNTRY | FCI ☐

9

| MNA | MOTHER'S NAME AT HER BIRTH | First | Middle | Last (Her maiden name) |
| FNA | FATHER'S NAME | First | Middle | Last |

10

PNO a. Has a Social Security number card ever been requested for the person listed in item 1? ☐ YES(2) ☐ NO(1) ☐ Don't know(1) If yes, when: ▶ MONTH | YEAR

b. Was a card received for the person listed in item 1? ☐ YES(3) ☐ NO(1) ☐ Don't know(1) If you checked yes to a or b, complete items c through e; otherwise go to Item 11.

SSN c. Enter the Social Security number assigned to the person listed in item 1. ☐☐☐ - ☐☐ - ☐☐☐☐

NLC d. Enter the name shown on the most recent Social Security card issued for the person listed in item 1. PDB e. Date of birth correction (See Instruction 10 on page 2) ▶ MONTH | DAY | YEAR

11 DON TODAY'S DATE ▶ MONTH | DAY | YEAR **12** Telephone number where we can reach you during the day. Please include the area-code. ▶ HOME | OTHER

ASD WARNING: Deliberately furnishing (or causing to be furnished) false information on this application is a crime punishable by fine or imprisonment, or both.

IMPORTANT REMINDER: SEE PAGE 1 FOR REQUIRED EVIDENTIARY DOCUMENTS.

13 YOUR SIGNATURE **14** YOUR RELATIONSHIP TO PERSON IN ITEM 1 ☐ Self ☐ Other (Specify)

WITNESS (Needed only if signed by mark "X") | WITNESS (Needed only if signed by mark "X")

DO NOT WRITE BELOW THIS LINE (FOR SSA USE ONLY).

SSN ASSIGNED ☐☐☐ - ☐☐ - ☐☐☐☐

DTC SSA RECEIPT DATE
NPN
BIC SIGNATURE AND TITLE OF EMPLOYEE(S) REVIEWING EVIDENCE AND/OR CONDUCTING INTERVIEW
DOC | NTC | CAN
TYPE(S) OF EVIDENCE SUBMITTED
☐ MANDATORY IN PERSON INTERVIEW CONDUCTED
IDN | ITV | DCL
DATE
DATE

Form SS-5 (5-84)   Destroy prior editions

221

Distribution of Social Security Numbers as of January 2, 1985: Highest Group Number Issued Within Each Area Number*

Area	Grp	Area	Grp	Area	Grp	Area	Grp	Area	Grp	Area	Grp	Area	Grp	Area	Grp
		059	68	118	68	177	64	236	25	295	80	354	72	413	45
001	68	060	68	119	68	178	64	237	45	296	80	355	72	414	45
002	68	061	68	120	68	179	64	238	45	297	80	356	72	415	43
003	66	062	68	121	68	180	64	239	45	298	80	357	72	416	19
004	82	063	68	122	66	181	64	240	43	299	80	358	72	417	19
005	80	064	68	123	66	182	64	241	43	300	80	359	72	418	19
006	80	065	68	124	66	183	64	242	43	301	80	360	72	419	19
007	80	066	68	125	66	184	64	243	43	302	80	361	72	420	19
008	66	067	68	126	66	185	64	244	43	303	92	362	94	421	19
009	64	068	68	127	66	186	64	245	43	304	92	363	94	422	19
010	66	069	68	128	66	187	64	246	43	305	92	364	94	423	19
011	66	070	68	129	66	188	64	247	59	306	92	365	94	424	17
012	64	071	68	130	66	189	64	248	59	307	92	366	94	425	51
013	64	072	68	131	66	190	64	249	59	308	92	367	94	426	51
014	64	073	68	132	66	191	64	250	57	309	92	368	94	427	49
015	64	074	68	133	66	192	64	251	57	310	92	369	94	428	49
016	64	075	68	134	66	193	64	252	49	311	92	370	94	429	57
017	64	076	68	135	78	194	64	253	49	312	92	371	94	430	57
018	64	077	68	136	78	195	64	254	49	313	92	372	94	431	55
019	64	078	68	137	78	196	64	255	49	314	92	373	94	432	55
020	64	079	68	138	76	197	64	256	49	315	92	374	94	433	55
021	64	080	68	139	76	198	64	257	47	316	92	375	94	434	55
022	64	081	68	140	76	199	64	258	47	317	92	376	94	435	55
023	64	082	68	141	76	200	62	259	47	318	74	377	94	436	55
024	64	083	68	142	76	201	62	260	47	319	74	378	94	437	55
025	64	084	68	143	76	202	62	261	99	320	74	379	94	438	55
026	64	085	68	144	76	203	62	262	99	321	74	380	94	439	53
027	64	086	68	145	76	204	62	263	99	322	74	381	94	440	84
028	64	087	68	146	76	205	62	264	99	323	74	382	94	441	84
029	64	088	68	147	76	206	62	265	99	324	74	383	92	442	84
030	64	089	68	148	76	207	62	266	99	325	74	384	92	443	84
031	64	090	68	149	76	208	62	267	99	326	74	385	92	444	84
032	64	091	68	150	76	209	62	268	82	327	74	386	92	445	84
033	64	092	68	151	76	210	62	269	82	328	74	387	92	446	82
034	64	093	68	152	76	211	62	270	82	329	74	388	92	447	82
035	54	094	68	153	76	212	06	271	82	330	74	389	92	448	82
036	52	095	68	154	76	213	06	272	82	331	74	390	92	449	69
037	52	096	68	155	76	214	06	273	82	332	74	391	92	450	69
038	52	097	68	156	76	215	06	274	82	333	74	392	92	451	69
039	52	098	68	157	76	216	06	275	82	334	74	393	92	452	69
040	76	099	68	158	76	217	06	276	82	335	74	394	92	453	69
041	76	100	68	159	64	218	06	277	82	336	74	395	92	454	69
042	76	101	68	160	64	219	06	278	82	337	74	396	92	455	69
043	76	102	68	161	64	220	04	279	82	338	74	397	92	456	69
044	76	103	68	162	64	221	68	280	82	339	74	398	92	457	69
045	76	104	68	163	64	222	66	281	82	340	74	399	92	458	69
046	76	105	68	164	64	223	33	282	82	341	74	400	25	459	69
047	76	106	68	165	64	224	33	283	82	342	72	401	25	460	69
048	76	107	68	166	64	225	33	284	82	343	72	402	25	461	69
049	74	108	68	167	64	226	33	285	82	344	72	403	25	462	69
050	68	109	68	168	64	227	33	286	82	345	72	404	25	463	69
051	68	110	68	169	64	228	33	287	82	346	72	405	25	464	69
052	68	111	68	170	64	229	33	288	82	347	72	406	23	465	69
053	68	112	68	171	64	230	31	289	82	348	72	407	23	466	69
054	68	113	68	172	64	231	31	290	80	349	72	408	45	467	69
055	68	114	68	173	64	232	27	291	80	350	72	409	45	468	04
056	68	115	68	174	64	233	27	292	80	351	72	410	45	469	04
057	68	116	68	175	64	234	27	293	80	352	72	411	45	470	04
058	68	117	68	176	64	235	25	294	80	353	72	412	45	471	04

472	04	495	88	518	11	541	11	564	81	587	49	610	00	706	18
473	04	496	88	519	11	542	11	565	81	588	00	611	00	707	18
474	02	497	88	520	04	543	11	566	81	589	30	612	00	708	18
475	02	498	88	521	43	544	11	567	81	590	30	613	00	709	18
476	02	499	88	522	43	545	83	568	81	591	30	614	00	710	18
477	02	500	88	523	43	546	83	569	81	592	30	615	00	711	18
478	06	501	04	524	43	547	83	570	81	593	30	616	00	712	18
479	06	502	02	525	53	548	83	571	81	594	28	617	00	713	18
480	06	503	04	526	99	549	83	572	81	595	28	618	00	714	18
481	06	504	04	527	99	550	81	573	81	596	00	619	00	715	18
482	06	505	13	528	49	551	81	574	76	597	00	620	00	716	18
483	06	506	13	529	49	552	81	575	27	598	00	621	00	717	18
484	04	507	11	530	08	553	81	576	27	599	00	622	00	718	18
485	04	508	11	531	96	554	81	577	11	600	16	623	00	719	18
486	90	509	88	532	96	555	81	578	08	601	14	624	00	720	18
487	90	510	88	533	96	556	81	579	08	602	00	625	00	721	18
488	90	511	88	534	96	557	81	580	19	603	00	626	00	722	18
489	88	512	88	535	94	558	81	581	99	604	00	700	18	723	18
490	88	513	88	536	94	559	81	582	99	605	00	701	18	724	28
491	88	514	86	537	94	560	81	583	99	606	00	702	18	725	18
492	88	515	86	538	94	561	81	584	97	607	00	703	18	726	18
493	88	516	04	539	94	562	81	585	51	608	00	704	18	727	10
494	88	517	04	540	11	563	81	586	78	609	00	705	18	728	14

*First three digits of the social security number are area numbers; second two digits are group numbers.

Group 00 is not a valid group -- it is for program purposes only.

Excerpt from
Validation and Screening Techniques for Social Security Numbers

VALIDATION OF SSN'S

Minimum information needed to validate an SSN is the person's name, sex, date of birth and the alleged SSN. Validation occurs only when the information on a current transaction exactly matches or can be reconciled with the information on the Alphident/Numident data bases or the microfilm subfiles of these systems. In certain circumstances, additional matching information is needed before validation can occur. If earnings are reported without an SSN or with an SSN or name that does not agree with these files and the correct SSN cannot be determined through internal screening operations, the employer or the worker is asked to furnish additional information to identify the record. The Internal Revenue Service (IRS) uses a similar system to validate SSN's of taxpayers.

MANUAL SCREENING OF DUPLICATE
AND ORIGINAL SSN APPLICATIONS

The electronic screening operation to which every application is subjected is capable of processing roughly 85 percent of all applications input by field offices. Through a sophisticated series of screening grids, the computer makes a decision: is this applicant already represented in the Alphident data base? If the decision is yes, the previously assigned SSN is identified and a replacement card is prepared and mailed. If the decision is no, a number is assigned and a card is printed and mailed.

However, the decision-making capability of the system is deliberately limited because some applications have identifying information common to others or conditions exist which should receive a clerical review. These applications produce worksheets which are processed manually by OCRO.

Worksheets to be screened are checked against the Alphident Microfilm File and the Alphident Microfiche File, using the name and date of birth shown on the application. If an SSN is not located for the name and date of birth shown, another search is made using dates of birth somewhat different from the one given on the application. If an SSN is still not located, certain other variations are checked, including name at birth or on the signature line if different from the name in item 1; acceptable variations of common first names; dropping middle name shown; substituting different middle initials; substituting maiden surname for middle given name for married females; substituting initials only in place of complete given names; etc. Once a "possible" SSN is located, verification can be made immediately since full identifying information is available on the Alphident files. See RM 00204.020 for procedures for handling "UTL" and "Investigate" items.

THE ALPHIDENT MICROFILM AND
MICROFICHE FILES

The electronic Alphident file is updated daily. If an SSN holder loses the social security card within the first days after it was issued, the number can be located and verified electronically.

The Alphident Microfilm File is an alphabetical file based on the Russell Soundex coding system. It contains essentially the same information as the electronic file.

Because the Alphident Microfilm File is updated only every 3 months, each week an accretion file is prepared on microfiche. This file contains all SSN assignments and corrections to our records processed during the preceding 12 weeks. This file is referred to when there is reason to believe that there was a recent SSN action for an individual.

Each record entry on both the Alphident Microfilm and the Alphident Microfiche Files consists of the following:

DATA	POSITIONS
Blank	1
Soundex	2-5
Blank	6
Applicant's Surname	7-27
Applicant's Given Name	28-43
Applicant's Middle Name	44-45
Month of Birth	56-57
Blank	58
Day of Birth	59-60
Blank	61
Century of Birth	62
Year of Birth	63-64
Blanks	65-66
SSN	67-77
Blank	78
Mother's Surname	79-91
Mother's Given Name	92-102
Mother's Given Initial	103
Blank	104
Sex/Race	105-106
Blank	107
Father's Surname	108-120
Father's Given Name	121-131
Father's Middle Initial	132
Blank	133
City/County of Birth	134-140
State/Country of Birth	141-142
Blanks	143-144
Form/Entry	145-146
Blanks	147-148
Reference Number	149-159
Blank	160

COMMON NAMES IN THE ALPHIDENT FILE

There are over 360 million records in the Alphident File, representing over 277 million SSN's assigned. Many of the names in the file are the same or are very similar. This is why it is extremely important to get complete and accurate identifying information on original applications and on requests for duplicate SSN cards. It is equally important to obtain information that is consistent with that on the original application. Applicants who have lost their original cards should be questioned closely to find out if any of the information on the current application is now different from that which they showed on their original application.

The latest tabulation of common surnames in the SSN file was made in 1974. Some examples of the number of times a common name could appear in Alphident are given below.

NAME	NUMBER OF ITEMS IN ALPHIDENT
Smith	2,382,509
Johnso(n)	1,807,263
Willia(ms)(mson)	1,568,939
Brown	1,362,910
Jones	1,331,205
Miller	1,131,861
Davis	1,047,848
Martin(ez)(son)	1,046,297
Anders(on)	825,648
Wilson	787,825

THE RUSSELL SOUNDEX CODE

By using the Russell Soundex Code system, searching for possible SSN's on the Alphident film and fiche in OCRO is accomplished quickly.

Here are the basic rules for using the Soundex Code.

Use the first letter of the surname, then code the remaining letters as follows:

LETTERS	CODE SYMBOLS
BPFV	1
CGJKQSXZ	2
DT	3
L	4
MN	5
R	6

Vowels are not coded, nor are the letters W, H, and Y. Two successive letters with the same code numbers are coded only once.

Example:
"Mack" is coded M-200. The "a" is not coded since it is a vowel. "c" falls under code symbol 2. "k" also falls under code symbol 2, but is not used since two successive letters with the same code sumbol are coded only once. Since the complete Soundex Code must consist of the first letter of the name followed by three numbers, we add enough zeros to complete the 3-digit code.

Here are some other examples:

1. Snyder - S-536
2. Way - W-000
3. Bear - B-600
4. Brown - B-650

LIMITATIONS IN OCRO SCREENING FOR SSN's

When an applicant has indicated a previous SSN in item 10 of the SS-5 and the correct number cannot be found in the electronic or OCRO screening operations, the data are returned via form SSA-4310 to the district office. This is because studies show that many such applicants are mistaken in stating they previously applied for a number, and it is not worthwhile spending additional time on the case unless different information can be found. When the district office receives a form SSA-4310 from OCRO, it should recontact the applicant for any different information that may be useful in screening. See RM 00204.020 A.1. Take appropriate action, but do not return the SSA-4310 to OCRO.

Upon recontacting the applicant, the district office may discover that a married woman obtained her original SSN under a first husband's name, but is now applying for the duplicate in her second husband's name; that a man who calls himself "Winslow" obtained his number earlier in life as "Buddy;" or that Mr. Kline's record was set up originally under "Cline." There is also a possibility that the applicant may be able to locate the previously issued SSN on an old pay stub or by asking a present or a past employer. This new information may enable OCRO to locate the original SSN. If the applicant is unable to give any information different from what was previously given and is unable to locate the alleged number, the district office has no other choice but to request assignment of an original SSN. However, this should be done only as a last resort, particularly if the person has earnings under the original number which might not be credited when the SSN holder applies for benefits.

These facts point up the need for obtaining the most accurate information possible during the initial interview with the applicant, whether it be for an original or duplicate SSN card; otherwise, multiple numbers may result. Any reasonable assistance should be extended to the applicant to help find out definitely what the alleged prior SSN is. (See RM 00202.025 I.10.)