EXPERIENCE WITH COMPUTER MATCHING OF NAMES

William Phillips, Jr. and Anita K. Bahn*
National Institute of Mental Health

For many years, epidemiologists, administrators and biostatisticians have felt the need for a large comprehensive case register for mental illness similar to those in use for cancer and tuberculosis. Such a register can provide unduplicated patient counts for rates of diagnosed incidence and prevalence of mental disorders, data on patient movement between facilities, changes in diagnosis and other longitudinal information not systematically available through any other means. In order to initiate such studies, on July 1, 1961 a psychiatric case register was established for the State of Maryland, in cooperation with the National Institute of Mental Health (1).

When planning for the project, the main problem foreseen was the linkage of records for the individual who receives services in more than one psychiatric facility. Psychiatric services to Maryland residents are provided by over 100 facilities and each facility has its own patient numbering system.

In exploring the possible methods of record linkage, name and address was considered essential. (State legislation now protects the confidentiality of this information and assures its use for research purposes only.) Although social security number has discriminating power and is now requested, it could not be used as the primary matching factor because of the large number of child patients without a social security number, the resistance of patients to furnishing this information, and the lack of social security numbers on reports already on file as part of the ongoing State reporting system.

A second major consideration was the eventual large size of the register due to the relatively few deaths in this patient population. For example, within five years we expect a cumulative file of 120,000 psychiatric experiences representing 60,000 to 80,000 different persons. Primary clerical matching of names, although feasible for the first year or two, would soon become unwieldy and unreliable. The anticipated size of the register warranted the use of computer methods for person matching.

Probability factors for computer decision based on the frequency of names and other characteristics of our unique population were not available pending detailed population analysis and large scale matching experience (2). We planned, therefore, to establish computer methods which would be largely "trial and error" and would attempt to duplicate clerical processes and judgment in checking on a number of identifying factors simultaneously in deciding on the match of a pair of records (3).

Initial rules for computer decision as to whether a match is positive, possible, or rejected were established on the basis of reasonableness, some clerical matching experience, and our general knowledge of the reliability of the reported data. It was decided that clerical review would be made initially of all "positive" and "possible" computer matches, not only to insure against false linkages and for resolution of possible matches, but also to review the results so that our methods could be improved.

Only those computer programs which are essential to the patient matching and record linkage will be described. Our first matching operation involved a comparison of reports of patients on the rolls of psychiatric facilities on July 1, 1961, in order to establish the psychiatric case register. We will later describe the operation of checking new reports with the established register to determine whether the record is for a previous registrant or represents an accretion to the master file.

Establishment of the register file

A total of 22,869 cases were enrolled in the various facilities on July 1, 1961. Our first step was the automatic assignment of a temporary or pseudo register number to each record in order to facilitate record retrieval, correction, and linkage. Beginning with the pseudo-number 5, numbers were assigned by an arithmetic progression of tens. During matching, if two records were considered to be a "positive" match, the pseudo-number of the second listed record was replaced by the pseudo-number, minus one, of the first record. If a second match was detected, the pseudo-number for this record was also replaced, but with the first pseudo-number less two. This process provided for the identification of ten matches for the same individual. (The maximum number turned out to be three.) At a later date, all records were sorted using this pseudo register number which resulted in all data records for the same individual falling in sequence. We were then able to combine all the data for each "individual" and assign a single permanent register number to each record. (All records for each individual are maintained in the master identity file as part of the permanent record and used in subsequent matching and updating processes.)

In making a check for duplication, the ideal method would be to check each record against all other records on all common factors of identity.

---

* Mr. William Phillips, Jr. is Digital Computer Programmer and Dr. Anita K. Bahn is Chief, Outpatient Studies Section, Biometrics Branch, National Institute of Mental Health.

This, of course, is not feasible with current equipment. It would mean a maximum of 261 million comparisons or $\left(\frac{22869^2 - 22869}{2}\right)$ of one record against another and possibly 6,000 hours of computer time (8.5 milliseconds × number of comparisons). Therefore, we had to group records into blocks of a size which can be handled by our computer.

We chose to group the records by a phonetic code, commonly referred to as the Russell Soundex code. This is a system whereby the consonant sounds of the surname are assigned numbers which are accumulated into a code designed to compensate for the common misspellings. For example, Brown, Browne and Braun all have the same code (1650). In usual practice, the first letter of the surname comprises part of the code and is not assigned a number. We varied this system by coding the first letter as well as the remaining letters so that names such as Cohn and Kohn would be compared (see Figure 1). The codes are all four digits in length.

The Soundex code was assigned by a computer program. For married women with maiden names, the record was reported differing only in the Soundex code assigned.

By use of the four digital Soundex code the file was divided into 1,007 different Soundex groups, of which 31 had more than 150 records. The largest group was "2520" with 519 records. Although such different names as Jones, James, and King, for example, were included in this group, our concordance rules about correspondence of letters in the name eliminated such obvious mismatches (see Figure 2).

We read all of the records of each group into core memory of the computer and cross-checked every record within the group. The first check of this program compares surname, address, first name, and birth year (see Figure 2). The tolerance rules established for concordance of these factors in a positive match were considered conservative. If any field was missing, the comparison of these fields was considered not in concordance. Whenever a match was accepted as "positive," further checking of the record was discontinued. If agreement within the specified tolerances was not achieved, but the match was considered as possible, a second check was automatically made on the basis of social security number and maiden name, factors which can aid in positive identification. The final group of factors or third automatic check, if agreement was still in doubt, consisted of sex, race, and complete birth date.

We made a total of approximately 1,700,000 comparisons of one record against the other. Total computer running time was four hours. At the completion of this program, 627 "positive" and 1,011 "possible" matches were presented in list form for clerical scrutiny. The print record contained the complete name, address, sex, race, birth date, facility code, patient case number, and, in addition, a pseudo register number.

Two clerks spent two days each examining these matches. These clerks were primarily key punch operators who were familiar with the codes but not with the matching program. They determined that 553 "positive" matches were truly positive, 169 "possible" matches were true linkages, and 736 "possible" matches were not true linkages. After this preliminary check, another clerk, who had a thorough knowledge of the logic of the program, spent two weeks checking the doubtful items in detail, referring back to case records or querying facilities where necessary. The final count of positive and false linkages by decision rule is summarized in Figure 3A. A total of four percent of the records or 604 of the positive and 201 of the possible matches were classified as "duplicates."

The clerks then prepared a final list of records requiring change. This included, in addition to discrepant information, corrections to the pseudo register number for "positive" matches determined to be non-matches and for "possible" matches determined to be "positive." An additional clerk day was required to punch, verify, and review the 298 cards used to make adjustments. The program to correct the file of 26,051 records (including 3,182 maiden name records) ran approximately ten minutes.

In reviewing the efficiency of this first unduplication program, it should be noted that few social security numbers were available for assistance in checking at the second stage of the program, and, actually, all social security number comparisons came out as unequal. Also, the month and day of birth were missing in a large percentage of our cases, which caused many linkages to be listed as possible instead of positive. Clerical determination that these cases were positive was aided by the fact that most of these patients were on the books of both the State mental hospital (i.e. on convalescent leave) and the clinic attached to that hospital (see Figure 3B).

Since the establishment of the register, we have been successful in obtaining many of the missing birth dates. This will make it possible to conduct a second primary grouping of the master file on the basis of birth date in order to associate "duplicate" records whose Soundex codes are dissimilar. This month and day of birth check has already been used in the updating programs which will be described next.

Updating the master file (first year's experience)

In updating the register master files with fiscal 1962 data, we began with 22,323 admissions to psychiatric care. First we extracted from the admission (detail) records only the information necessary for our linkage checks. Pseudo register numbers were again assigned to each record

for later association, and in addition we added another field to the linkage search record for inserting a located permanent register number (see Figure 4).

A preliminary matching program by facility code and case number searched for readmitted registrants who could be identified by their unit case number.

A person-matching program was used to link the remaining admissions with the master file (see Figure 5). We learned from our previous check that the addition of sex as a grouping factor would not impair the efficiency of our program and would shorten the computer running time. In addition, some changes in final decision rules were made (from "reject" to "possible") to permit clerical verification of our logic.

We read into computer memory core all of the records from the master file for the same Soundex-sex group. We then read the detail records for the same group into core memory, one by one, and compared each record to every record in the master group. When a match which we considered to be a "positive" linkage was detected, we extracted the register number from the master record and rewrote the detail record on a located register number file. We also prepared a print record showing all information for both the detail and master records for later clerical verification. Further computer checking of these records was discontinued.

When the computer program detected a possible linkage, a print record was prepared showing all details and computer checking was continued. As with our program to establish the register, there could be several "possible" linkages for each detail record. If no "positive" linkage was detected, the detail record was rewritten on an unlocated register number tape file to be processed through the next program.

In the first person-matching program used for the updating process, the computer made 1,501,690 record comparisons and classified 2,685 records as positive linkages and 3,219 records as possible linkages (see Figures 6 and 7). Each set of print records (master and detail) were examined clerically. We changed a total of 317 computer decisions (29 "positive" to "reject" and 288 "possible" to "accept") (Figures 6 and 7). Twelve of the 29 positive changes appear to be twins (A0630 category). There were 2,944 net linkages from this program. This figure is inflated slightly because of a number of duplicate linkages based on both the maiden name and the married name for the same person. This duplication will not interfere with our processing as we had planned for this eventuality. The actual number of person-linkages was 2,850.

Our analysis has pointed out several desirable refinements to this program. Two of the major modifications are the discontinuance of print records for the A0000 category whenever we have agreement on month and day of birth or social security number and the changing of A0232 to a positive linkage. These changes will reduce the amount of clerical work.

Clerical scrutiny of the listings required 30.5 hours of clerical time. In addition, four clerical days were spent in checking further into the 112 linkages where a decision could not be made from the listing, and six hours were spent in punching, verifying, and reviewing the 317 cards to correct the "located" and "unlocated" tape files.

The final rejects from this first updating program were then processed through a second program in which month of birth, day of birth, and sex were used as primary grouping factors (see Figure 8). These factors were chosen in order to check cases where the names were entirely different or so misspelled that the Soundex code did not permit them to be compared. The same general processing principles as in the previous program were used (i.e. each detail record was checked to every master record of the same month of birth, day of birth, and sex).

The adjusted unlocated register number file from the Soundex check was used as input to this program along with the master identity file. The computer listed 300 matches of which 219 were classified as "positive" and 81 as "possible" (see Figure 8). Running time for the program was one hour and 20 minutes. Again, we checked all linkages clerically and made adjustments to both the located and unlocated files. This required 6.5 clerk hours.

The yield of this program was very meager. We had a net of only 85 linkages. A few of these linkages were unique because of the fact that we had no name on the admission record (about one percent of the records have no name) and others had completely different surnames. There are several factors which could account for this small yield: the large number (2,000) of cases with month and day of birth still missing and the previous detection of most linkages by the Soundex program. This month and day of birth check may become more valuable in time when there are more possibilities of name change due to adoption of children or remarriage where maiden name is not reported.

After checking the input against the master file for a previously assigned register number, we had remaining a file of approximately 19,000 admission records which were not linked. These records represent new admissions to psychiatric service since inception of the register. As there are undoubtedly duplications within this file, we are planning to use person-matching programs similar to the Soundex check for persons on the rolls July 1, 1961 plus the month and day of birth check. These programs will include the further refinements pointed out by the 1962 updating programs.

Discussion

We envision a continually expanding case register which may ultimately contain over a quarter of a million names. At present, register maintenance and updating require the matching each year of over 20,000 additional records and 30,000 resident death certificates. It is expected that the number of records to be matched annually will increase due to the opening of new facilities. The development of computer programs for primary person matching seems warranted, therefore.

A series of rules for this operation has been established on judgmental grounds. A record is kept of the outcome of each decision rule so that its yield of positive and false matches can be determined and the rule modified by experience. Because of the large amount of missing data in several fields, we will have to distinguish missing information and other unequal comparisons. The large number of linkages with disagreement in address reflects in part differences in the punching of addresses. We have standardized this punching as much as possible, but abbreviations used in the past are still causing difficulties. It is also our intention to revise our programs in the future to decrease the number of print records and thereby reduce the amount of clerical work.

Our first computer matching operation of 23,000 records with each other to establish the register required 4.2 computer hours and 120 clerical hours for residual matching. Based upon the experience of another register one-fourth the size, it is estimated that approximately six man months of clerical time would have been required if the matching operation was entirely clerical, that is, for a clerk to review each record against the name of other records in an alphabetical listing.

In the linkage of new admissions to our file, we benefited from our first duplication check. The computer program for the linkage of records by Soundex grouping ran for three hours and 23 minutes and 68.5 hours of clerical time were required. For the check by month and day of birth grouping, one hour and 20 minutes of computer time was required, and clerical personnel spent 6.5 hours in scrutinizing the linkage and preparing correction punch cards.

Although the programing and other costs of making these checks are high in proportion to the number of cases, our costs will decrease over the years with refinements to our programs and improvement in the accuracy of our data. From the results of sample studies based upon clerical review of the alphabetical master file, we estimate that we are missing between .3 and .5 percent of the linkages. We are planning further sample studies to improve our computer methods for linking these misses. A primary sort on the basis of address is also planned.

In addition to routine statistical checks of each master record for internal consistency of data, there will be a continuous review of sample files to detect false linkages.

As soon as we have completed the updating of our master files with fiscal 1962 data, we are planning a series of "death clearance" programs. We have obtained a duplicate set of punch cards for all Maryland deaths during fiscal 1962. It is our intention to run these records through the same programs used to update our register and obtain the cause and date of death for any of our patients who have died.

There is still much work and experimentation to be undertaken to improve the efficiency of the maintenance of such a file by an electronic computer. Of course, additional identifying information on patients such as birth place and mother's maiden name would increase the discriminatory power of the computer. Such information cannot now be requested. However, we believe that the greater accuracy, consistency, and efficiency of our present computer program as compared with clerical operation for a register this size justify continued experimentation. Furthermore, these methods permit the matching of large rosters of individuals obtained from other sources, such as welfare and criminal records, to our master file.

We believe our methods to be generally applicable to other types of person-matching operations.

References

(1) Phillips, William, Jr., Gorwitz, Kurt, and Bahn, Anita K.: Electronic Maintenance of Case Registers. Pub. Health Rep. 77: 503-510, June 1962.

(2) Newcombe, H. B., Kennedy, J. M., Axford, S. J., James, A. P.: Automatic Linkage of Vital Records. Science 130: October 16, 1959.

(3) Phillips, William, Jr., Bahn, Anita K., Miyasaki, Mabel: Person-Matching by Electronic Methods. Communications of the ACM 5: 404-407, July 1962.

FIGURE 1

Phonetic Code Used in Maintaining Register

The consonants of the surname are assigned numbers according to the following schedule and rules.

Code	Letters
1	B, F, P, V
2	C, G, J, K, Q, S, X, Z
3	D, T
4	L
5	M, N
6	R
Not coded	A,E,I,O,U,W,H,Y

Rule 1

The code for any name consists of 4 digits. If a name does not have sufficient coded consonants, zeros are added to complete the code. (e.g. Lee: 4000). If there are more than 4 coded consonants, the code is truncated. (e.g. Malinauskas: 5452).

Rule 2

If 2 or more consonants which have the same coded number come together, they are coded as only one letter.

Phillips is coded: 1412

P is coded 1	I is not coded
H & I are not coded	P is coded 1
LL is coded 4	S is coded 2

Dickson is coded: 3250

D is coded 3, I is not coded, C K and S all have the same code value 2, and occuring together they are coded as one letter, O is not coded, N is coded 5, and a zero is added to complete the code.

Rule 3

Consonants having the same code number but separated by one or more vowels (a,e,i,o,u,y) are coded individually.

Dicicio is coded: 3220
Wyman is coded: 5500

Rule 4

W and H do not separate consonants. If two consonants having the same code are separated by a W or H, they are coded as one consonant.

Sachs - 2200

FIGURE 2

Soundex Check for Linkage

July 1, 1961 Population

		FIRST CHECK					SECOND CHECK			THIRD CHECK			
Reference Code	Soundex code	Surname	First name	Addr.	Birth year range	Decision	Soc. Sec.	Maid. name	Decision	Sex & race	Birth mo. & day	Birth year	Decision
A0000	0	0	0	0	0	Accept							
A0100	0	0	0	0	1	Possible	0	0	Accept				
A0110							0	1	Accept				
A0120							1	0	Accept				
A0130							1	1	Possible	0	0	1	Accept
A0131										0	1	1	Reject
A0132										1	0	1	Possible
A0133										1	1	1	Reject
A0200	0	0	0	1	0	Possible	0	0	Accept				
A0210							0	1	Accept				
A0220							1	0	Accept				
A0230							1	1	Possible	0	0	0	Accept
A0231										0	0	1	Accept
A0232										0	1	0	Possible
A0233										0	1	1	Possible
A0234										1	0	0	Accept
A0235										1	0	1	Possible
A0236										1	1	0	Reject
A0237										1	1	1	Reject
A0300	0	0	0	1	1	Reject							
A0400	0	0	1	0	0	Possible	0	0	Accept				
A0410							0	1	Accept				
A0420							1	0	Accept				
A0430							1	1	Possible	0	0	0	Accept
A0431										0	0	1	Accept
A0432										0	1	0	Possible
A0433										0	1	1	Reject
A0434										1	0	0	Possible
A0435										1	0	1	Possible
A0436										1	1	0	Reject
A0437										1	1	1	Reject
A0500	0	0	1	0	1	Possible	0	0	Accept				
A0510							0	1	Accept				
A0520							1	0	Accept				
A0530							1	1	Possible	0	0	1	Possible
A0531										0	1	1	Reject
A0532										1	0	1	Possible
A0533										1	1	1	Reject
A0600	0	0	1	1	0	Possible	0	0	Accept				
A0610							0	1	Accept				
A0620							1	0	Accept				
A0630							1	1	Possible	0	0	0	Accept
A0631										0	0	1	Possible
A0632										0	1	0	Possible
A0633										0	1	1	Reject
A0634										1	0	0	Possible
A0635										1	0	1	Possible
A0636										1	1	0	Reject
A0637										1	1	1	Reject

(Cont'd.)

FIGURE 2
Soundex Check for Linkage
(Cont'd.)
July 1, 1961 Population

		FIRST CHECK					SECOND CHECK			THIRD CHECK			
Reference code	Soundex code	Surname	First name	Addr.	Birth year range	Decision	Soc. Sec. #	Maid. name	Decision	Sex & race	Birth mo. & day	Birth year	Decision
A0700	0	0	1	1	1	Reject							
A0800	0	1	0	0	0	Accept							
A0900	0	1	0	0	1	Possible	0	0	Accept				
A0910							0	1	Accept				
A0920							1	0	Accept				
A0930							1	1	Possible	0	0	1	Accept
A0931										0	1	1	Reject
A0932										1	0	1	Possible
A0933										1	1	1	Reject
A1000	0	1	0	1	0	Reject							
A1100	0	1	0	1	1	Reject							
A1200	0	1	1	0	0	Possible	0	0	Accept				
A1210							0	1	Accept				
A1220							1	0	Accept				
A1230							1	1	Possible	0	0	0	Possible
A1231										0	0	1	Possible
A1232										0	1	0	Possible
A1233										0	1	1	Reject
A1234										1	0	0	Possible
A1235										1	0	1	Possible
A1236										1	1	0	Reject
A1237										1	1	1	Reject
A1300	0	1	1	0	1	Reject							
A1400	0	1	1	1	0	Reject							
A1500	0	1	1	1	1	Reject							

NB: 0 indicates agreement; 1 indicates discrepancy between the records

Tolerance Rules for Concordance:

Surname - In a one-to-one correspondence of the first 8 letters, only one disagreement allowed

First name - In a one-to-one correspondence of the first 8 letters, only one disagreement allowed

Address - Agreement on street number and first 8 letters of street name

Birth year range -

If current age is:	Range must be within:
0 - 17	2 years
18 - 29	5 years
30 - 49	10 years
50 and over	15 years

Complete agreement required for social security number, maiden name, sex, race, birth month and day, and birth year

FIGURE 3

Psychiatric Case Register (Maryland)
Book Population July 1, 1961

A. Results of Soundex Duplication Check

Major Decision Group or Reference Code	"Positive" Linkages from Computer			"Possible" Linkages from Computer
	Total	Determined Not Correct	Net Positive Linkages	
A0000	357	0	357	6
A01XX	2	0	2	1
A02XX	220	2	218	2/ 918
A04XX	14	1/ 10	4	2
A05XX	0	0	0	0
A06XX	22	10	12	90
A08XX	11	0	11	0
A09XX	0	0	0	0
A12XX	1	1	0	0
Total	627	23	604	3/ 1,011

1/ These were determined to be twins.

2/ This substantial number of "possible" matches was due to the large number of records with missing month and day of birth.

3/ Of these possible linkages, 201 were determined to be positive. There are no counts by detailed reference code. However, the majority were in the A0232 group.

B. Type of Duplications Detected during Processing

Number of Psychiatric Cases on Rolls - July 1, 1961		22,869
Number of Positive Linkages Detected		805
State Hospital Leave and Clinic Care	615	
On Veterans Administration Hospital and Clinic Books	50	
On the Books of an Inpatient Facility and 2 Clinics	8	
On books of Two Inpatient Facilities	23	
In Private Hospital and on Clinic Books	7	
In State Hospital and on Clinic Books	65	
On Books of Two Clinics	37	
Net Number of Patients on Register		22,064

FIGURE 4

TAPE RECORDS USED IN LINKAGE CHECKS

MASTER RECORD USED IN LINKAGE CHECKS

REGISTER NUMBER	FACILITY CODE	PATIENT CASE NUMBER	CODE	SOUNDEX CODE	SEX	RACE	BIRTH			NAME- LAST - FIRST - MIDDLE
							YR	MO	DY	

MASTER RECORD (con't)

STREET NUMBER	STREET NAME	CITY OR TOWN	ZONE	STATE	MAIDEN NAME	SOCIAL SECURITY NUMBER	STATUS CODE	CENSUS TRACT

DETAIL RECORD USED IN LINKAGE CHECK

FACILITY CODE	PATIENT CASE NUMBER	CODE	PSEUDO REGISTER NUMBER	LOCATED REGISTER NUMBER	SOUNDEX CODE	SEX	RACE	BIRTH			NAME - LAST - FIRST-
								YR	MO	DY	

DETAIL RECORD (con't)

MIDDLE	STREET NUMBER	STREET NAME	CITY OR TOWN	ZONE	STATE	MAIDEN NAME	SOCIAL SECURITY NUMBER

FIGURE 5

Soundex Check for Linkage

Fiscal 1962 Updating

		FIRST CHECK					SECOND CHECK			THIRD CHECK				
Reference code	Soundex code & sex	Surname	First name	Addr.	Birth year range	Decision	Soc. Sec. #	Maid. name	Decision	Race	Birth mo. & day	Birth year	Decision	COMPUTER COUNTS
A0000	0	0	0	0	0	Accept								00001187
A0100	0	0	0	0	1	Possible	0	0	Accept					None
A0110							0	1	Accept					00000001
A0120							1	0	Accept					00000002
A0130							1	1	Possible	0	0	1	Accept	00000011
A0131										0	1	1	Possible	00000020
A0132										1	0	1	Possible	None
A0133										1	1	1	Possible	None
A0200	0	0	0	1	0	Possible	0	0	Accept					00000008
A0210							0	1	Accept					00000096
A0220							1	0	Accept					00000192
A0230							1	1	Possible	0	0	0	Accept	00000894
A0231										0	0	1	Accept	00000093
A0232										0	1	0	Possible	00000227
A0233										0	1	1	Possible	00000690
A0234										1	0	0	Accept	00000017
A0235										1	0	1	Possible	00000003
A0236										1	1	0	Reject	00000030
A0237										1	1	1	Reject	00000358
A0300	0	0	0	1	1	Reject								00002956
A0400	0	0	1	0	0	Possible	0	0	Accept					00000002
A0410							0	1	Accept					00000003
A0420							1	0	Accept					00000002
A0430							1	1	Possible	0	0	0	Accept	00000045
A0431										0	0	1	Accept	None
A0432										0	1	0	Possible	00000008
A0433										0	1	1	Reject	00000034
A0434										1	0	0	Possible	None
A0435										1	0	1	Possible	None
A0436										1	1	0	Reject	None
A0437										1	1	1	Possible	None
A0500	0	0	1	0	1	Possible	0	0	Accept					None
A0510							0	1	Accept					None
A0520							1	0	Accept					None
A0530							1	1	Possible	0	0	1	Possible	00000001
A0531										0	1	1	Reject	00000092
A0532										1	0	1	Possible	None
A0533										1	1	1	Reject	None
A0600	0	0	1	1	0	Possible	0	0	Accept					None
A0610							0	1	Accept					00000005
A0620							1	0	Accept					00000027
A0630							1	1	Possible	0	0	0	Accept	00000077
A0631										0	0	1	Possible	00000095
A0632										0	1	0	Possible	00002127
A0633										0	1	1	Reject	00032516
A0634										1	0	0	Possible	00000002
A0635										1	0	1	Possible	00000044
A0636										1	1	0	Reject	00001391
A0637										1	1	1	Reject	00022685
A0700	0	0	1	1	1	Reject								00182702
A0800	0	1	0	0	0	Accept								00000023
A0900	0	1	0	0	1	Possible	0	0	Accept					None
A0910							0	1	Accept					None
A0920							1	0	Accept					None
A0930							1	1	Possible	0	0	1	Accept	None
A0931										0	1	1	Reject	None
A0932										1	0	1	Possible	None
A0933										1	1	1	Reject	None
A1000	0	1	0	1	0	Reject								00005283
A1100	0	1	0	1	1	Reject								00013624
A1200	0	1	1	0	0	Possible	0	0	Accept					None
A1210							0	1	Accept					None
A1220							1	0	Accept					None
A1230							1	1	Possible	0	0	0	Possible	00000002
A1231										0	0	1	Possible	None
A1232										0	1	0	Possible	None
A1233										0	1	1	Reject	00000003
A1234										1	0	0	Possible	None
A1235										1	0	1	Possible	None
A1236										1	1	0	Reject	None
A1237										1	1	1	Reject	None
A1300	0	1	1	0	1	Reject								00000015
A1400	0	1	1	1	0	Reject								00312449
A1500	0	1	1	1	1	Reject								00921648
													TOTAL	1,501,690

NB: 0 indicates agreement;
1 indicates discrepancy between the records

Tolerance Rules for Concordance:

Same as Figure 2.

FIGURE 6

"Positive" Decision Results
Soundex Check Program - 1962 Updating

Reference code	Soundex code & Sex	Surname	First name	Addr.	Birth year range	Soc. Sec.#	Maid. name	Race	Birth mo. & day	Birth year	Computer linkages	Clerical rejects	Net linkages	Clerk time hours
A0000	0	0	0	0	0						1187	0	1187	5.00
A0110	0	0	0	0	1	0	1				1	0	1	.02
A0120	0	0	0	0	1	1	0				2	0	2	.02
A0130	0	0	0	0	1	1	1	0	0	1	11	0	11	.25
A0200	0	0	0	1	0	0	0				8	0	8	.25
A0210	0	0	0	1	0	0	1				96	0	96	.35
A0220	0	0	0	1	0	1	0				192	0	192	.25
A0230	0	0	0	1	0	1	1	0	0	0	894	0	894	2.00
A0231	0	0	0	1	0	1	1	0	0	1	93	0	93	.35
A0234	0	0	0	1	0	1	1	1	0	0	17	0	17	.25
A0400	0	0	1	0	0	0	0				2	0	2	.02
A0410	0	0	1	0	0	0	1				3	0	3	.02
A0420	0	0	1	0	0	1	0				2	0	2	.02
A0430	0	0	1	0	0	1	1	0	0	0	45	0	45	.35
A0431	0	0	1	0	0	1	1	0	0	1	-	-	-	-
A0500	0	0	1	0	1	0	0				-	-	-	-
A0510	0	0	1	0	1	0	1				-	-	-	-
A0520	0	0	1	0	1	1	0				-	-	-	-
A0600	0	0	1	1	0	0	0				-	-	-	-
A0610	0	0	1	1	0	0	1				5	0	5	.03
A0620	0	0	1	1	0	1	0				27	15	12	.50
A0630	0	0	1	1	0	1	1	0	0	0	77	14	63	1.25
A0800	0	1	0	0	0						23	0	23	.20
A0900	0	1	0	0	1	0	0				-	-	-	-
A0910	0	1	0	0	1	0	1				-	-	-	-
A0920	0	1	0	0	1	1	0				-	-	-	-
A0930	0	1	0	0	1	1	1	0	0	1	-	-	-	-
A1200	0	1	1	0	0	0	0				-	-	-	-
A1210	0	1	1	0	0	0	1				-	-	-	-
A1220	0	1	1	0	0	1	0				-	-	-	-
										TOTAL	2685	29	2656	11.13

NB: 0 indicates agreement;
1 indicates discrepancy between the records

FIGURE 7

"Possible" Decision Results
Soundex Check Program - 1962 Updating

Reference code	Soundex code & Sex	Surname	First name	Addr.	Birth year range	Soc. Sec.#	Maid. name	Race	Birth mo. & day	Birth year	Computer possible linkages	Clerical non-linkages	Clerical Net linkages	Clerk Time hours
A0131	0	0	0	0	1	1	1	0	1	1	20	20	0	.25
A0132	0	0	0	0	1	1	1	1	0	1	-	-	-	-
A0133	0	0	0	0	1	1	1	1	1	1	-	-	-	-
A0232	0	0	0	1	0	1	1	0	1	0	227	35	192	5.50
A0233	0	0	0	1	0	1	1	0	1	1	690	622	68	3.50
A0235	0	0	0	1	0	1	1	1	0	1	3	0	3	.08
A0432	0	0	1	0	0	1	1	0	1	0	8	0	8	.25
A0434	0	0	1	0	0	1	1	1	0	0	-	-	-	-
A0435	0	0	1	0	0	1	1	1	0	1	-	-	-	-
A0437	0	0	1	0	0	1	1	1	1	1	-	-	-	-
A0530	0	0	1	0	1	1	1	0	0	1	1	1	0	.02
A0532	0	0	1	0	1	1	1	1	0	1	-	-	-	-
A0631	0	0	1	1	0	1	1	0	0	1	95	92	3	.50
A0632	0	0	1	1	0	1	1	0	1	0	2127	2115	12	8.75
A0634	0	0	1	1	0	1	1	1	0	0	2	2	0	.02
A0635	0	0	1	1	0	1	1	1	0	1	44	44	0	.20
A0932	0	1	0	0	1	1	1	1	0	1	-	-	-	-
A1230	0	1	1	0	0	1	1	0	0	0	2	0	2	.02
A1231	0	1	1	0	0	1	1	0	0	1	-	-	-	-
A1232	0	1	1	0	0	1	1	0	1	0	-	-	-	-
A1234	0	1	1	0	0	1	1	1	0	0	-	-	-	-
A1235	0	1	1	0	0	1	1	1	0	1	-	-	-	-
										TOTAL	3219	2931	288	19.09

NB: 0 indicates agreement;
1 indicates discrepancy between the records

FIGURE 8

DATE OF BIRTH CHECK FOR LINKAGE-CHECKING

1962 UPDATING

TO BE APPLIED TO ALL "REJECTS" AND
UNACCEPTED "POSSIBLES" FROM SOUNDEX CHECK

Reference code	Birth mo. & day; sex	Social Sec. #	Birth year	Maiden name	First name	Address	Birth year range	Decision	Computer Counts	Clerical Rejects	Net Linkages
B01	0	0						Accept	7	0	7
B02	0	1	0	0				Accept	16	0	16
B03	0	1	0	1	0			Accept	196	142	54
B04	0	1	0	1	1	0		Possible	9	1	8
B05	0	1	0	1	1	1		Reject			
B06	0	1	1	0	0			Accept	0	0	0
B07	0	1	1	0	1			Possible	30	30	0
B08	0	1	1	1	0	0	0	Accept			
B09	0	1	1	1	0	0	1	Possible	4	4	0
B10	0	1	1	1	0	1		Reject			
B11	0	1	1	1	1	0	0	Possible	38	38	0
B12	0	1	1	1	1	0	1	Reject			
B13	0	1	1	1	1	1		Reject			
								TOTAL	300	215	85

NB: 0 indicates agreement; 1 indicates discrepancy between the records

Tolerance Rules for Concordance:

Social Security Number - Complete agreement
Maiden name - Complete agreement in either maiden names or in cross-check with surname
Birth year, first name, address, birth year range - Same as in Figure 2