AN AUTOMATED PROCEDURE FOR MATCHING RECORD CHECK AND HOUSEHOLD-REPORTED HEALTH CARE DATA

Philip C. Cooley and Brenda G. Cox, Research Triangle Institute

The National Medical Care Expenditure Survey (NMCES) collected medical care data for the calendar year 1977 for the civilian, noninstitutionalized population of the United States. Data were collected in six survey rounds during 1977 and early 1978. NMCES was composed of a Household Survey, a Medical Provider Survey (MPS), and a Health Insurance/Employer Survey. Complete survey data were collected for over 38,000 persons in more than 13,000 households.

The purpose of the matching exercise that will be described in this paper was to identify the same set of visits to medical providers that were reported from two different sources: the provider of medical care, and the consumer of this care, that is, the household. The data that were common to both types of records were used as the matching parameters. These data included total charge and its components: visit date, medical provider (MID), and an indicator of the types of services performed. After the matched record pairs were determined and linked, data unique to each record type (e.g., diagnosis is unique to the provider record) were merged into a single analysis record. Matching of the two data sets was necessary to assess the level of agreement between household and provider reports of health care utilization.

The results of the matching operation can be summarized as follows. Matching the medical provider data set and the household data set resulted in reported visits of three types:

1. Matched visits. These are those visits reported in the household interview that were judged to have been matched to visits reported by the respondent's medical providers.
2. Unmatched visits reported by the household. These are the visits that were reported in the household interview to which no medical provider reported visit could be matched under the rules for matching.
3. Unmatched visits reported by medical providers. These visits are those reported by medical providers that could not be matched to household data under the matching rules.

The unmatched visits may either be visits actually reported in one survey and not in the other or they may be visits reported in both surveys, but which could not be matched using the selected matching rules.

The number of matched visits with respect to the number of unmatched visits is directly related to the strictness of the rules for matching. Requiring an exact PID (Participant Identification Number) and MID (Medical Provider Identification Number) match and an approximate match on the date of visit (± 30 days) and total charge (± minimum of $100 or 50 percent of the maximum of the two charges) would result in approximately 30 percent of the provider-reported visits matching with household-reported visits. Thus, strict matching rules would imply that the matched visits would indeed be correctly linked but that many visits reported by both the medical provider and the household would be unmatched due to discrepancies between the data supplied. Recognizing that households may not be able to accurately recall the exact date of visits and expenditures associated with each visit implies that the rules for matching must be relaxed from requiring exact matches. However, too lenient a system of matching data would result in visits being matched that are not the same.

1. Pre-Match Data Preparation

The Household Survey data used in this matching process were the data in the Summary File. For each interview round, a Summary File was created for all Household Survey participants containing the most important data variables. This file was used during data collection to produce computer-generated summaries for the participants and interviewers to verify and correct. This file also served other uses, such as a data source for the work described here [Woodside et al., 1981].

A number of edits were also implemented for the Medical Provider Survey data. The edits included checks of the person and provider identification numbers and duplicate record checks. None of these edits were specifically performed to facilitate matching, although they did improve the overall matching performance [Batts et al., 1981].

2. Preliminary Matching Activities

Initially a series of matching experiments was performed on a sample of visits using the software system referred to as UNIMATCH and also with the Statistical Analysis System (SAS) [Shah, 1978]. These early experiments demonstrated the inadequacies of SAS as a software tool in the area of matching. However, UNIMATCH exhibited some extremely desirable qualities such as efficiency, as well as great versatility and generality for specifying matching rules. Some initial matching experimentation was done using UNIMATCH but none of the experiments were evaluated because a "truth set" was unavailable for assessing matching logic performance. A standard was eventually developed. In the interim, some undesirable qualities of UNIMATCH were recognized and eventually UNIMATCH was abandoned as the matching experiment software.

The primary inadequacies of UNIMATCH that were identified were the inability of UNIMATCH to process variable block files and the inability of UNIMATCH to simultaneously split both the Summary File and the Medical Provider File into a set of matched and unmatched components. These deficiencies could have been circumvented, but the development of a FORTRAN matching program was believed to be a more appropriate alternative. Consequently a matching program was developed in FORTRAN. This program incorporated many UNIMATCH features and it corrected for the previously discussed UNIMATCH limitations. Specifically, it provided for point and interval matching capabilities, the use of tolerances, and minimum acceptance matching criteria. The FORTRAN matching program was used for all of the ensuing matching experimentation including the post-match edits.

3. Truth Set Experimental Concepts

The Truth Set experiment was initiated to provide an objective method for making comparisons between different automated (i.e., machine) matching rules. The principal assumption behind the experiment was that individuals possessing a thorough knowledge of the data collection forms (i.e., the questionnaires) and health care data in general could accurately match household and medical provider data. Consequently, a sample of surveyed individuals (PIDs) was selected and all visit data associated with each sample individual were retrieved from both household and provider sources. A set of four rules was then defined, hereafter called the Tight Rules. These rules were felt to be sufficiently restrictive to guarantee that all matches occurring as a result of these rules would be correct matches. The four rules were implemented and the matched records were segregated from the unmatched records. The residual (unmatched) set of records was then listed; this listing was given to staff of the National Center for Health Services Research who hand-matched the household and provider visit data. These hand-matched data along with the machine matches obtained using the Tight Rules were henceforth assumed to be the correctly matched visit pairs for this sample.

The linkage pointers between the hand-matched (judged to be correct) visit pairs for the sample, henceforth referred to as the Truth Set, were then keyed and subsequently used to evaluate the effect of different matching rules in the following manner. An automated rule was implemented (i.e., provider data were matched to household data) and the automated match linkages were compared against the Truth Set. The resulting automated match linkages that were also in the Truth Set were considered correct matches. Matched records with linkages not included in the Truth Set were considered mismatches. Records not matched, but which should have been according to the Truth Set, were defined as incorrect non-matches. Records that matched neither under the automated rule nor in the Truth Set were defined as correct non-matches. Finally match-mismatch statistics were computed and the rule was assessed.

4. Truth Set Implementation

The sample of individuals used to establish the Truth Set was selected in the following manner. A file consisting of all possible person/providers pairs occurring in the Medical Provider Survey was systematically sampled. This file contained 24,254 entries and was used by the NMCES automated control system for various activities, such as addressing labels used for MPS mailouts. The sampling rate was one in 60 beginning at a randomly selected record. All of the visit data associated with the 400 sample individuals selected by this procedure were retrieved from the Medical Provider Survey and household Summary File data bases. There were 3,123 visit records from the Medical Provider data and 4,862 visit records from the Household Survey data associated with these individuals.

Nine hundred fifty-two (952) records in each of the two files were computer-matched using the Tight Rules (referred to above) that were felt to be sufficiently restrictive to guarantee true matches. These 952 records were matched in three separate steps. The Tight Rules are summarized in Table 1. The TYPE variable used in these tables differentiates between hospital stay visits (TYPE=2), doctor in-hospital visits (TYPE=6), and medical provider visits (TYPE=3). TYPE was used to force the record matching to occur within the three visit type categories.

The 952 records matched using the Tight Rules were removed from the original sample visit records from each of the two files. The remaining records, consisting of the difficult to match records, were then hand-matched by NCHSR staff. The key variables were listed for each file to aid in the hand-matching operation. NCHSR staff identified the sequence number of the matching

Table 1. Summary of Tight Matching Rules

Step	Type[a]	PID	MID	Date	Charge	Other
1	3	Exact	Exact	±1 day	±min ($25, 20%)	Telephone[b]
2	3	Exact	Exact	±1 day	Missing	None
3	2	Exact	Exact	Exact in interval[c]	None	None

[a]Type 2 = hospital stay visits; type 3 = medical provider visits; type 6 = doctor in-hospital visits.

[b]A Summary File record with a non-zero charge and an indicator that a telephone call was made could only be matched with a fee for service record that explicitly identified a telephone call as part of the service rendered.

[c]A Summary File Type 2 record could only match with an Inpatient provider record if the household visit date fell in the interval defined by the hospital admission and discharge dates.

Table 2. Truth Set Statistics

	MPS	Summary
PIDs	400	400
Visits	3,123	4,862
Tightmatched Records	952	952
Unmatched Records	2,171	3,910
Type 2	68	132
Type 3	1,935	3,395
Type 6	168	420
Truth Set Records	2,171	3,910
Duplicates[a]	90	-
Type 2	5	-
Type 3	84	-
Type 6	1	-
Non-Matches[a]	798	-
Type 2	6	-
Type 3	767	-
Type 6	25	-
Possible Matches[a]	1,283	-
Type 2	57	-
Type 3	1,084	-
Type 6	142	-

[a]Statistics not computed for Summary File.

record in each file. The sequence number of the Summary File record that was judged to be the match to each provider record, or a code indicating that no match existed, was included in a machine-readable record file. This number also became part of the provider residual file (those records difficult to match) so that correct matches could be ascertained and overall match rate statistics computed during the development of specific matching rules.

The composition of the Truth Set visit sample is illustrated in Table 2. Note that some records were judged by NCHSR staff to be duplicates of others and that some MPS records were judged to have no matching counterparts in the Summary File.

5. Development of Matching Rules

Using a variety of matching rules, an investigation of the matching errors was initiated (see Radner et al., 1980 for a comprehensive discussion of this problem). The following terminology is introduced to facilitate this discussion. A given set of rules can be implemented in a different order with different effects on overall performance being observed. Therefore, a "pass" is a specific rule that has been implemented and is preceded by a number of other passes. A matching "scheme" is defined as a collection of passes that terminates as a result of some specification. The stopping criterion used for most schemes was that the final pass was the last pass that produced more correct matches (according to the Truth Set) than mismatches.

The first matching scheme that was implemented was based on suggestions presented by NCHSR staff members and was applied to Type 3 records (medical provider visits), which were judged by NCHSR to be more difficult to match than the other record types. The matching scheme consisted of nine passes. Of 1,026 potentially valid matches, 843 occurred. However, only 549 of the 843 were judged to be correct by the Truth Set; there were 294 mismatches. It was evident that a significant number of mismatches occurred during the early passes of this scheme. Therefore additional schemes were developed in an attempt to cut down on the early occurrence of mismatches, specifically by relaxing the charge and date variables more slowly than in Scheme 1. The results indicated improved performance compared with Scheme 1 [Cooley, 1980].

The performance of the last matching scheme is presented in Table 3. All schemes that were investigated progressed until all additional passes produced more mismatches than matches, but the last scheme displayed the best performance characteristics. Refinements to this scheme that improved performance beyond the thirteenth pass could not be found. Therefore, this scheme proved to be the basis for the production matching activities.

Similar experiments were performed for Type 2 (hospital stay visits) Type 6 (doctor in-hospital visits) and records. The results for Type 6 and Type 2 records are displayed in Tables 4 and 5, respectively. Note that the tables examine the records matched, and indicate by pass the number of correct matches and the number of mismatches (incorrect matches). The second part of each of Tables 3 thru 5 examines the matched records as well as those records not matched (non-matches).

Table 3. Type 3 Records
Final Matching Scheme

Pass	Total Matches	Correct Matches	Mismatches	Tolerances[a,b,c,d] MID	Date	Charge
1	154	150	4	Y	0	50%
2	144	136	8	Y	0	--
3	77	70	7	Y	2	--
4	84	73	11	Y	14	25%
5	71	55	16	Y	7	--
6	105	92	13	N	0	50%
7	30	25	5	Y	31	25%
8	96	71	25	N	0	Missing
9	47	33	14	Y	14	--
10	39	22	17	Y	±1 month[e]	25%
11	31	18	13	Y	24	--
12	17	9	8	Y	61	50%
13	24	13	11	N	3	50%
Total	919	767	152			

Match-Mismatch Statistics

	Matched	Did Not Match	Total
Truth Set			
Matched	767[f]	152[g]	919
Did Not Match	317[h]	699[i]	1,016
Total	1,084	851	1,935

Correct Matches and Nonmatches: 1,466 = 767 + 699
Total Match Rate: 76% = (1,466/1,935) * 100
Correct Matches: 767
Correct Match Rate: 71% = (767/1,084) * 100

[a] Y = MID rule was employed in the matching pass. N = exact MID match was not used.
[b] The Date and Charge columns identify tolerances allowed in the appropriate MPS parameters. For example, a 14 under Date means a match occurred if the household visit date was within 14 days of the MPS visit date. The -- means that rule was not used for that pass.
[c] The minimum of $100 or 50% of the larger reported charge (MPS or household) was used to
[d] define the matching interval.
Charge value was required to be a legitimate "missing" code.
[e] Records with missing days could be accepted as matching candidates.
[f] Correct matches.
[g] Mismatches.
[h] Incorrect non-matches.
[i] Correct non-matches.

These tables indicate the number of correct matches, mismatches, and non-matches. For example, in Table 3, out of 919 MPS records that were matched, 767 were judged to be correct matches and the remaining 152 were mismatches. Of the 1,016 MPS records that were not matched, 699 of these non-matches had no partner in the household data and were judged to be correct non-matches. The remaining 317 records were non-matches that should have matched, or incorrect non-matches. From these data two match rates were calculated.

Table 4. Type 6 Records
Final Matching Scheme

Pass	Total Matches	Correct Matches	Mis-matches	Tolerances[a,b,c,d] MID	Date	Charge
1	43	42	1	Y	7	50%
2	19	18	1	Y	31	50%
3	14	13	1	Y	3	--
4	13	12	1	Y	7	--
5	11	8	3	N	31	25%
6	12	9	3	Y	--	25%
7	14	10	4	Y	14	--
Total	126	112	14			

Match-Mismatch Statistics

	Matched	Did Not Match	Total
Truth Set			
Matched	112[e]	14[f]	126
Did Not Match	30[g]	12[h]	42
Total	142	26	168

Correct Matches and Nonmatches: 124 = 112 + 12
Total Match Rate: 74% = (124/168) * 12
Correct Matches: 112
Correct Match Rate: 79% = (112/142) * 100

[a] Y = MID rule was employed in the matching pass.
[b] N = exact MID match was not used.
The Date and Charge columns identify tolerances allowed in the appropriate MPS parameters. For example, a 14 under Date means a match occurred if the household visit date was within 14 days of the MPS visit date. The -- means that rule was not used for that pass.
[c] The minimum of $100 or 50% of the larger reported charge (MPS or household) was used to define the matching interval.
[d] Charge value was required to be a legitimate "missing" code.
[e] Correct matches.
[f] Mismatches.
[g] Incorrect non-matches.
[h] Correct non-matches.

The first calculates the percent of correct matches and correct non-matches out of the total sample. The second match rate is based on only those records that should have matched. With respect to the data in Table 3, the values for the two match-mismatch rates are 76 percent and 71 percent respectively. Note that the correct match rate is based on only the set of matched records; the total match rate is based on all records. The former indicates how well the applied rules performed; the latter rate indicates how far the applied rules were carried out.

The performance of the total matching effort on all Truth Set records including the Tight Rules is presented in Table 6. Table 6 indicates that if the Truth Set sample is representative of the visit set in the total MPS-household NMCES sample data, a match rate exceeding 80 percent is expected.

Table 5. Type 2 Records
Final Matching Scheme

Pass	Total Matches	Correct Matches	Mis-matches	Tolerances[a,b,c,d] MID	Date	Charge
1	26	24	2	Y	31	--
2	16	15	1	N	7	--
3	12	9	3	Y	--	--
Total	54	48	6			

Match-Mismatch Statistics

	Matched	Did Not Match	Total
Truth Set			
Matched	48[e]	6[f]	54
Did Not Match	9[g]	5[h]	14
Total	57	11	68

Correct Matches and Nonmatches: 53 = 48 + 5
Total Match Rate: 78% = (53/68) * 100
Correct Matches: 48
Correct Match Rate: 84% = (48/57) * 100

[a] Y = MID rule was employed in the matching pass.
[b] N = exact MID match was not used.
The Date and Charge columns identify tolerances allowed in the appropriate MPS parameters. For example, a 14 under Date means a match occurred if the household visit date was within 14 days of the MPS visit date. The -- means that rule was not used for that pass.
[c] The minimum of $100 or 50% of the larger reported charge (MPS or household) was used to define the matching interval.
[d] Charge value was required to be a legitimate "missing" code.
[e] Correct matches.
[f] Mismatches.
[g] Incorrect non-matches.
[h] Correct non-matches.

Table 6. Match-Mismatch Statistics, All Schemes and Tight Rules

	Matched	Did Not Match	Total
Truth Set			
Matched	1,879[a]	172[b]	2,051
Did Not Match	356[c]	716[d]	1,072
Total	2,235	888	3,123

Correct Matches and Nonmatches: 2,595 = 1,879 + 716
Total Match Rate: 83% = (2,595/3,123) * 100
Correct Matches: 1,879
Correct Match Rate: 84% = (1,879/2,235) * 100

[a] Correct matches.
[b] Mismatches.
[c] Incorrect non-matches.
[d] Correct non-matches.

6. Production Matching Results

The final schemes discussed above suggested that correct match rates could exceed 80 percent of all matched records if these schemes were adopted for production purposes. Furthermore, for reasons enumerated later (Section 8), the 80 percent figure was believed to understate the actual match rate. Additional experimentation did not demonstrate a superior scheme. Consequently, a variant of these schemes was implemented.

The salient feature of the Production Scheme was that the three record types were matched simultaneously using the scheme designed for the Type 3 records. The relative difficulty of matching Type 3 records was witnessed by the progressively slower relaxation of the matching rules to achieve the same match rate that the other record types experienced. By applying the Type 3 Scheme to the Type 2 and Type 6 records, improved matching occurred. However, provisions were made for the final few passes for the Type 2 and 6 records. These passes were characterized by matching rules that appeared to be record-type specific and inappropriate for Type 3 records. The resulting Production Scheme is illustrated in Table 7. In this scheme, the first three passes are the first three of the Tight Rules. Passes 4 through 16 correspond to the thirteen passes of the Type 3 final scheme. The final passes 17 through 19 represent rules specific to Type 2 and Type 6 records and were not applied to Type 3 records.

Note that the Summary File record count reflects the full sample of household records and not just those that appear in the MPS sample. Thus about one-half of the Summary File record have PIDs that do not occur in the MPS sample and cannot be matched. Note also that out of 51,650 total MPS records, 34,949 or 68 percent of the total, were matched during the production runs. This should be compared with the Truth Set records, where 60 percent of the total was machine-matched. This lower figure probably occurred because of the accumulation of the minor problems that developed during Truth Set construction and the additional matching rule refinements that were implemented for the production matches. This difference provides additional evidence that the match rate results displayed in Tables 3 through 5 understate the corresponding production match results.

Upon completion of the 19 passes defined in Table 7, the residual Type 2 and Type 6 records were listed and hand-matched by NCHSR staff. This hand-matching was then implemented on the residual 19 pass outputs, affecting 464 records. The 464 hand-matched records consisted of 99 Type 2 and 365 Type 6 records.

During the Truth Set experimentation, a relatively large number of duplicate records were observed. These duplicate records were not resolved by the duplicate edit checks occurring prior to matching. This prompted two proposals that attempted to resolve duplicates that occurred in the matched MPS file and the unmatched (residual) MPS file.

Briefly, the first post-match edit consisted of identifying records in the MPS matched record set (35,413 records) and the MPS residual file (16,237 records) with equal PID, DATE, and CHARGE

Table 7. Production Matching Statistics

Pass No.	Matches	Remaining MPS	Remaining SUMMARY	Rules[a,b,c,d] Type[e]	MID	Date	Charge
0	0	51,650	198,898	--	--	--	--
1	9,992	41,658	188,906	A	Y	1	Min (20%, $25)
2	5,105	36,553	183,801	A	Y	1	Missing
3	1,354	35,199	182,447	2	Y	int.[f]	--
4	3,092	32,107	179,355	A	Y	0	50%
5	1,834	30,273	177,521	A	Y	0	--
6	1,644	28,629	175,877	A	Y	2	--
7	2,149	26,480	173,728	A	Y	14	25%
8	1,810	24,670	171,918	A	Y	7	--
9	1,700	22,970	170,218	A	Y	0	50%
10	745	22,225	169,473	A	Y	31	25%
11	1,735	20,490	167,738	A	N	0	Missing
12	982	19,508	166,756	A	Y	14	--
13	612	18,896	166,144	A	Y	m±1[g]	25%
14	886	18,010	165,258	A	Y	24	--
15	334	17,676	164,924	A	Y	61	50%
16	360	17,316	164,564	A	N	3	50%
17	174	17,142	164,390	2	N	7	--
18	125	17,017	164,265	2	Y	--	--
19	316	16,701	163,949	6	N	31	25%

[a] Y = MID rule was employed in the matching pass. N = exact MID match was not used.

[b] The Date and Charge columns identify tolerances allowed in the appropriate MPS parameters. For example, a 14 under Date means a match occurred if the household visit date was within 14 days of the MPS visit date. The -- means that rule was not used for that pass.

[c] The minimum of $100 or 50% of the larger reported charge (MPS or household) was used to define the matching interval.

[d] Charge value was required to be a legitimate "missing" code.

[e] A implies rule applies to all record types.

[f] A Summary File Type 2 record could only match with an Inpatient provider record if the household visit date fell in the interval defined by the hospital admission and discharge dates.

[g] Records with missing days could be accepted as matching candidates.

variables. For each of the pairs of matched records (556 in number) in the MPS match and residual files, those records possessing more information (i.e., the nonreject, nonpartial records) were identified and moved if necessary into the matched file (17 records were moved). The other records were removed from the residual file and retained as a separate file.

The second post-match edit step identified records from the matched and residual MPS files with equal PID and DATE variables. (There were 1,432 records in each file.) Both files were retained as separate files and removed from match and residual files. The authors' assumption was that only an unknown portion of the 1,432 records identified by this process represented duplicates. By isolating these records, further investigations could be undertaken to determine whether the records were duplicates, and appropriate action could be taken.

Table 8. Final Counts by Match Category
Match Category Code

Match Category Code	Number
01 = Matched	33,981
02 = Matched (originally category 01) but put into a separate category because other unmatched records were determined either to be duplicates or to contain additional data for the same visits.	1,432
03 = Not matched, but determined during post-match edits to be very similar to a duplicate of records that had matched category 03.	1,432
04 = Not matched, duplicates of category 01 records.	556
05 = Not matched	14,249

The set of all matched and unmatched (residual) visit records can be divided into five categories. These five, identified in Table 8, are:

01. The 35,413 less 1,432 MPS match records (i.e. 33,981 matched records),
02. The 1,432 post-match MPS match records (i.e. potential duplicates in the set of match records),
03. The 1,432 second post-match MPS residual records (also potential duplicates in the set of unmatched records),
04. The 556 first post-match MPS residual records, and
05. The remaining 16,237 residual records less the 556 first post-match residual records (i.e. 04) and less the 1,432 second post-match MPS residual records (i.e. 03), that is, a file of 14,249 records.

7. The Matching Stopping Criterion and Its Effect on Unreported Visits

The stopping criterion used in the final schemes during the Truth Set experiment was to terminate when the mismatch rate for a given pass exceeded the match rate for all subsequent matching rules. The effect of this stopping criterion was to match fewer records than were matched by hand in the Truth Set; that is, the termination procedure understated the true number of matched visits. Specifically, 2,235 records out of 3,123 (72 percent) were matched by hand and 2,051 records (66 percent) were matched by machine (see Table 6). To estimate the number of unreported visits, it was necessary to assume that the Truth Set was a representative sample with respect to the application of the stopping criterion. Hence out of a total sample of 3,123 records, 184 fewer records were matched by machine than by hand (5.89 percent).

Generalizing this result to the complete MPS and Summary File sample implies that there are 3,042 unreported visits (i.e., MPS visits that should not have been matched to a household visit) out of the 51,650 records matched through the 19 matching passes. Adjusting for the 464 (assumed totally correct) hand-matched Type 2 and Type 6 records implies that 2,578 visits (3,042 - 464) were not reported by the matching process, but should have been. The post-match edits did not alter this estimate, since the calculations assumed the existence of duplicates throughout. In summary, the automated scheme under-reports visits. The authors' best estimate of the number of unreported visits is 5.89 percent of the total number of records minus the number of hand matches, i.e., 2,578 visit records.

8. Additional Sources of Bias Affecting the Estimated Match Error Rate

The following sources can be identified that bias the match error rate estimates when applied to the complete MPS sample:

- The large number of duplicates in the sample that were partially resolved through the post-match edits.
- The error that occurred in the implementation of one of the Tight Rules in the Truth Set, which incorrectly matched 14 records.
- The implementation of final schemes for Type 2 and Type 6 records in the production run using tighter rules than originally specified in these schemes.
- The calculation of error rates assuming unweighted sampling units.
- The assumption that the Tight Rules matched records with 100 percent accuracy.

The first four sources cause the estimate to be understated. The last source overstates the estimate. Additional follow-up efforts should be considered to precisely measure the overall match rate and unreported visits of the Production Scheme.

9. Acknowledgments

The research in this paper was done for the National Center for Health Services Research (NCHSR) under contract No. HRA-230-76-0268. The views expressed in this paper are those of the authors and no official endorsement by NCHSR is intended or should be inferred.

10. References

Batts, James R. and Linda Nixon (1981). NMCES Medical Provider Survey File Construction. National Center for Health Services Research, Federal Center Building No. 2, Room 8059B No. 4, 3700 East-West Highway, Hyattsville, MD 20782.

Cooley, Philip (1980). NMCES Matching of MPS and Household Summary Data Methodology Report. National Center for Health Services Research, Federal Center Building No. 2, Room 8059B No. 4, 3700 East-West Highway, Hyattsville, MD 20782.

Radner, Daniel B., Rich Allen, Maria E. Gonzalez, Thomas B. Jabine, and Hans J. Muller (1980). Report or Exact and Statistical Matching Techniques. U.S. Department of Commerce.

Shah, Babu (1978). Strategy for Development of a Matching Procedure for Provider Check Data For NMCES. National Center for Health Services Research, Federal Center Building No. 2, Room 8059B No. 4, 3700 East-West Highway, Hyattsville, MD 20782.

Woodside, M. Beebe and Barbara A. Moser (1981). NMCES Household Survey Computer Data Processing Operations Methodology Report. National Center for Health Services Research, Federal Center Building No. 2, Room 8059B No. 4, 3700 East-West Highway, Hyattsville, MD 20782.