

QUALITY ASSURANCE FOR THE CLERICAL MATCHING OPERATION IN THE 1990 POST-ENUMERATION SURVEY

Kirsten West, Carol Corby, and Gary Van Nest*, U.S. Bureau of the Census
Washington, DC 20233

Introduction

This paper describes the quality assurance plan for clerical matching of the 1990 Post-Enumeration Survey (PES). Because of the complexity of the matching operation, and the need for highly accurate output from it, an intensive quality control plan is of importance.

Formal quality assurance procedures were implemented for the clerical matching between the PES and the Census for the first time during the 1988 Dress Rehearsal. The plan involved the independent match of a work unit by two groups of matching clerks. The operation was aided by a computer system that allowed the matching clerks to key their match codes directly into a data base. The system checked their codes for validity, consistency and completeness, thus preventing many minor errors.

The computerized system as well as the specific steps of the quality assurance of the clerical matching process are outlined below. Experiences from the 1988 Dress Rehearsal are discussed along with recommendations for improvements to the system.

The Post-Enumeration Survey

The Post-Enumeration Survey is designed to evaluate the coverage of the decennial census of housing and population characteristics. Coverage error is the error in the count of persons or housing units resulting from cases missed during census enumeration or counted erroneously either through duplication or erroneous inclusion in a wrong geographical area.

The evaluation of the census is based on two samples. The first is an independently selected probability sample of the target census population. This sample is the population or P-sample. The other sample is the enumeration or E-sample. It is a sample of persons in housing units enumerated in the census. It is drawn directly from the census for the same areas as the P-sample.

The PES interview collects names, characteristics, and census day addresses for all residents in the P-sample. Also, names and characteristics for persons who moved from the sample address between the census and the PES interview are obtained. A person-by-person match of these data to the census data determines the "coverage status" in the census. The results from the sample cases are weighted-up to estimate coverage for the entire target population, using dual-system estimation technique.

Computer Matching

The PES and census data are initially matched by computer. All persons reported in the PES interview in a sample block are computer matched to the persons enumerated in the census in the PES block and its surrounding blocks.

After the computer matching, the results are printed out on computer match forms (CMF) and clerically reviewed. Each household goes on a separate form. Matched and possibly matched PES and census persons are printed side by side on the same line. Non-matched persons are printed on separate lines on the form.

If the entire household contains no matches or possible matches, all PES household members are printed on one match form with the census side blank. The census household members are printed on another match form with the PES side blank.

Clerical Matching

The clerical matchers are trained to review the cases which have been assigned a status of possibly matched or non-matched by the computer matcher. Many cases not matched by the computer are resolved quickly by the clerical matchers because they can use the information from other household members to resolve the match status of an individual person. (The computer matches by person, not household.) Also, the clerks can easily resolve the match status for cases where the computer failed because of keying errors. The clerks have access to more information than the computer, including original handwritten answers plus special notes made by the interviewers.

To assist in the clerical search for matching persons or households in surrounding blocks, listings of persons (by surname) as well as listings by household (in street address order) are available to the clerks.

Matches

Cases that are given a status of matched by the computer are not targeted for formal clerical review, because the computer matching error rate has been shown to be small in previous tests. However, a separate study is done to ensure that no problems occur during the computer matching.

In partially matched households, where erroneous computer matches are most likely to be a problem, the matches may be reviewed as

part of the overall process of matching the remaining members of the household. If a false match status should be discovered by the clerks, they are instructed to correct the error.

Possible matches

Using the matching guidelines, the clerks review the paired cases assigned a status of possibly matched by the computer matcher. The PES and census data are printed on the match form. The clerical review may result in a conversion to match, a conversion to a non-match or a decision to maintain the possible match status. All possibly matched cases are sent to follow-up interviewing for more information. Any additional matches or possible matches discovered by the clerks are transcribed to the match forms.

Non-matches

The clerks review all the non-matched cases. The clerks can code a pair as matched when the differences fall within their matching guidelines, i.e., the discrepancies are minor and explainable. For example, a person may not have been matched by the computer because he/she was listed with an incorrect name. A PES interviewer may have recorded a surname that does not correspond with the census surname. When the clerk reviews the case it may be obvious from other information that one of the surnames is a maiden name. The non-match status can be converted to a match status.

If PES and census persons in the non-matched households are linked together as matched or possibly matched, the census household data and person data are transcribed to the computer match form with the PES household by the clerical matching clerks.

Other matching operations

In addition to possible and non-match cases, the clerical review also involves cases that cannot go through the computer matcher. For example, the census day addresses for movers recorded on the PES form are matched to the census. To match the persons who lived at a different address on census day, the address is geocoded, and the questionnaires in the census block and surrounding blocks are examined. The computer matcher cannot be used because the census names are not keyed for the entire site. In other words, the names are not available to the computer matcher for addresses outside the P-sample blocks.

The clerical matching is also extended to matching late census data. There are different reasons why census data may not be available when the computer matching operation starts. Some census data are obtained after the

computer matching has begun for a District Office. Some census operations do not end until after the computer matching has started, and some census questionnaires may not be data captured in time to be included.

The quality assurance plan

A proposal for formal quality assurance procedures was developed for the clerical matching in the 1988 Dress Rehearsal. The proposal involved inspection of each matching clerk's work. The inspection was to be accomplished through an independent rematch of the cluster by another matching clerk, next, by a comparison of the match codes, and finally, by adjudication of differences between the codes by a matching expert. In order to minimize the extra clerical work required by the quality control (QC) itself, and thus to minimize the opportunities for errors in the inspection process, the comparison between the two sets of match codes was to be completed by computer.

The plan was implemented as follows: A block cluster was assigned to a clerical matching group. This matching group consisted of two clerks: a clerical matching group clerk (CMG) and a special matching group clerk (SMG-1). In this group, the CMG clerk matched the cluster according to strict guidelines. The SMG-1 clerk then verified the CMG clerk's work, making this a dependent operation. Next, the cluster was assigned to another special matching group clerk (SMG-2), who independently matched the cluster. The SMG-2 clerk matched the records from the beginning like the CMG clerk, but used the latitude and flexibility in matching available to the special matching group. Each set of final codes were compared and differences were adjudicated by the PES technicians.¹ All clusters were double matched in this way.

The matching operation was aided by the computer system. This system allowed the clerks to enter their match codes into a data base as soon as they completed each cluster. The system checked their codes for validity, consistency and completeness as the codes were entered. After all three clerks' code had been entered, the system compared the codes and printed lists of differences for adjudication by the PES technicians. The final adjudicated codes were entered by the technicians and checked for consistency and completeness.

At the completion of the clerical matching for a cluster, a file was produced. This file contained all the match codes used by the clerks, and a summary of the errors made by the clerks. The error summary was forwarded to the supervisory clerk who then provided feedback or retrained the clerks as appropriate.

The steps outlined above were carried out for nonmover matching of block clusters before

follow-up. This includes within block matching, duplicate search, and surrounding block search. These quality control steps were also carried out for nonmover after follow-up matching activities.

Description of the software system

A menu driven software system performs the tasks outlined above. The software for the QC includes a database file containing the person records that were printed onto the CMF. Each person line on the CMF corresponds to one record in the file. On the CMF, when a clerk assigns a match code to a person, the code is recorded next to the computer generated match code. In other words, a code is entered in the appropriate column on the row for the person. At times, the data for a census person must be copied from one row to another and the codes must be recorded on a new line. An example of a key-in menu allowing the clerks to perform these and similar matching steps is presented in Figure 1.

Figure 1. SMG AFTER FOLLOW-UP
'KEY-IN' MENU

Do you want to:

- 1 - key in match code on regular record
- 2 - split a regular record
- 3 - merge two records
- 4 - key in match code for half record
- 5 - key in data for added census person
- 6 - undo a regular match code
- 7 - undo a split record
- 8 - undo a merged record
- 9 - undo a half record match code
- 10 - undo a census data add
- 11 - change code on already coded record
- 12 - change outcome code for PES household
- 13 - look at any record
- 14 - look at all records in a household
- 15 - check for uncoded records (edit and close-out)
- 99 - exit to main menu

Please enter the option here : _____

The keying activities are accomplished as follows:

1. From the 'keyin' menu, type of record is selected.
2. The computer prompts for the record number of the person to be coded.
3. The record number is entered.
4. The record is displayed on the screen as it appears on the CMF.
5. The clerk visually verifies that the correct record has been found, then enters the match code and returns to the 'key-in' menu.

As stated previously, in addition to providing the menu driven key-in programs used by the CMG and SMG matching clerks and the PES technicians, the program performs invalid and missing code checks. The program also compares the two sets of match codes entered by the SMG clerks for the PES technician review.

Requirements for the system

Equipment: The quality assurance system requires that terminals and line printers are available. In 1988, fourteen VAX computer terminals and one low speed printer were available for the clerks.

Staff: Unlike most QC operations, which have a QC staff separate from the regular production staff, this quality assurance plan calls for one large SMG staff to perform the two matching operations. All matching clerks are given the same training and instructions. This ensures that everyone works with the same procedures and applies them in a consistent manner. With no separate QC staff, bottlenecks are avoided. As a result, the processing flow is smooth.

The system is designed to be user friendly. In 1988, hands on terminal training was accomplished by one hour training sessions for groups of six clerks. In total, forty-seven CMG and SMG clerks and five control clerks received terminal training.

Terminal time: Keying for each matching activity is expected to take from 15 minutes to an hour for each block cluster depending on the size of the cluster and the number of matches already accomplished by the computer matcher. On the average, approximately one hour of keying is expected for every eight hours of matching time spent by each clerk.

Results from the 1988 Dress Rehearsal

The quality assurance system detects transcription errors, errors occurring during merging or splitting of records, (matching or unmatching of persons) incorrect assignment of match codes and misinterpretation of matching situations. Specifically, the following types of errors are recorded:

1. **Keying error.** Edit checks are performed to discover invalid match codes, i.e., typing errors within the range of acceptable match codes.
2. **Incorrect match code.** The right person records may be paired, but the match code is unacceptable. For example, the choice of P (possible match) or M (match) for matched pairs may be wrong.
3. **Incorrect pairing or separation of persons.** Persons who do not match might have been paired by the clerk. For example, a father on the census side may have been paired with a son on the PES side, or a

brother on the census side may have been paired with a different brother on the PES side. In other words, different children within the same household have inadvertently been mis-matched.

4. **Erroneous assignment of duplication code.** Erroneously finding duplicates in surrounding block or failing to find duplicates in surrounding block.
5. **Erroneous change to status code.** The PES person status as a mover or a nonmover is erroneously changed.
6. **Erroneous change to PES interview outcome code.** For example, erroneously changing from a complete interview with a household member to a complete interview with a proxy.
7. **Incorrect assignment of address match.** Erroneous matching of addresses, or failing to match two addresses.
8. **Missed or erroneously added census person.** Erroneous matching to a census person from a surrounding block, or missing a match that could have been made with a person from a surrounding block.
9. **Incorrect assignment of geocodes.** For example, failure to determine correct reason for nonmatched based on geocode.

The results obtained in the 1988 Dress Rehearsal are shown in Tables 1, 2 and 3. These results indicate how frequently each of the different types of error occurred in the before and after follow-up matching operations. Table 1 represents Columbia, Table 2 St. Louis and Table 3 Washington State. Also, when reviewing the results, it should be kept in mind that the counts represent the number of errors made by both groups of clerks (recall that each cluster underwent two matching steps: SMG-1 and SMG-2).

Table 1. Number of Clerical Errors by Type of Error and Clerical Stage. 1988 PES Dress Rehearsal - Columbia.

<u>Type of Error</u>	<u>Clerical Stage</u>			
	Before		After	
	Followup	Followup	Followup	Followup
	SMG 1	SMG 2	SMG 1	SMG 2
Keying error	6	2	3	1
Wrong match code	114	85	92	71
Split or merge error	17	24	0	0
Dup code error	19	36	0	1
Status code error	55	5	9	13
Outcome code error	1	6	0	0
Address match	133	343	3	0
Census add error	0	0	0	0
Geocode error	--	--	36	64

Total Records in Before Follow-Up Matching:
11,208

Total Records in After Follow-Up Matching:
1,689

Table 2. Number of Clerical Errors by Type of Error and Clerical Stage. 1988 PES Dress Rehearsal - St. Louis.

<u>Type of Error</u>	<u>Clerical Stage</u>			
	Before		After	
	Followup	Followup	Followup	Followup
	SMG 1	SMG 2	SMG 1	SMG 2
Keying error	14	18	5	12
Wrong match code	262	214	276	150
Split or merge error	5	16	2	0
Dup code error	29	27	0	1
Status code error	81	83	39	18
Outcome code error	5	5	6	6
Address match	328	343	3	3
Census add error	3	7	0	0
Geocode error	--	--	41	26

Total Records in Before Follow-Up Matching:
17,228

Total Records in After Follow-Up Matching:
3,201

Table 3. Number of Clerical Errors by Type of Error and Clerical Stage. 1988 PES Dress Rehearsal - Washington State.

<u>Type of Error</u>	<u>Clerical Stage</u>			
	Before		After	
	followup	followup	followup	followup
	SMG 1	SMG 2	SMG 1	SMG 2
Keying error	0	0	2	3
Wrong match code	69	5	15	19
Split and merge error	2	0	0	0
Dup code error	4	1	0	0
Status code error	92	78	3	4
Outcome code error	0	0	0	0
Address match	136	202	0	0
Census add error	1	0	0	0
Geocode error	--	--	61	20

Total Records in Before Follow-Up Matching:
3,682

Total Records in After Follow-Up Matching:
585

The frequency by which the errors occurred are consistent across the three sites. Overall, considering the number of records that were handled, few clerical matching errors were made. The lowest occurrence of error is found in the census add category - erroneous matching of a census person from a surrounding block or missing a match that could have been made with a person from a surrounding block. Few keying errors were made. Similarly, the splitting and merging of records did not result in many errors. There were few erroneous assignments of duplication codes or erroneous changes to the PES interview outcome code.

When a match code was changed from, for example, an N (non-match) to an M (match) or from a P (possible match) to an M (match) it would be counted as an error. Most of the errors recorded in the wrong match code category represent such occurrences. Address match errors primarily involve failure to distinguish between eight different non-match codes: N1, N2, N3 and N4 match codes on the PES side and E1, E2, E3 and E4 match codes on the census sides. Most geocode errors can be attributed to absence of a code.

During the adjudication process the PES technicians corrected all the errors made by the clerks and detected by the comparison. The QC was not limited to editing of entered data. Checks were also made for incorrect applications of procedure.

Recommendations for 1990

As previously stated, the computer assisted quality assurance system was implemented for the first time during the 1988 Dress Rehearsal. Based on the experience during this operation several observations can be made with regard to the system itself and its implementations in 1990:

1. Random error - independent review. In 1988, it was the assumption that the system should deal with random rather than systematic errors. Time and practicality concerns guided this decision. Finding random errors only requires independent rematch. The tracing of systematic errors, on the other hand, requires expert review of matching results, i.e., a dependent review. It requires that a substantial number of experts are available to perform the review. Thus, the independent review was selected as the best approach. It is the recommended approach for 1990.

2. The scope of the quality assurance system. In the Dress Rehearsal, it was the assumption that the mandate of the quality assurance system was broad. The system should not be limited to catching only errors that might affect the dual system estimation. Rather, the system was designed so that any type of mistake could be caught and immediately corrected. While patterns of errors made by individual clerks failed to emerge, the system on several occasions revealed the need for further elaboration of procedures. For example, address matching appeared to cause some misunderstandings. (When to use N2, N3 or N4). When this problem was discovered, corrective training was provided to the entire SMG staff.

For 1990, it is recommended that the same 'catch all' approach be adopted. The capability of the system to produce a clean file for a cluster proved valuable in 1988. Such files can be available as soon as the data are finished with QC, because the errors are corrected during adjudication.²

3. Special case follow-up flags. In 1990, the menu item setting follow-up flags should be utilized. This will expedite the processing of cases that need to go through special follow-up interviewing.

4. Assignment control. In 1990, the system should also function as an assignment control system. The software package can provide a menu driven program to be used by the assignment clerks and the supervisory clerks. The program can keep track of assignments and queues for the assignment clerks. It can provide a record of each clerk's QC history to determine when inspection is required.

5. Staffing requirements. It is recommended that additional PES technicians be hired for 1990. In 1988, two technicians were available. A predominant amount of the technician-time was devoted to the adjudication process. If the number of technicians were increased to four or five per processing office, one person could be in charge of maintaining the data base, one or two persons could assist the matching clerks, and the remaining two persons could concentrate their efforts on the adjudication process.

6. Print-out capability. To help the technicians with adjudication it is recommended that more streamlined print-outs of differences in clerical codes be produced.

7. New menu items. Menu items should be created to produce print-outs that would facilitate limited dependent reviews by experts. These print-outs could indicate where systematic errors may be occurring and pinpoint which clusters are in need of further review. The system should generate summary results of, for example:

- a. match codes
- b. added persons in surrounding blocks
- c. late census data - how many, and where
- d. counts of persons by mover/non-mover status code
- e. non-matches with various demographic characteristics. The results should be presented by District Office or other census geography

Endnotes

1. The position PES technician was created in 1988. PES techs are responsible for answering questions to the clerks. They are given two weeks of intensive training before matching begins. The training includes the theory of coverage measurement surveys and the reasoning behind the matching concepts.

2. By methodological design, if the SMG-1 level and the SMG-2 level clerks make identical errors, the mistake might go undetected.

* This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.