MATCHING FOR CENSUS COVERAGE CHECKS

Walter M. Perkins and Charles D. Jones, Bureau of the Census

Perhaps it is true that "a rose is a rose is a rose" - this is a little outside our field. We do, however, disagree with anyone who makes the same claim for a match. It has been our job to try to answer the same question for many different individuals, "Was this person counted in the 1960 Census?" In this process, we have frequently met the creature called a "match" - and sometimes its opposite the "nonmatch". These are no problem - it is the gradations and variants that are troublesome. For example, we do find a creature that can only be described as "probably a match - but maybe not" and that other category "seemingly not a match - but possibly is." Then there is that rather mysterious category - "the impute;" the census interviewer has been told that a given address is occupied, but has been unable to find anyone home on repeated visits. The electronic computer, in its infinite wisdom, has imputed at this address a young couple plus a five year old daughter. On a revisit, however, we find the occupant at the time of the census was an elderly man. Was he counted - or wasn't he? It would seem clear that he wasn't; yet, from another point of view, it appears that he has been counted - not wisely, but too well!!

Speaking more seriously, we believe that the gradations in matching - and the variants - rule out the possibility of developing a single optimum matching procedure. The effects of a false match on the one hand, and the failure to make a valid match on the other hand, can each have a drastically different effect on the reliability of one survey as compared with another; and such value differences may make a matching procedure efficient in one survey, yet entirely inappropriate for another. What this group has been doing is developing and improving a variety of matching procedures, where each procedure is designed for a defined set of survey needs and objectives.

With this in mind, we want to take most of our time here to consider the needs and objectives of the coverage evaluation study, as related to matching. We will first give a summary of our matching methods - enough to make it clear that they differ significantly from others discussed at this session. For any that are interested, we have prepared a handout that describes in greater detail the matching procedures used in the matching coverage check studies.

Summary of Matching Procedures

The bases for the coverage evaluation were samples of persons who should have been counted in the census, selected from independent sources. We then check to see if these persons could be found in the census enumerations.

The matching was done by clerical coding, and professional review on a selective basis - the computer was not used. The information available for the match was quite variable. At the minimum, it consisted of the sample person's name and address; at the maximum, it included not only the sample person whose enumeration status was being checked, but the name of every person in his household at the time of the census, and the race, sex, age, marital status, and relationship of each person to the head of the household.

When any match - even vaguely possible - was found in census enumeration records, a two-part code was assigned as a measure of the degree of match. The first part of the code is based on the minimum information only - name and address. Here the definitions could be clear-cut, and there was relatively little chance for marginal cases. Independent verification indicated a very low error rate in these codings.

The second part of the code considers all information other than the name and address of the sample person. The question was "What additional evidence, if any, is provided by such information that the given census enumeration is, or is not, an enumeration of the sample person?" The code allowed for five categories of decision, ranging from "very strong additional evidence that the sample person was enumerated" to "very strong additional evidence that the census household being checked did not include the sample person." Clearly, the categories of decision were subjective. They could be illustrated, but not precisely defined, since it simply was not feasible to write into instructions the great variety of data combinations that might be encountered. What we did in the training of coders was not only to provide a variety of examples drawn from an early sample of actual match situations, but also to justify in specific detail each illustrative code. Some of these illustrative examples and justifications have been included in the handout materials. The names and addresses in the examples have been changed to protect the confidentiality of Census records; yet the variety of matching situations presented is authentic. We were satisfied with our training results for this coding, inasmuch as independent verification of the codes showed a high degree of consistency.

Match-Related Requirements of Coverage Evaluation

With what has just been said as a background, we should like to consider for a moment what the needs and objectives of a coverage evaluation study are - and how these affect matching:

(1) In our record check studies, we start with a sample of persons obtained from a source other than the 1960 Census. As we said earlier, in matching these persons against the 1960 Census, we seek to determine which ones have been missed in the census enumeration. Right here, we have an important difference between matching for coverage evaluation and many other matching studies. For most matching projects, the matched cases are the useful product, i.e., it is the matched cases that are studied, comparing information from the two matched sources. The focus

of a coverage study, on the other hand, is on persons that are unmatched. (The problem in these coverage studies, of course, is to distinguish between genuinely-missed persons and persons who are unmatched for other reasons.)

In most matching projects, by far the majority of cases turn out to be matches. Consequently, when it is the unmatched rather than the matched that is the base for the results, matching errors automatically become more serious percentagewise. Thus, if the "true" missed rate in a population is about 3 percent, the error of failing to match even 1 percent of the sample, when the persons involved are actually enumerated, biases the missed rate about a third!

(2) From one point of view, all errors made in the matching process can be put in two categories - positive errors, i.e., matches made between the sources when the persons involved are actually different, and negative errors, i.e., matches not made for sample persons who are actually listed in the source being searched. In coverage evaluation, these two types of match error are equally serious, since each distorts the missed rate to an equal degree - even though in opposite directions. The estimate of the missed rate is biased precisely to the degree that the expected number of positive match errors differs from the expected number of negative match errors.

In matching projects that are not used for coverage evaluation, positive and negative match errors are not necessarily of equal importance. Typically, the positive match error has the worse effect, since the false match puts a case in sample that does not belong, thus introducing a false association between the data from the two sources being matched. The negative match error, on the other hand, results merely in a loss of sample data - i.e., an increase in the non-response rate. Clearly, under these conditions, the two types of matching error are in no sense compensating.

(3) The two characteristics so far discussed - (a) the focus on unmatched cases rather than matched cases, and (b) the equal importance of positive and negative match errors - are basic to coverage evaluation studies. The 1960 record check studies had, in actual operation, a third characteristic that affected the matching results - a great range in the amount of information available for the match. For the most part this gradation was a reflection of the different sources from which our samples of names were drawn, and of the great variation in how current those sample lists were. One sample, for example, provided for the census search a name, in some cases an age, and a mailing address current as of January 1960. At the other extreme, one sample originally provided a great deal of information about the sample person, his personal characteristics, residence, other members of his household, their characteristics, and even neighbors and their personal characteristics - all this with one slight drawback.....the list was 10 years old!! In this case, of course, we did not put the information directly into a census matching operation, but first went through a location procedure in which we attempted to learn whether the person was still alive in April 1960, his current residence, and details about the household in which he currently lived. When the location procedure was successful, we typically had a good deal of detailed information upon which a census enumeration identification could be based.

Supporting Evaluation Procedures

Once the match-related requirements of a survey are clearly recognized, it becomes easier to choose between alternative matching procedures - it is also easier to see when no matching procedure, by itself, can possibly do the job required. This was the case in the coverage evaluation program. The sources of our lists of sample persons provided us with addresses as well, but in a significant number of cases these addresses were not where the sample persons were living at the time of the 1960 Census. Under these circumstances, any matching procedure whatever would return the sample persons as unmatched, not because they were missed in the census, but because they were enumerated at addresses that were not searched. Such negative matching errors would substantially overstate the estimate of the missed rate.

A field reconciliation operation was set up to reduce the number of erroneous decisions that would otherwise result from the matching operation. Field reconciliation was carried out by letter, by telephone, and by personal visit. All unmatched persons plus all problem cases were sent for reconciliation. Only the "sure" matches were treated as final decisions. Broadly speaking, three types of reconciliation were done, as follows:

(a) In all cases, careful questioning was done to determine all other addresses where the sample persons should or would have been enumerated. (This probing, obviously, was designed to reduce the number of negative matching errors we spoke of a moment ago.)

(b) In some cases, especially where a rural address was involved, we had difficulty in conducting a satisfactory census search because it was difficult to determine the exact geographic location of the address, and consequently, which Census Enumeration District should be examined. Here, the goal of the reconciliation was to pinpoint the location of the address on a map; this geographic location was then readily translated into an equivalent Census Enumeration District.

(c) Cases in which the matching operation found in the census a "probable" or, perhaps, just a "possible" misspelling of the sample person's name, were

also included in reconciliation. Here, the goal of reconciliation was to search for the persons with the name as spelled in the census. If no such named person was found, or that named person was identified as the sample person, this was considered as confirming evidence for a match.

Summary

As stated earlier, we are happy to supply any interested person with a detailed description of the matching procedures used in the record check coverage studies of the 1960 Census. We hope it is clear that we are not recommending these procedures for indiscriminate use in surveys that require matching. In general, we suspect the procedures we followed entail costs that are higher per person matched or searched for than the costs of other systems of matching. In developing or selecting a system of matching, an essential criteria is the function that the matching system must perform in the given survey. In the paper, we have discussed the three requirements of coverage evaluation that are particularly critical to the matching of the three, undoubtedly the most important is the fact that coverage evaluation matching focuses on unmatched rather than matched cases.

BIBLIOGRAPHY

(1) U.S. Bureau of the Census. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Record Check Studies of Population Coverage. Series ER60, No. 2. Washington, D.C., 1964.

(2) U.S. Bureau of the Census. Procedural Studies of the 1950 Censuses, No. 1, Infant Enumeration Study: 1950.

(3) Census-NOVS Study of Death Certificates Matched to Census Records, Milbank Memorial Fund Quarterly, April 1959, Vol. XXXVII, No. 2, pp. 144-153.

(4) U.S. Bureau of the Census. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Accuracy of Data on Housing Characteristics. Series ER60 No. 3. Washington, D.C., 1964.

(5) U.S. Bureau of the Census. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Accuracy of Data on Population Characteristics as Measured by Reinterviews. Series ER60, No. 4. Washington, D.C., 1964.

(6) U.S. Bureau of the Census. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Accuracy of Data on Population Characteristics as Measured by CPS-Census Match. Series ER60, No. 5. U.S. Government Printing Office, Washington, D.C., 1964.

(7) U.S. Bureau of the Census. Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: The Employer Record Check. Series ER60, No. 6. U.S. Government Printing Office, Washington, D.C., 1965.

HANDOUT

MATCHING FOR CENSUS COVERAGE CHECKS

In a number of census coverage checks, records for persons selected from record sources independent of the 1960 Census were compared to the census records for those persons, their households, and their addresses. For each comparison, a two-part code was clerically assigned to indicate the degree of information agreement between the two sources. The first part of that code was for address-name agreement while the second part of the code was for agreement of all information between the sources not used in the address-name coding. A large number of examples was used to train the clerks on how to assign composite codes - especially for the assignment of the second part of that code. The definitions given for the second part of the code were in relative terms and required some subjective interpretation by the coders. The examples served to guide their subjective interpretation.

Once the codes had been assigned, a sample of cases were reviewed in order to categorize the cases by MATCHED or NON-MATCHED by composite code. the MATCHED category consisted of those cases where the evidence was clear that the sample persons had been enumerated in the census. The NON-MATCHED included, in addition to census missed persons, all the matching problem cases. All NON-MATCHED cases were subjected to a reconciliation process carried out by office review, further field work, and additional census search for possible matches.

In this handout are the address-name codes with their descriptions, definitions of terms used in those descriptions, the supplemental information codes with their descriptions, some examples used for training the coders how to make composite code assignments (Illustrations 1-10), how these composite codes were used to classify each case as MATCHED or NON-MATCHED (Illustration 11), distribution of MATCHED and NON-MATCHED cases for some of the studies (Illustration 12), and the work sheet used to record all information pertinet to the matching operation used for these matching coverage studies (Illustration 13).

ADDRESS-NAME CODES

CODE	DESCRIPTION OF COMPARABILITY BETWEEN THE TWO SOURCES
A	Same address and same name
B	Same address and similar name or similar address and same name
C	Similar address and similar name or noncontradictory address and same name
D	Same or similar address and noncontradictory name or different address and same name
HA, HB, HC, HD	Same criteria as for Codes A-D above respectively except that the name part of the comparison was made on some person in the household other than the sample person (these codes were frequently used when the names for the sample person, as recorded in the two sources, were quite different, or there was no possible match for the sample person in the census household being examined).
X	All other address-name comparisons than those described above.

DEFINITIONS OF TERMS USED IN DESCRIPTIONS

SAME NAME

Two names are same if:

(a) They are identical with at least one given name present;

(b) The last and one given name are identical with

(1) no disagreement with respect to the other name or initial or

(2) only minor variation with respect to the other given name;

(c) An accepted contraction or nickname is given in one source for a first name; last name being identical.

SIMILAR NAME

Two names are similar if they differ too much to be considered "same" and the difference can be attributed to:

(a) Careless spelling or an error in interpreting handwriting;

(b) Phonetic spelling of the name

NONCONTRADICTORY NAME

Two names are noncontradictory if the last names are same or similar and:

(a) There is no first name in one of the sources; or

(b) Given names and initials between the two sources agree on initial basis.

SAME ADDRESS

Two addresses are same if:

(a) They are identical (this includes rural addresses if such addresses are numerically specific);

(b) The difference between the sources can be attributed to omission of street type, compass point, or apartment designation.

SIMILAR ADDRESS

URBAN

There are two parts of an urban address (a) street number and (b) street name. Two urban addresses are similar if:

(a) street number is identical while street name is similar where similar street name is described in the same terms as for a similar name for persons;

(b) street name is identical while street number is similar where similar street number is described as a number where the difference could arise from:

(1) a mistake in writing a digit in the number,

(2) an omission of a digit in a number, or

(3) the transposition of digits in a number.

RURAL

Two rural addresses are considered to be similar if:

(a) They are the same but are not numerically specific;

(b) They differ in description with some identical descriptive terms in both sources but the complete descriptions between the two sources are not contradictory.

NONCONTRADICTORY ADDRESS

Two addresses are noncontradictory if they differ in description but are not in opposition in the meaning (this term is used only for rural addresses where addresses often are described using different civil level descriptions for the road name).

DIFFERENT ADDRESS

Two addresses are different if they differ too greatly to be considered either same, similar, or noncontradictory.

SUPPLEMENTAL INFORMATION CODE

CODE	DESCRIPTION OF COMPARABILITY BETWEEN THE TWO SOURCES
++	Excluding information used in the address-name coding, the remaining evidence is very strong that the sample person is included in the Census household being compared. Minimum requirements for this code is that there be other household members in both sources and that the personal characteristics (age, race, sex, relationship to head, marital status) of the sample person and the household information essentially agree between the sources.
+	Excluding information used in address-name coding, there remains significant additional information that the sample person is included in the Census household being compared. Some items to consider are: (a) Sample person's age as given between the two sources is within 1 year of agreement; (b) Household composition, though not in complete agreement between the two sources, has more positive agreements than negative disagreements; (c) "Race" and "marital status" are not not sufficient for significance but add positive evidence if race is other than white or Negro or marital status is other than married or never married and these agree between the sources.
0	Excluding information used in address-name coding, the remaining evidence, or balance, provides no significant amount of additional evidence that the sample person is or is not included in the household being compared. Some instances of use of this code are: (a) When only address and name is given between the sources; (b) When information on personal characteristics is available but is not sufficient to be considered significant evidence; (c) When additional information is available but there are enough contradictions in the available information to offset any positive significant evidence in both sources.
?	Excluding information used in address-name coding, there is a significant amount of remaining evidence that the sample person is not included in the household being compared. This contradictory evidence can be of the form "probable different household" or some disparity in personal characteristics for the possible sample person match between the two sources.
??	Excluding information used in address-name coding, there is very strong evidence that the sample person is not included in the Census household being compared. The contradictory evidence here can take the form "different household" or different personal characteristics--especially large differences in age where the possible matching sample person has an age that would place him in another generation from the sample person in the independent record source.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 1

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE
INDEPENDENT RECORD SOURCE	1830 W. Flat St.		
	*Cavallo, Sue P.	Hd.	-
	Russo, Jack L.	Father	51
	_____, May B.	Mother	47
	_____, Charles J.	Brother	23
	_____, Jill A.	Sister	14
	_____, Margaret	Sister	7
	_____, Doris H.	Daughter	20 Months
CENSUS RECORD	1830 Flat 3rd		
	Russo, Jack L.	Hd.	51
	_____, May B.	Wife	47
	_____, Charles T.	Son	23
	**_____, Sue T.	Daughter	21
	_____, Jill A.	Daughter	15
	_____, Margaret T.	Daughter	7

*Sample Person **Possible Sample Person

Code "HA++"

Justification

"HA" - "1830 W. Flat St. vs. 1830 Flat, 3rd." is same address.
- "Russo, Jack L." given in both sources is same name for a household member.

"++" - Both sources show the same name for the following members of the Russo family; May B. Charles (J.vs.T), Jull A. and Margaret (T). The ages and relationships are the same for the respective members of the two households. The sample person"s name, "Cavallo, Sue P." (Russo) might reasonably be the married name of Russo, Sue T.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 2

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE
INDEPENDENT RECORD SOURCE	1318 Steeple St.		
	Holster, James K.	Hd.	42
	_____, Edna	Wife	39
	_____, Allen	Son	18
	_____, Keven	Son	15
	Smith, Dave	Step-Son	12
	*_____, Louise	Step-Dau.	10
CENSUS RECORD	1314 Steeple		
	Holster, James	Hd.	42
	_____, Edna	Daughter	39
	_____, Allen	Son	18
	_____, Keven	Son	15
	_____, John	Son	12
	**_____, Jill	Daughter	10

*Sample Person **Possible Sample Person

Code "HB++"

Justification

"HB" - "1318 Steeple St. vs. 1314 Steeple" are considered similar addresses.

- "James K. Holster vs. James Holster" are considered same name for a household member.

"++" - Note that we have exactly 6 persons in the two households being compared, with ages that can be exactly matched - year for year. Four names (including the one already coded in the HB) are the same. True, the other two names (including the "possible" sample person) are entirely different, but these might represent changed names - before and after adoption.

The relationships shown are different in 3 cases, and appear to be unreliably recorded in the Census. Certainly is is unreasonable to show "Edna" as a 39-year old daughter of the 42-year old head.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 3

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE
INDEPENDENT RECORD SOURCE	1601 Wayne		
	*Hayes, Hattie	Hd.	63
	_____, Sue Ann	Grand Daughter	7
	Henson, Kari Lynn	Grand Daughter	6
CENSUS RECORDS	1601 Wayne		
	**Haze, Hattie	Hd.	63
	Henson, Carolyn	Rel.	8
	Haze, Suzan	Daughter	6

*Sample Person **Possible Sample Person

Code "B++"

Justification

"B" - "1601 Wayne" is shown in both sources

- "Hayes, Hattie" vs. "Haze, Hattie" are considered
- similar names.

"++" The same 3-person household composition is shown on both sources (a head and two children) with the "possible" sample person listed as head and 63 years old.

The 7 year old granddaughter shown in the source (Sue Ann Hayes) has a phonetically-similar name and the same age as the 6 year old daughter shown in the census (Suzan Haze). The relationship, granddaughter vs. daughter, might be considered noncontradictory.

The second child is shown by both sources as having the same last name, Henson, (Which is other than the last names of the other household members) and phonetically-similar first names. The age and relationship comparison are "noncontradictory" as defined.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 4

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE	MARITAL STATUS
INDEPENDENT RECORD SOURCE	Apple Tree Road (Rural)			
	Burr, Daag K.	Hd.	70	
	*____, Betty	Wife	70	
CENSUS RECORD	Route 52, 1st. house after State 25			
	Burr, Daag K.	Hd.	70	Married
	**____, Betty T.	Wife	70	Married

*Sample Person **Possible Sample Person

Code "C++"

Justification

"C" - "Apple Tree Road (Rural) "vs. "Route 52, 1st. house after State 25" are considered noncontradictory addresses; "Burr, Betty" vs. "Burr, Betty T" are considered same name.

"++" - Both sources list two persons in the household - Burr, Daag K., age 70, as head, and the "possible" sample person as his wife, age 70. The rather uncommon name "Daag" should be considered to have a particularly high weight as evidence.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 5

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE	MARITAL STATUS
INDEPENDENT RECORD SOURCE	3932 Steel St. *Grant, Troy Gene	Hd.	65	
CENSUS RECORD	3932 Steel **Grant, Gene T.	Hd.	65	Married

*Sample Person **Possible Sample Person

Code "A+"

Justification

"A" - "3932 Steel St. vs. 3932 Steel" are considered same address.

"Grant, Troy Gene" vs. "Grant Gene T." are considered same name.

"+" - Both census and source show household consisting of one person of the same age.

Note that, for matches involving only the personal characteristic of the sample person, this is about as good a match as can be expected. Knowledge of race and sex would not contribute appreciably to the match.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 6

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE	MARITAL STATUS
INDEPENDENT RECORD SOURCE	128 Dallas St.			
	Milano, Eugene S.	Hd.	22	
	*_____, Lucille G.	Wife	21	
CENSUS RECORD	128 Dallas - Basement			
	Milano	Hd.	-	Married
	**_____, ?	Wife	-	Married

*Sample Person **Possible Sample Person

Code "D+"

Justification

"D" - Note that "128 Dallas St." vs. "128 Dallas - Basement" are considered same address.

"Milano, Lucille G." vs. "Milano,?" are considered noncontradictory names.

"+" The source and the Census both show households consisting of exactly 2 persons, a married couple.

This is a very week "+", which might reasonably have been considered as "0". If the households being compared had had more than 2 people (with agreement in number and relationship), or if the Census had reported age for at least one of the couple that agreed with the source within one year, then the code would be definitely "+".

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 7

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE	MARITAL STATUS
INDEPENDENT RECORD SOURCE	Middlebrook State School for the Mentally retarted Michael, Carl B., M.D. *Hanes, John	 Superintendent Inmate		
CENSUS RECORD	Middlebrook State School **Hanes, John M.	 Inmate	45	Never

*Sample Person **Possible Sample Person

Code "BO"

Justification

"B" -"Middlebrook State School for the Mentally Retarded vs. Middlebrook State School" are considered similar address.

"Hanes, John vs. Hanes, John M." are considered same name.

"O" -The additional information consists only of the fact that the "possible" sample person is an inmate. This is certainly not very distinctive for an institutionaladdress. (The agreement of address is of course, already included in code "B"). If the source had shown an age for the sample person which was within one year of the age given in the census household, a code "+" would have been assigned.

Note: This case illustrates a possible deficiency in what was transcribed from the census records. In this case, the records would be reexamined to see if "Michael, Carl B." was included in the census as being the superintendent, and, if not, to transcribe the name of the person recorded as being in charge.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 8

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE	MARITAL STATUS
INDEPENDENT RECORD SOURCE	1498 West 168th St. Apt. 2B			
	Spurlock, Stanley	Hd.		
	_____, Babs	Wife		
	Stile, Agnes	Inmate		
	*Jackson, Carl	Lodger	23	Married
	_____, Alice	Lodger's Wife	23	Married
CENSUS RECORD	1417 W. 166th St. Apt. 3D			
	**Jackson,	Hd.	-	Married
	Jackson,	Wife	-	-

*Sample Person **Possible Sample Person

Code - "XO"

Justification

"X" - "1498 West 168th St., Apt. 2B vs. 1417 W. 166 St., Apt. 3D" are considered different addresses. Note that the difference arises from three "similarities" - the house number, the street number and the apartment number. Had any two of these been the same, the two addresses would have been considered similar.

"Jackson, Carl vs. Jackson" are considered noncontradictory names.

"O" The only evidence that "Jackson, Carl" is "Jackson, ___" beyond the poor name-address comparison is the fact that in both sources the "possible" is shown as married, wife present. This fact is not sufficient in itself to merit a "+", and, in any case, the fact that the couple is shown as living by themselves in one source and as living as lodgers in someone else's household in the other source is offsetting evidence in the "?" direction.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 9

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE	MARITAL STATUS
INDEPENDENT RECORD SOURCE	2845 Lee St.			
	*Spencer, Samuel G.	Hd.	34	Married
	_____, Geraldine B.	Wife	34	Married
CENSUS RECORD	2845 Lee St.			
	Spencer, Geraldine B.	Hd.		
	**_____, Samuel	Son		

*Sample Person **Possible Sample Person

Code "A?"

Justification

"A" - "2845 Lee St." is present in both sources.

- "Spencer, Samuel G." vs. "Spencer, Samuel" are considered same name.

"?" - Geraldine B. Spencer is listed as 34 year old wife of Samuel G. Spencer in the source and and as his mother in the Census. This is significant evidence that the two "Samuel Spencers" may be different people - possible father and son.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 10

The examples given in these Illustrations were used for training coders on how to make the composite code assignments. The examples come from actual cases encountered in the match. For this handout, however, the addresses and names have been altered to protect the confidentiality of census records.

	ADDRESS AND NAMES	RELATIONSHIP TO HEAD	AGE	MARITAL STATUS
INDEPENDENT RECORD SOURCE	1354 Main			
	*Smith, Charles M.	Hd.	20	
	Myers, Eva	Grandmother	87	
CENSUS RECORD	1354 Main, Apt. 4			
	Meyers, Eva.	Hd.	88	Never

*Sample Person

Code "HB??"

Justification

"HB" -"1354 Main vs. 1354 Main, Apt. 4" are considered <u>same address</u>.

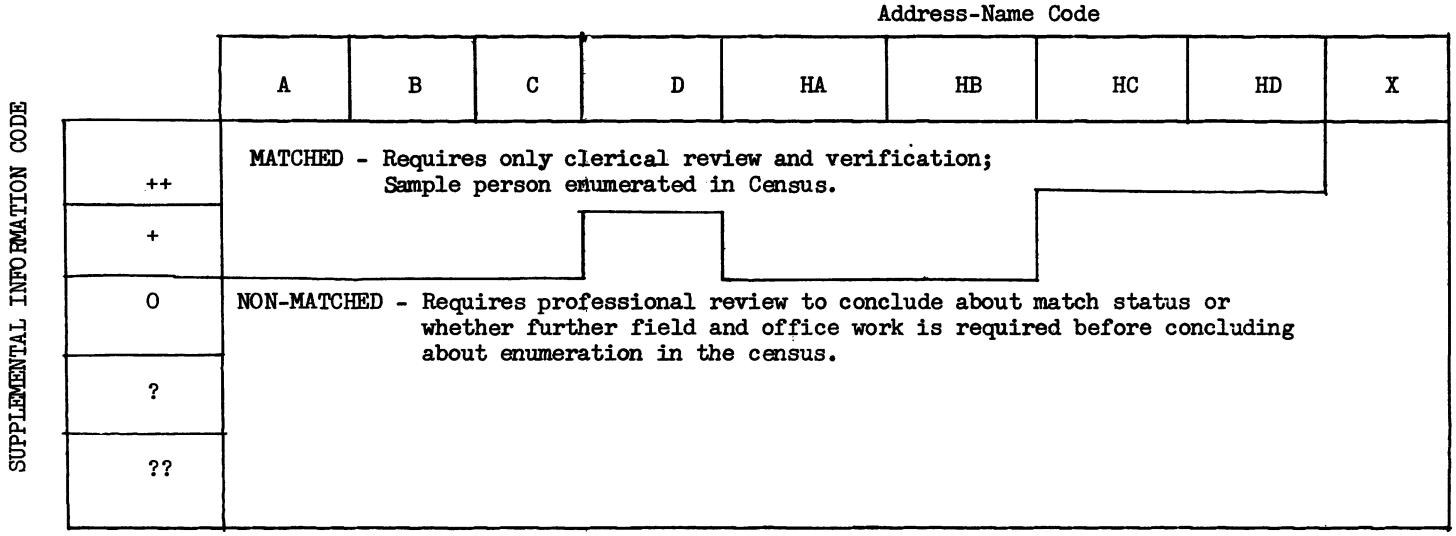
"Myers, Eva vs. Meyers, Eva" are considered <u>similar names</u> (Sample person not involved)

"??" -The sample person, Smith, Charles M., clearly does not appear in the Census household.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 11

After the composite codes were assigned, a sample of cases were reviewed in order to categorize the cases as MATCHED or NON MATCHED by composite code. The table below shows how the codes were divided into these categories.



MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 12

Percentage Distribution of "MATCH" and "NON-MATCH" cases for 4 record check studies reported in ER 60 No. 2 a/

	STUDY			
	TOTAL 4 Studies reported in ER 60 No.2 a/	Sample of Birth Records for Children born during inter-censal period 4/1/50 to 4/1/60	Sample of persons selected from 1950 Census and 1950 PES b/ records	Sample of Persons selected from alien registration records; January 1960
TOTAL	100.0	100.0	100.0	100.0
MATCHED (only clerical review and verification required)	84.0	86.6	82.7	62.8
NON-MATCH (professional review required to conclude about match status or whether further search or field work is needed)	16.0 c/	13.4 c/	17.3 c/	37.2 c/
Approximate number of cases for which census matching was attempted	6200	3700	2300	200

a/ U.S. Bureau of the Census, Evaluation and Research Program of the U.S. Censuses of Population and Housing, 1960: Record Check Studies of Population Coverage. Series ER 60 No. 2, Washington, D.C. 1964.

b/ U.S. Bureau of the Census, the POST-Enumeration Survey: 1950. Bureau of the Census Technical Paper No. 4. Washington, D.C. 1960.

c/ These percentages are not missed rates. Rather, they reflect the total group of matching problems from which the missed cases are finally culled by the process of professional review, further Census search, and field follow-up.

MATCHING FOR CENSUS COVERAGE CHECKS

ILLUSTRATION 13

Form 60-28-1.16 U.S. DEPARTMENT OF COMMERCE Bureau of the Census	1. Control Number	2. Best Match Code
WORKSHEET TO SEARCH CENSUS RECORDS FOR SAMPLE PERSONS	3. Source	4. Date on Source

5. INFORMATION ON INITIAL SOURCE

a. Address	City	State

Line No. b	Name c	Relation to Head d	Sex e	Race f	Date of birth Month g	Date of birth Year h	Age i	Marital Status j
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								

6. FIRST POSSIBLE MATCH - CENSUS

a. Address (Listing Book)	b. Address (FOSDIC Book) Same as Listing Book

c. ED Number	d. L.B. Page	e. L.B. Line	f. FOSDIC Page	g. Key	h.	i.	j.	k.	l.

Line No. m	Name n	Relation to Head o	Sex p	Race q	Date of birth Month r	Date of birth Year s	Age t	Marital Status u
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								