



UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Washington, DC 20233-0001


October 9, 2002


MASTER FILE

DSSD A.C.E. REVISION II MEMORANDUM SERIES #PP - 5

PRED CENSUS AND SURVEY MEASUREMENT STAFF MEMORANDUM SERIES:
CSM-A.C.E. Revision II-01R

MEMORANDUM FOR: Maureen P. Lynch
Assistant Division Chief, Coverage Measurement Processing,
Decennial Statistical Studies Division (DSSD)

From: Mary Mulry *signed 10/9/02* 
Chair, A.C.E. Revision II Quality Indicators Group Leader
Statistical Research Division

Through: David Hubble *signed 10/9/02* 
Assistant Division Chief, Evaluations, Planning, Research, and
Evaluation Division (PRED)

Prepared By: Susanne L. Bean
Mathematical Statistician, PRED

Subject: A.C.E. Revision II – Preparation of the P-sample Input File for the
Duplicate Links Evaluation

Attached is the specification for the preparation of the P-sample input file for the Duplicate Links Evaluation. This will be used for the study which examines the quality of the computer duplicates with administrative records. Please direct any comments or questions to Susanne Bean, 301-763-9590.

cc: DSSD A.C.E. Revision II Memorandum Series Distribution List

1. INTRODUCTION

The initial goal of this evaluation is to use the Census Numident File and the Statistical Administrative Records System (StARS) 2000 to determine the effectiveness of the methodology used in the Further Study of Person Duplication in Census 2000. For this evaluation, all A.C.E. person records will be matched to the administrative records to estimate the amount of person duplication.

This document defines the steps and file layouts required to prepare the P-sample file (which includes a record for every P-sample person and inmover in the A.C.E. sample) which will be used as the input to the administrative records matching process. This file is for matching purposes and does not include all the information necessary for data analysis.

2. PROCESS

The P-sample Input file will contain person records from the A.C.E. interview process. In order to facilitate the matching process, selected data will be extracted from the multiple source data files and compressed into a single national file.

a. Universe

All person records from the A.C.E. interviewing are to be included in the files with the exception of persons from Puerto Rico (defined by LCO code of 2150 to 2158).

- i. P-sample inputs: All persons from the A.C.E. interviews that went to person clerical matching are to be included. The data items are to be recoded as they were for person clerical matching.
- ii. Inmover inputs: All persons that went into the Missing Data Input processing that did not go through clerical matching are to be included. These persons must satisfy *each* of these two conditions:
 - the person has a mover flag (MOVERPER) of 1 (nonmover) or 2 (inmover); exclude infants born after Census Day (RSC equal to B); AND
 - the person comes from a housing unit with final outcome code for A.C.E. Interview Day (FINOUTI) of 1, 2, or 3 (representing an interview).

More information on the source files can be found in the DSSD Census 2000 Procedures and Operations Memorandum Series:

- Chapter S-CO-03, “Requirements for Person Matching Review and Coding System (PerMaRCS) File Transfers”, by Susan M. Odell and Gary VanNest, dated December 6, 1999.

- Chapter S-PT-09, “2000 Census Accuracy and Coverage Evaluation Person Computer Matching Specifications”, by Damon Smith, dated December 29, 1999.

- Chapter Q-40, “Accuracy and Coverage Evaluation Survey: Specifications for Data Requirements for Missing Data Input and Output Files”, by Michael Ikeda, dated January 19, 2001.

b. Intmaster Variables

i. For each person record that went to person clerical matching , create a P-sample Input File record setting SOURCE to ‘P’.

ii. Create a P-sample Input File record setting SOURCE to ‘I’ for each person record that satisfies *each* of these conditions:

- the person has a mover flag (MOVERPER) of 1 (nonmover) or 2 (inmover); exclude infants born after Census Day (RSC equal to B); AND
- the person comes from a housing unit with final outcome code for A.C.E. Interview Day (FINOUTI) of 1, 2, or 3 (representing an interview); AND
- the person did not go to person clerical matching.

iii. Extract variables listed in Attachment A.

(1) Recode SEX as follows:

<u>Original value</u>	<u>New value</u>
D,R	blank
1	M
2	F

(2) If the MOB contains a value other than ‘01’ through ‘12’ then blank fill.

(3) If DOB contains a value other than ‘01’ through ‘31’ then blank fill.

(4) If YOB contains the value ‘D’ or ‘R’ then blank fill.

(5) If Age contains the value ‘D’ or ‘R’ then blank fill.

(6) Eliminate the leading zero from the AGE field.

(7) Obtain FIPS state code for the cluster corresponding to the CLUSTER1 field.

c. P-sample Input File

The resulting file from this process, DUPEVALQI.DAT, should be an ASCII file with the layout described in Attachment B.

Variables to Extract

Extract the following variables from the IntMaster person records:

Variable	Description	Format	Position
CLUSTER1	Cluster Number & Check Digit	A6	1 - 6
MSN1	ICM Map Spot Number	A5	7 - 11
WMSN1	ICM Within MSN	A4	12 - 15
UPERID	Unique person Number	A2	19 - 20
FNAME	First Name	A13	154 - 166
MINIT	Middle Initial	A1	167 - 167
LNAME	Last Name	A26	168 - 193
SEX	Sex BLANK IS VALID 1 = MALE 2 = FEMALE D = DON'T KNOW R = REFUSAL	A1	507 - 507
MOB	Month of Birth	A2	508 - 509
DOB	Day of Birth	A2	510 - 511
YOB	Year of Birth	A4	512 - 515
AGE	Age	A4	268 - 271

Pull the following variables from the IntMaster housing unit records where the housing unit record CLUSTER1/MSN1/WMSN1 matches the person record CLUSTER1/MSN1/WMSN1:

Variable	Description	Format	Position
HNO	House num	A10	442 - 451
STRNM	Street name (RR, Box, Hwy info for Rural Address)	A33	455 - 487
UNIT	Unit / Within Structure ID	A20	488 - 507
ZIP5	Zip Code	A5	631 - 635
LCO	Local Census Office	A4	650 - 653

Layout Name : ACELAY:DUPEVALQI
 Description : File for duplicate evaluation for indicators group
 Total Length : 145
 Date Created : 06-14-2002

#	Field	Field description	length	Positions		
				Beg	-	End
		* KEY 0: CLUSTER/MSN/WMSN/ PER *				
1.	CLUST	Cluster Number & Check Digit	6	1	-	6 CHAR
2.	LCO	Local Census office	4	7	-	10 CHAR
3.	MSN	ICM Map Spot Number	5	11	-	15 CHAR
4.	WMSN	ICM WITHIN MSN NUMBER	4	16	-	19 CHAR
5.	IPER	Unique person number	2	20	-	21 CHAR
		*Name and Demographics *				
6.	SOURCE	"P" OR "I" P-sample or In mover	1	22	-	22 CHAR
7.	FNAME	First Name (Parsed)	13	23	-	35 CHAR
8.	MNAME	Middle Initial (Parsed)	1	36	-	36 CHAR
9.	LNAME	Last Name (Parsed)	26	37	-	62 CHAR
10.	SEX	Sex Blank = Nonresponse,D,R M = Male F = Female	1	63	-	63 CHAR
11.	AGE	Age	3	64	-	66 CHAR
12.	YEAR	Year of Birth	4	67	-	70 CHAR
13.	PMOB	Month of Birth	2	71	-	72 CHAR
14.	PDOB	Day of Birth	2	73	-	74 CHAR
15.	HNO	House num	10	75	-	84 CHAR
16.	SNAME	Street name	33	85	-	117 CHAR
17.	DNO	Within structure ID	20	118	-	137 CHAR
18.	ZIP	Zip code	5	138	-	142 CHAR
19.	FIPST	Fips state code	2	143	-	144 CHAR