



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

December 31, 2002

MASTER FILE

DSSD A.C.E. REVISION II MEMORANDUM SERIES # PP-13

**PRED CENSUS AND SURVEY MEASUREMENT STAFF MEMORANDUM SERIES:
CSM-A.C.E.REVISION II-06R**

MEMORANDUM FOR: Donna Kostanich
Chair, A.C.E. Revision II Planning Group
Decennial Studies Statistics Division

From: Mary Mulry *signed 12/31/02* *MM*
Chair, A.C.E. Revision II Quality Indicators Group
Statistical Research Division

Through: David Hubble *signed 12/31/02* *DH for*
Assistant Division Chief, Evaluations
Planning, Research, and Evaluations Division (PRED)

Prepared by: Donald Keathley
Mathematical Statistician, PRED

Subject: Procedure for Raking Outmovers to Inmovers in the A.C.E.
Revision II for the Evaluation of using Inmovers to Estimate the
Number of Census Outmovers

This document contains the specifications for raking outmovers to inmovers in the 2000 Accuracy and Coverage Evaluation (A.C.E.) Revision II. We wanted to compare the A.C.E. Revision II dual system estimates (DSE) that used inmovers to estimate the number of census outmovers to the A.C.E. Revision II DSEs that would estimate the census outmovers using raked outmover counts. Section I. provides a background as to why we want to rake outmovers to inmovers; section II. lists the input files and variables we need to do the raking; section III spells out the raking procedure; section IV lists the output files that both DSSD and PRED need for further processing; section V. specifies two PRED activities: verification of DSSD's implementation of the raking specifications and a supplemental replicate raking project.

Contact Don Keathley at donald.h.keathley@census.gov or at 301-457-8084 if you have any questions or comments.

cc. DSSD A.C.E. Revision II Memorandum Series Distribution List
Ruth Ann Killion (PRED)
David Hubble

I. Background

The Accuracy and Coverage Evaluation Survey (A.C.E.) was designed to measure and possibly correct for net coverage error in Census 2000. The A.C.E., as originally designed and conducted, was flawed. The Census Bureau's Executive Steering Committee on A.C.E. policy (ESCAP) recommended twice not to correct the census counts. There are, however, concerns about differential coverage error in Census 2000 data. While the Census 2000 data products will not be corrected, it is possible that improvements could be made to the intercensal and population estimates used for survey controls. This is the motivation for correcting errors in the A.C.E. data and developing improved estimates of the net undercount. These estimates are referred to as A.C.E. Revision II estimates. It's hoped that A.C.E. Revision II estimates will give a better picture of Census 2000 coverage and, in turn, improve operations for Census 2010 and methods for measuring coverage.

One of the possible errors in the A.C.E. is the use of in-mover counts to estimate the number of out-movers. Out-movers are persons who lived at a particular address on census day but did not live at the same address during the A.C.E. interview; in-movers are persons who did not live at a particular address on census day but who did live at the address during the A.C.E. interview. This method assumes that the out-mover counts from the A.C.E. underestimate the true out-mover counts. This is because the A.C.E. used proxy interviews to identify out-movers and that the respondents of these interviews didn't always identify everyone who had moved away after census day.

Theoretically, the number of in-movers in the United States should equal the number of out-movers. The A.C.E. used post-stratum-level in-mover counts, however, and it's possible that these counts are biased estimates of the corresponding post-stratum numbers of out-movers. In turn, this would bias the 2000 A.C.E. dual system estimates (DSE).

We're interested in learning how much of an effect the use of the in-mover counts had on the DSEs. One way of doing this is to apply the distribution of out-mover counts to the in-mover counts using a raking procedure. The procedure would be based on various demographic and geographic variables (not necessarily the post-stratum definitions). Then we can compare the DSEs that use the in-mover counts against the DSEs that use the raked out-mover counts, by post-stratum, FIPS state code, and region. We'll do this for the A.C.E. Revision II.

This document provides these raking specifications. DSSD will implement these specifications; PRED will verify DSSD's implementation.

II. Input Files

DSSD needed the data from two input files for the raking.

- A. Margs.sas7bdat
- B. Cells.sas7bdat

Margs and Cells are referenced in sections V.A.2.c. and d., respectively - we used them as input into the raking verification program also. Attachments C and D show layouts of margs and cells, respectively.

III. Procedure

A. Introduction

The DSE in the A.C.E. Revision II contains three sets of outmover and inmover counts. These are for:

- i. A.C.E. production, accounting for post-stratum definition changes
- ii. Same as i., except using A.C.E. Revision II sample weights and post-stratum definitions
- iii. Same as ii., except using A.C.E. Revision II sample coding

We originally wanted to rake all three outmover counts. The A.C.E. Revision II DSE post-stratified the mover counts in ii. and iii. on tenure only (owner, non-owner), however. So, we decided to rake the outmovers in i., only.

B. Marginals

We decided to use two sets of marginals for the raking procedure. These marginals are:

- 1. 8 age-sex \otimes 7 domain groups (56 cells), where we defined the individual groups on:
 - a. Age-sex groups
 - (1) age 0-9, both sexes
 - (2) age 10-17, both sexes
 - (3) age 18-29, male
 - (4) age 18-29, female
 - (5) age 30-49, male

- (6) age 30-49, female
- (7) age 50+, male
- (8) age 50+, female

b. Domain groups

- (1) American Indian or Alaska Native on a reservation
- (2) American Indian or Alaska Native not on a reservation
- (3) Hispanic
- (4) Non-Hispanic Black
- (5) Native Hawaiian or Pacific Islander
- (6) Non-Hispanic Asian
- (7) Non-Hispanic White or 'some othe race'

2. 2 tenure \otimes 4 MSA/TEA \otimes 3 return rate groups (24 cells), where we defined the individual groups on:

a. Tenure

- (1) Owner
- (2) Non-owner

b. MSA/TEA

- (1) Large MSA MO/MB
- (2) Medium MSA MO/MB
- (3) Small MSA & Non-MSA MO/MB
- (4) All other TEAs

c. Return Rate

- (1) Domains 3, 4, and 7, from section B.1.b. above
 - (a) High - tracts with the top quarter of census form return rates
 - (b) Low - all other tracts
- (2) Domains 1, 2, 5, and 6 - no return rate stratification

C. Collapsing

There was a chance that we wouldn't have been able to rake on the full set of marginals above due to either small cell sizes, excessively long raking iteration

program runs, or both. If that had been the case, then we would've collapsed the marginals on the age-sex and MSA/TEA groups:

1. Age-sex
 - a. age 0-17, both sexes
 - b. age 18-29, males
 - c. age 18-29, females
 - d. age 30+, males
 - e. age 30+, females

2. MSA/TEA
 - a. large and medium MSA MO/MB
 - b. small MSA MO/MB and all other TEAs

We didn't encounter any problems with the raking, however, using the full set of marginals. So, we didn't collapse.

D. Procedure

1. Assign each full-A.C.E. sample inmover and outmover to the appropriate age-sex \otimes domain \otimes tenure \otimes MSA/TEA \otimes return rate cell.
2. Sum the number of outmovers and inmovers in each cell. Keep the outmover and inmover counts separate (we would've collapsed the cells on the age-sex groups, MSA/TEA groups, or both, if that had been necessary).

Example: **Because of the number of cells in the actual raking process, we're using a mythical set of cells and counts for illustrative purposes.**

Suppose the summations resulted in the following counts, where:

V_1 = marginal set 1 (corresponding to the 56 age-sex \otimes domain cells above)

V_2 = marginal set 2 (corresponding to the 24 tenure \otimes MSA/TEA \otimes return rate cells above)

$V_1 \in \{1, 2, 3, 4\}$

$V_2 \in \{1, 2, 3\}$

e.g., V_1 could be [age \in {0-49, 50+}] and [domain \in {white, non-white}]
 V_2 could be collapsed MSA/TEA (Large, Medium, Other).

Table 1. Inmover Totals

V_1	V_2			Total
	1	2	3	
1	56	67	74	197
2	78	67	99	244
3	44	42	30	116
4	71	23	87	181
Total	249	199	290	738

Table 2. Original Outmover Totals

V_1	V_2			Total
	1	2	3	
1	23	44	37	104
2	46	54	41	141
3	87	62	52	201
4	54	48	53	155
Total	210	208	183	601

3. Rake on the V_1 marginals (the 56 age-sex \otimes domain cells)
 - a. Divide the inmover V_1 marginals (totals) by the outmover V_1 marginals (totals).

$$R_{1,i} = \frac{\text{inmover } V_{1,i}}{\text{outmover } V_{1,i}}$$

where $i \in$ (set of V_1 classifications)

- b. Multiply the outmover values in each V_1 row (individual cells only, i.e., not the marginals) by the appropriate $R_{1,i}$ value.

Example, cont'd: Calculate the $R_{1,i}$ s:

$$\begin{aligned} R_{1,1} &= 197 / 104 = 1.894231 \\ R_{1,2} &= 244 / 141 = 1.730496 \\ R_{1,3} &= 116 / 201 = 0.577114 \\ R_{1,4} &= 181 / 155 = 1.167742 \end{aligned}$$

Multiply the individual V_1 cell values in Table 2. by the appropriate $R_{1,i}$ s. For example, multiply each individual cell value in the $V_1 = 1$ row by 1.894231. Table 3. shows the results of the rake.

Table 3. Raked Outmover Totals 1

V_1	V_2			Total
	1	2	3	
1	43.56731	83.34615	70.08654	197
2	79.60284	93.44681	70.95035	244
3	50.20896	35.78109	30.00995	116
4	63.05806	56.05161	61.89033	181
Total	236.4372	268.6257	232.9372	738

4. Rake on the updated V_2 marginals (the 24 tenure \otimes MSA/TEA \otimes return rate cells)
- a. Divide the inmover V_2 marginals (totals) by the *updated* outmover V_2 marginals

$$R_{2,j} = \frac{\text{inmover } V_{2,j}}{\text{updated_outmover } V_{2,j}}$$

where $j \in (\text{set of } V_2 \text{ classifications})$

- b. Multiply the outmover values in each V_2 column (individual cells) by the appropriate $R_{2,j}$ value.

Example, cont'd: Calculate the $R_{2,j}$ s:

$$R_{2,1} = 249 / 236.4372 = 1.053134$$

$$R_{2,2} = 199 / 268.6257 = 0.740808$$

$$R_{2,3} = 290 / 232.9372 = 1.244971$$

Multiply the individual V_2 cell values in Table 3. by the appro $R_{2,j}$ s. For example, multiply each individual cell value in the $V_2 = 2$ column by 0.740808. Table 4. shows the results of the rake.

Table 4. Raked Outmover Totals 2

V_1	V_2			Total
	1	2	3	
1	45.88221	61.74348	87.2557	194.88139
2	83.83245	69.22613	88.33113	241.38971
3	52.87675	26.50692	37.36152	116.74519
4	66.40859	41.52347	77.05165	184.98371
Total	249	199	290	738

5. Rake on the V_1 marginals (the 56 age-sex \otimes domain cells)
 - a. Divide the inmover V_1 marginals (totals) by the *updated* outmover V_1 marginals (totals).

$$R_{1,i} = \frac{\text{inmover } V_{1,i}}{\text{updated_outmover } V_{1,i}}$$

where $i \in (\text{set of } V_1 \text{ classifications})$

- b. Multiply the outmover values in each V_1 row (individual cells only, i.e., not the marginals) by the appropriate $R_{1,i}$ value.

Example, cont'd: Calculate the new $R_{1,i}$ s:

$$R_{1,1} = 197 / 194.88139 = 1.010871$$

$$R_{1,2} = 244 / 241.38971 = 1.010814$$

$$R_{1,3} = 116 / 116.74519 = 0.993617$$

$$R_{1,4} = 181 / 184.98371 = 0.978465$$

Multiply the individual V_1 cell values in Table 4. by the appropriate $R_{1,i}$ s. For example, multiply each individual cell value in the $V_1 = 4$ row by 0.978465. Table 5. shows the results of the rake.

Table 5. Raked Outmover Totals 3

V_1	V_2			Total
	1	2	3	
1	46.38101	62.41471	88.20428	197
2	84.73898	69.97471	89.28631	244
3	52.53924	26.33772	37.12304	116
4	64.97845	40.62925	75.3923	181
Total	248.6377	199.3564	290.0059	738

6. Repeat steps D.4. and D.5. until all of the raked outmover marginals (Totals) equal the inmover marginals -or- until there have been 50 rakes (25 rakes for each marginal).

Example, cont'd: It would take 23 total rakes (twenty additional rakes from Table 5.) to match the raked outmover marginals to the inmover marginals in Table 1.

Table 6. Raked Outmover Totals 8

V_1	V_2			Total¹
	1	2	3	
1	46.46124	62.31655	88.22219	197
2	84.86656	69.84901	89.28443	244
3	52.60429	26.28339	37.11234	116
4	65.06791	40.55105	75.38104	181
Total	249	199	290	738

¹ Totals don't sum exactly because of rounding.

E. Post-Raking Distributions

We want to distribute the raked outmover counts across the P-Sample post-strata. This will allow us to calculate dual system estimates for each post-stratum, FIPS state code, and region.

1. P-Sample post-stratum

The individual cells in the raking procedure only approximate the P-Sample post-strata (see Attachment A for the post-strata definitions). Do the following to obtain post-stratum level raked outmover counts.

a. Domains 1, 2, and 6

These domains are not stratified by MSA/TEA nor by return rate group (return rate category = none) - sum the raked outmover counts across the 4 MSA/TEA groups for each domain \otimes age-sex \otimes tenure group

b. Domains 3 and 4

These two domains are collapsed on MSA/TEA and age-sex group:

(1) Collapsed MSA/TEA groups

- (a) large and medium MSA MO/MB
- (b) small MSA & non-MSA MO/MB, all other TEAs

Sum the raked outmover counts for groups (1)(a) and (1)(b) separately for each domain \otimes age-sex \otimes tenure \otimes return rate group.

(2) age-sex (for *non-owners*, collapsed MSA/TEA group (b) above, and low return rate only- see Attachment A)

- (a) 0-9, both sexes
- (b) 10-17, both sexes
- (c) 18+ male
- (d) 18+ female

Sum the collapsed raked outmover counts for groups (2)(a) through (2)(d) separately for the two domains.

c. Domain 5

Domain 5 is not stratified by MSA/TEA nor by return rate group (return rate category = none). Furthermore, it collapses the age-sex groups to the groups in b.(2) above.

- (1) Sum the raked outmover counts across the 4 MSA/TEA groups for each age-sex \otimes tenure group
- (2) Sum the collapsed raked outmover counts separately for groups b.(2)(a) through b.(2)(d) for each tenure group

d. Domain 7, owners

These cells are post-stratified further on region. Additionally, age-sex is collapsed in four low-return rate cells.

(1) Region

Apportion the raked outmover counts across the regions. For each MSA/TEA \otimes return rate \otimes age-sex group:

- (a) Determine the distributions of the pre-raked outmovers among the four regions
- (b) Apportion the raked outmover counts across the regions using these distributions.

Example: Suppose, in the northeast, for the following MSA/TEA \otimes return rate \otimes age-sex group:

MSA/TEA = large MSA MO/MB
return rate = high
age-sex = 0-9

the raked outmover count = 1,000. Suppose that the regional distribution of pre-raked outmovers in this group was:

Northeast = 0.21
Midwest = 0.27
South = 0.30
West = 0.22

Apportion the raked outmover count for this group using the above distribution, where:

$$\begin{aligned}\text{Northeast} &= (0.21 \cdot 1,000) = 210 \\ \text{Midwest} &= (0.27 \cdot 1,000) = 270 \\ \text{South} &= (0.30 \cdot 1,000) = 300 \\ \text{West} &= (0.22 \cdot 1,000) = 220\end{aligned}$$

(2) Age-sex

Four of the MSA/TEA \otimes region \otimes low-return rate groups collapse on age-sex - the collapsing is the same as shown in b.(2) above. These groups are:

- (a) large MSA MO/MB, midwest
- (b) large MSA MO/MB, west
- (c) medium MSA MO/MB, northeast
- (d) small MSA & non-MSA MO/MB, northeast

Sum the collapsed raked outmover counts for each collapsed age-sex group in b.(2) separately for each of these four low-return rate groups.

e. Domain 7, non-owners

The definitions for the marginals and post-strata for this domain and tenure were identical. Thus, no adjustments were necessary.

2. FIPS state code, Regions

We don't need to collapse the raked outmover counts across FIPS state codes or regions before we compute DSEs. This is because, first, we compute DSEs at the post-stratum level; then we compute FIPS state- and region-level DSEs using the post-stratum level DSEs.

F. Raking Termination Exception

We were on a tight schedule during the implementation of this specification. It was possible that the raking program would take a prohibitive amount of time to complete. To compensate for this, we limited the number of rakings to 50 (25 for each marginal).

IV. Output File

DSSD produced four output files from the raking process:

- A. Imtest1.sas7bdat - referenced in sections V.A.2.c. and V.B.2.d.
- B. Covpb.sas7bdat - see below
- C. Covpc.sas7bdat - referenced in section V.A.2.e.
- D. Moverct.sas7bdat - referenced in section V.B.2.f.

Covpb has the same layout as covpc - attachment G shows an abbreviated layout of covpc. The difference between the two files is that covpb contains DSSD's total replicate inmover and raked outmover counts for each cluster while covpc contains the individual clusters' inmover and raked outmover counts. Covpb is located in dssd_se_sam:[d_olson.reace] on the vax.

V. PRED Activities

- A. Verification

PRED verified DSSD's implementation of the raking specifications. PRED's verification consisted of:

1. writing its own raking program

We used one program to verify the both the raking procedure itself and the post-stratum distributions of raked counts. This program is:

- verif2.sas - post-stratum distribution verification (Attachment B)

This program is in pred_cover1:[pred.raking] on the vax.

2. using various files from DSSD

DSSD provided us with the following sas files for the verification process:

- a. margs.sas7bdat

Margs contained the inmover marginal counts for each of the 56 domain \otimes age-sex and 24 tenure \otimes MSA/TEA \otimes return rate groups.

b. cells.sas7bdat

Cells contained the outmover counts for each of the 1,344 individual raking cells.

c. imtest1.sas7bdat

Imtest1 contained the raked counts that DSSD produced, by the 1,344 individual raking cells.

d. rakemap.sas7bdat

Rakemap contained the mapping of the 1,344 individual raking cells to the P-Sample post-strata.

e. covpc.sas7bdat

Covpc contained DSSD's unraked and raked outmover counts by P-Sample post-strata.

Margs, cells, imtest1, and rakemap are in dssd_se_sam:[d_olson.zuw]; covpc is in dssd_se_sam:[d_olson.reace]. See Attachments C-G, respectively, for the layouts of these files.

3. comparing the raked values

There are four general parts to verif2.sas: part one rakes the outmover counts; part two compares these counts to what DSSD produced, by raking cell; part three distributes the raked counts to the post-strata, and part four compares these distributions to DSSD's distributions. See the comments in Attachment B for the details.

4. analyzing the results

For the raking cell comparisons, all of the verif2.sas-produced raked counts matched the DSSD raked counts on the first go-around. For the distribution of raked counts comparison, only the domain 7, owners, low-return rate, 18+ post-strata produced discrepancies (8 post-strata). It turned out that DSSD had made an error in the addition of counts for these post-strata. DSSD quickly corrected this error, after which PRED approved their results.

B. Supplemental Replicate Raking

A loss function analysis is accompanying the A.C.E. Revision II. Part of this analysis involves the use of 32 replicates. In turn, we need to compute 32 sets of raked outmover-to-inmover values, one set for each replicate.

1. Replicates

The replicates are cluster based, i.e., a cluster is either in or not in a given replicate.

2. Input Files

a. repfile.sas7bdat

Repfile contains the replicate sample assignments for each cluster. See Attachment H for the layout.

b. imargct.sas7bdat

Imargct contained the inmover marginal counts for each of the 56 domain \otimes age-sex and 24 tenure \otimes MSA/TEA \otimes return rate groups, by cluster. See Attachment C for the layout.

c. ocellct.sas7bdat

Ocellct contained the outmover counts for each of the 1,344 individual raking cells, by cluster. See Attachment D for the layout.

d. imtest1.sas7bdat

Imtest1 contained the raked counts that DSSD produced, by the 1,344 individual raking cells (same as in the verification section - see Attachment E for the layout).

e. rakemap.sas7bdat

Rakemap contained the mapping of the 1,344 individual raking cells to the P-Sample post-strata (same as in the verification section - see Attachment F for the layout).

f. moverct.sas7bdat

Moverct contained DSSD's raked outmover counts by P-Sample post-strata and cluster. See Attachment I for the layout.

Repid is in the pred_cover1:[pred.k_bench.reace.inconsist] directory on the vax; the files in b.-f. are in dssd_se_sam:[d_olson.zuw].

3. Procedure

We used replicates.sas (see Attachment J) to rake the outmover counts and to distribute these counts among the post-strata for each replicate; replicates.sas is in pred_cover1:[pred.raking]. All of the steps in this section are in replicates.sas - see the Attachment for additional comments.

a. Raking

- (1) Merge the files in V.B.2.
- (2) Using a modification of the verification program in section V.A., calculate the raked outmover counts for each replicate. Use the procedure specified in section III.

b. Post-raking Distributions

Follow the procedure in section III.E.1. for distributing the raked outmover counts in each replicate to the P-Sample poststrata.

4. Output Files

There were 33 output files, one total sample file and one file for each replicate. These files are

- a. rep_in0.out - for the total sample
- b. rep_in1.out - rep_in32.out - for replicates 1-32, respectively

All of these files are in pred_cover1:[pred.raking].

Each file has the following layout:

Replicate number	1-2 (2.0)
P-Sample post-stratum	5-8 (\$char4.)
Raked outmover count	10-35 (26.15)

VI. References

Haines, Dawn (2002), "*A.C.E. Revision II - Computer Specifications for Defining Full P- and E-Sample Post-Strata,*" draft, dated December 13, 2002

Table 1. P-Sample Age/Sex Groupings

Age	8 Groups		4 Groups	
	Male	Female	Male	Female
0-9				
10-17				
18-29				
30-49				
50+				

Table 1. shows the age-sex breakdowns for the full P-Sample post-strata. The numbers in the High Return Rate and Low Return Rate columns in Table 2. (next page) correspond to the groups in Table 1., i.e., 8 = 8 groups, 4 = 4 groups.

Table 2. Full P-Sample Post-Stratum Groups and Number of Age/Sex Groupings

Race/Hispanic Origin Domain Number	Tenure	MSA/TEA	High Return Rate				Low Return Rate			
			NE	MW	S	W	NE	MW	S	W
Domain 7 Non-Hispanic White or “Some other race”	Owner	Large MSA MO/MB	8	8	8	8	8	4	8	4
		Medium MSA MO/MB	8	8	8	8	4	8	8	8
		Small MSA & Non-MSA MO/MB	8	8	8	8	4	8	8	8
		All Other TEAs	8	8	8	8	8	8	8	8
	Non-Owner	Large MSA MO/MB	8				8			
		Medium MSA MO/MB	8				8			
		Small MSA & Non-MSA MO/MB	8				8			
		All Other TEAs	8				8			
Domain 4 Non-Hispanic Black	Owner	Large MSA MO/MB	8				8			
		Medium MSA MO/MB	8				8			
		Small MSA & Non-MSA MO/MB All Other TEAs	8				8			
	Non-Owner	Large MSA MO/MB	8				8			
		Medium MSA MO/MB	8				8			
		Small MSA & Non-MSA MO/MB All Other TEAs	8				4			
Domain 3 Hispanic	Owner	Large MSA MO/MB	8				8			
		Medium MSA MO/MB	8				8			
		Small MSA & Non-MSA MO/MB All Other TEAs	8				8			
	Non-Owner	Large MSA MO/MB	8				8			
		Medium MSA MO/MB	8				8			
		Small MSA & Non-MSA MO/MB All Other TEAs	8				4			
Domain 5 Native Hawaiian or Pacific Islander	Owner	4				4				
	Non-Owner	4				4				
Domain 6 Non-Hispanic Asian	Owner	8				8				
	Non-Owner	8				8				
Domain 1 American Indian or Alaska Native, on a reservation	Owner	8				8				
	Non-Owner	8				8				
Domain 2 American Indian or Alaska Native, not on a reservation	Owner	8				8				
	Non-Owner	8				8				

VERIF2.SAS

```
options nodate nocentre;

libname x 'dssd_se_sam:[d_olson.zuw]';
libname z 'dssd_se_sam:[d_olson.reace]';
libname y '[]'; /* pred_cover1:[pred.raking] */

/*****/

/*          *****          */
/*          RAKING          */
/*          *****          */

/*****/

/* The <data margs2> data step reads in the inmover
   marginals for the full p-sample: marg56_1 -
   marg56_56 and marg24_1 - marg24_24. */

data margs2; set x.margs;
where clust = '000000';
/* Full p-sample subsetting - one record is read in
   from x.margs and one record is read to margs2. */

/*****/

/* The <data cells2> data step reads in the outmover
   cell values: cell1 - cell1344. */

data cells2(drop=i); set x.cells;
where clust='000000';
/* Full p-sample subsetting - one record is read in
   from x.cells and one record is read to cells2. */

array c{1344} c1-c1344;
array cell{1344} cell1-cell1344;

do i = 1 to 1344;
  c{i} = cell{i};
end;

/*****/

/* <data rake1> rakes the outmovers to the inmover
   marginal distribution. */

data rake1; merge margs2 cells2;

array out56(56) out56_1 - out56_56 (56*0);
array out24{24} out24_1 - out24_24 (24*0);
array marg56{56} marg56_1 - marg56_56;
```

VERIF2.SAS

```
array marg24{24} marg24_1 - marg24_24;
array fact56{56} fact56_1 - fact56_56;
array fact24{24} fact24_1 - fact24_24;
array cell{1344} cell1 - cell1344;
array c{1344} c1-c1344;

do u = 1 to 25; /* 25 rakes, one for each set of marginals. */

do i = 1 to 56; /* sum the original (for u=1) or raked (for
                u>1) outmover cell counts to the 56 */
  out56{i} = 0; /* domain-by-age/sex groups. */
  do j = 1 to 24;
    out56{i} = out56{i} + c{j+((i-1)*24)};
  end;
end;

do k = 1 to 56; /* calculate the raking factor */
  fact56{k} = marg56{k}/out56{k};
end;

do i = 1 to 56; /* apply the appropriate raking factor to */
  do j = 1 to 24; /* the most recently-calculated outmover
                  cell counts */
    c{j+((i-1)*24)} = c{j+((i-1)*24)} * fact56{i};
  end;
end;

do i = 1 to 24; /* sum the raked outmover cell counts to the 24 */
  out24{i} = 0; /* tenure-by-MSA/TEA-by-return rate groups */
  do j = 0 to 1320 by 24;
    out24{i} = out24{i} + c{i+j};
  end;
end;

do k = 1 to 24; /* calculate the raking factor */
  fact24{k} = marg24{k}/out24{k};
end;

do i = 1 to 24; /* apply the appropriate raking factor to the */
  do j = 0 to 1320 by 24; /* most recently-calculated outmover
                          cell counts */
    c{i+j} = c{i+j} * fact24{i};
  end;
end;

end;

/*****/
```

VERIF2.SAS

```
/*          ***** */
/*          END OF RAKING */
/*          ***** */
```

```
/***/
```

```
/***/
```

```
/*          ***** */
/*          VERIFICATION OF THE RAKING */
/*          ***** */
```

```
/***/
```

```
/* <data check> transposes the raked data to a 1,344
   record-level file, one record per outmover cell,
   with the unraked and raked counts for each cell. */
```

```
data check(keep=cellno marg56 marg24 unraked raked);
  set rake1;
```

```
  array cell{1344} cell1-cell1344;
  array c{1344} c1-c1344;
  array m56{56} m56_1-m56_56;
  array m24{24} m24_1-m24_24;
```

```
  retain cellno 0;
```

```
  do a = 1 to 56;
    do b = 1 to 24;
```

```
      marg56=a; marg24=b;
      unraked = cell{b+((a-1)*24)};
      raked   = c{b+((a-1)*24)};
      cellno+1;
```

```
      output check;
```

```
    end;
  end;
```

```
/***/
```

```
/* <data full> reads in the raked outmover counts
   that DSSD produced, for the full p-sample */
```

```
data full; set x.imtest1;
  where clust = '000000';
```

VERIF2.SAS

```
/* Full p-sample subsetting - one record is read in
   from x.margs and one record is read to x.imtest1 */

/*****/

/* <data check2> merges the raked outmover counts in
   work.rake1 with the DSSD raked values from work.full */

proc sql;
  create table check2 as
  select *
  from work.check as c, work.full as f
  where c.marg56 = f.d1 and
        c.marg24 = f.d2
  order by marg56, marg24;

/*****/

/* <data check3> compares the two sets of raked and unraked
   outmover counts and prints the non-matches to separate files */

data check3 u r; set check2 end=eof;
  retain u r 0;

if unraked ne cell then do; /* this if-then compares the unraked */
  u + 1; output u; /* outmover counts from the two files - */
end; /* post script - all comparisons matched
      i.e., work.u was an empty data set */

attrib rpred rdssd format=comma11.2;
rpred=round(raked,.01);
/* rpred = raked outmover count from work.rake1 */
rdssd=round(fit,.01);
/* rdssd = raked outmover count from DSSD (x.imtest1) */

if rpred ne rdssd then do; /* this if-then compares the raked */
  r + 1; output r; /* outmover counts from the two files - */
end; /* post script - all but two comparisons
      matched, i.e., there were two records
      in work.r - the cause was rounding, so
      the results were okay. */

if eof then output check3;

/*****/

/* the three proc prints display the comparison
   results from work.check3. */
```


VERIF2.SAS

```
proc print data = check3;
  var u r;

proc print data = u;
  var marg56 d1 marg24 d2 unraked cell;

proc print data = r;
  var marg56 d1 marg24 d2 rpred rdssd;

/*****/

/*          *****/
/*          END OF RAKING VERIFICATION          */
/*          *****/

/*****/

/*****/

/*          *****/
/*          RAKED COUNT POST-STRATUM DISTRIBUTION          */
/*          *****/

/*****/

/* <sql check4> merges work.check2 with x.rakemap to obtain
   each outmover cell's full p-sample post-stratum code
   (pfullshort) */

proc sql;
  create table check4 as
  select *
  from work.check2 as c full join x.rakemap as r
  on c.marg56 = r.danum and
     c.marg24 = r.tnnum
  order by pfullshort, marg56, marg24;

/*****/

/* <data covpc> reads in the unraked mover counts for each
   full p-sample post-stratum - this was necessary for computing
   the adjustment/apportionment factor for the post-strata in
   domain 7, owners. */

data covpc; set z.covpc;
  where clust = '000000';
  rename pgrp=pfullshort;
  /* Full p-sample subsetting - one record is read in
     from z.covpb and one record is read to covpb. */
```

VERIF2.SAS

```
proc sort data = covpc;
  by pfullshort;

/*****/

/* <data all, collapsed> merges the two previously-created
   data sets, by full p-sample post-stratum. This is where
   the raked outmover counts, calculated above in work.rake1,
   are distributed/apportioned to the post-strata */

data all collapsed; merge check4(in=a) covpc(in=c);
  by pfullshort;
  if a or c;

  retain out rake unrake 0;

  /* 1. apportions the raked outmover counts to the
     four regions in the domain 7, owner, low return
     rate post-strata with collapsed age/sex
     categories.
     2. apportions the raked outmover counts to the
     four regions for the remaining post-strata in
     domain 7, owners
     3. distributes the remaining raked outmover counts
     among the non-(domain 7, owners) post-strata */

/* 1 */
if pfullshort in('0671','0672','0871','0872',
  '1371','1372','2171','2172') then do;
  rake = rake + raked;
  unrake = unrake + unraked;
  if last.pfullshort then do;
    if unrake = 0 then factor = 0;
    else factor = rake / unrake;
    outmover = factor * ypa4; /* ypa4 is the unraked mover count */
    out = outmover; /* for the given region */
  end;
end;

/* 2 */
else if marg56 ge 49 and marg24 le 12 then do;
  if unraked = 0 then factor = 0;
  else factor = raked / unraked;
  outmover = factor * ypa4; /* ypa4 is the unraked mover count */
  out = out + outmover; /* for the given region. */
end;

/* 3 */
else do;
  factor = 1; /* no apportionment here - just collapsing */
```

VERIF2.SAS

```
    outmover = raked;
    out = out + outmover;
end;

output all; /* full file, with running tabs, etc. */

if last.pfullshort then do;
    output collapsed; /* 480-record full p-sample post-stratum file */
    out = 0; rake = 0; unrake = 0;
end;

/*****

/* ***** */
/* END OF RAKED COUNT POST-STRATUM DISTRIBUTION */
/* ***** */

/*****

/*****

/* ***** */
/* POST-STRATUM DISTRIBUTION VERIFICATION */
/* ***** */

/*****

/* the two proc sort-proc print sets print out the results from
the previous data step - this is an internal verification, to
make sure the program ran correctly - they did */

proc sort data = all;
  by marg56 marg24 pfullshort;

proc print data = all;
  var marg56 marg24 pfullshort raked rake unraked unrake ypa4 factor outmover;

proc sort data = collapsed;
  by pfullshort;

proc print data = collapsed;
  var pfullshort rake unrake factor ypa4 out;

/*****

/* <data covpc> reads in the DSSD-computed raked outmover counts,
   by full p-sample post-stratum */

data covpc; set x.covpc;
```

VERIF2.SAS

```
where clust = '000000'; /* one-record read in and out */
rename pgrp=pfullshort;
keep pgrp ypa5;

proc sort data = covpc;
  by pfullshort;

/*****/
/* <data gnarley, bogus> merges this program's apportionment/
  distribution of raked outmover counts with those from DSSD
  and then compares them, by full p-sample post-stratum */

data gnarly bogus; merge collapsed(in=c)
  covpc(in=v);
  by pfullshort;
  if c or v;

dssd = round(ypa5,.1);
  /* dssd = raked outmover counts, by full p-sample
  post-stratum, from work.covpc */
pred = round(out, .1);
  /* pred = raked outmover counts, by full p-sample
  post-stratum, from work.collapsed */

if dssd ne pred then output bogus; /* output non-matches */
else output gnarly; /* output matches */

proc print data = bogus;
  var pfullshort dssd pred ypa5 out; /* non-matches - two present, due
  to tenths being off by 1 (e.g.,
  xxx.1 vs. xxx.0) */

proc print data = gnarly;
  var pfullshort dssd pred ypa5 out; /* matches - the remaining
  full p-sample post-stratum
  raked counts matched */

/*****/

/* *****/
/* END OF POST-STRATUM DISTRIBUTION VERIFICATION */
/* *****/

/*****/
```

Layout for Margs.sas7bdat and Imargct.sas7bdat (variables used by PRED only)

Each marg56_# and marg24_# contains the marginal in-mover count for the combination in the Description column below, e.g., marg56_1 contains the in-mover count for domain 1, persons 0-9.

The difference between the two files is that margs.sas7bdat contains DSSD replicate counts for each cluster; imargct.sas7bdata contains the raw in-mover counts. We were able to use margs.sas7bdata for clust = '000000' only (the full p-sample)..

Variables	Description		Format
Clust	Cluster		\$char6
	Domain	Age-Sex	
Marg56_1	1	0-9, both	8.0
Marg56_2	1	10-17, both	8.0
Marg56_3	1	18-29, Male	8.0
Marg56_4	1	18-29, Female	8.0
Marg56_5	1	30-49, Male	8.0
Marg56_6	1	30-49, Female	8.0
Marg56_7	1	50+, Male	8.0
Marg56_8	1	50+, Female	8.0
Marg56_9	2	0-9, both	8.0
Marg56_10	2	10-17, both	8.0
Marg56_11	2	18-29, Male	8.0
Marg56_12	2	18-29, Female	8.0
Marg56_13	2	30-49, Male	8.0
Marg56_14	2	30-49, Female	8.0
Marg56_15	2	50+, Male	8.0
Marg56_16	2	50+, Female	8.0
Marg56_17	3	0-9, both	8.0
Marg56_18	3	10-17, both	8.0
Marg56_19	3	18-29, Male	8.0
Marg56_20	3	18-29, Female	8.0
Marg56_21	3	30-49, Male	8.0
Marg56_22	3	30-49, Female	8.0
Marg56_23	3	50+, Male	8.0
Marg56_24	3	50+, Female	8.0

Layout for Margs.sas7bdat and Imarget.sas7bdat (variables used by PRED only)

Variables		Description	Format
Marg56_25	4	0-9, both	8.0
Marg56_26	4	10-17, both	8.0
Marg56_27	4	18-29, Male	8.0
Marg56_28	4	18-29, Female	8.0
Marg56_29	4	30-49, Male	8.0
Marg56_30	4	30-49, Female	8.0
Marg56_31	4	50+, Male	8.0
Marg56_32	4	50+, Female	8.0
Marg56_33	5	0-9, both	8.0
Marg56_34	5	10-17, both	8.0
Marg56_35	5	18-29, Male	8.0
Marg56_36	5	18-29, Female	8.0
Marg56_37	5	30-49, Male	8.0
Marg56_38	5	30-49, Female	8.0
Marg56_39	5	50+, Male	8.0
Marg56_40	5	50+, Female	8.0
Marg56_41	6	0-9, both	8.0
Marg56_42	6	10-17, both	8.0
Marg56_43	6	18-29, Male	8.0
Marg56_44	6	18-29, Female	8.0
Marg56_45	6	30-49, Male	8.0
Marg56_46	6	30-49, Female	8.0
Marg56_47	6	50+, Male	8.0
Marg56_48	6	50+, Female	8.0
Marg56_49	7	0-9, both	8.0
Marg56_50	7	10-17, both	8.0
Marg56_51	7	18-29, Male	8.0
Marg56_52	7	18-29, Female	8.0
Marg56_53	7	30-49, Male	8.0
Marg56_54	7	30-49, Female	8.0
Marg56_55	7	50+, Male	8.0
Marg56_56	7	50+, Female	8.0

Layout for Margs.sas7bdat and Imarget.sas7bdat (variables used by PRED only)

Variables	Description			Format
	Tenure	MSA/TEA	Return Rate	
Marg24_1	Owner	Large	Low	8.0
Marg24_2	Owner	Large	High	8.0
Marg24_3	Owner	Large	None	8.0
Marg24_4	Owner	Medium	Low	8.0
Marg24_5	Owner	Medium	High	8.0
Marg24_6	Owner	Medium	None	8.0
Marg24_7	Owner	Small	Low	8.0
Marg24_8	Owner	Small	High	8.0
Marg24_9	Owner	Small	None	8.0
Marg24_10	Owner	Other TEA	Low	8.0
Marg24_11	Owner	Other TEA	High	8.0
Marg24_12	Owner	Other TEA	None	8.0
Marg24_13	Non-Owner	Large	Low	8.0
Marg24_14	Non-Owner	Large	High	8.0
Marg24_15	Non-Owner	Large	None	8.0
Marg24_16	Non-Owner	Medium	Low	8.0
Marg24_17	Non-Owner	Medium	High	8.0
Marg24_18	Non-Owner	Medium	None	8.0
Marg24_19	Non-Owner	Small	Low	8.0
Marg24_20	Non-Owner	Small	High	8.0
Marg24_21	Non-Owner	Small	None	8.0
Marg24_22	Non-Owner	Other TEA	Low	8.0
Marg24_23	Non-Owner	Other TEA	High	8.0
Marg24_24	Non-Owner	Other TEA	None	8.0

Layout for Cells.sas7bdat and Ocellct.sas7bdat (variables used by PRED only)

The variables in cells.sas7bdat and ocellct.sas7bdat contain Cell1 - Cell1344. These are the individual raking cells. They map sequentially to the marg56_# and marg24_# values from Attachment C, where the first 24 cell# map to marg56_1, the next 24 to marg56_2, etc. Within a marg56_#, the mappings are also sequential, starting with marg24_1 and ending with marg24_24. **Each cell# contains the unranked outmover count for that cell number, e.g., cell3 contains the unranked outmover count for the marg56_1 (domain 1, persons 0-9), marg24_3 (owners, large MSAs, no return rate category) combination.**

The difference between the two files is that cells.sas7bdat contains DSSD replicate counts for each cluster; ocellct.sas7bdata contains the raw, unranked outmover counts. We were able to use cells.sas7bdata for clust = '000000' only (the full p-sample)..

Variable	Description	Format	
Clust	Cluster	\$char6.	
	Marg56_#, where #=	Marg24_#, where #=	
Cell1	1	1	8.0
Cell2	1	2	8.0
Cell3	1	3	8.0
.	.	.	
Cell22	1	22	8.0
Cell23	1	23	8.0
Cell24	1	24	8.0
Cell25	2	1	8.0
Cell26	2	2	8.0
.	.	.	
Cell47	2	23	8.0
Cell48	2	24	8.0
Cell49	3	1	8.0
.	.	.	
Cell72	3	24	8.0
Cell73	4	1	8.0
.	.	.	
Cell1344	56	24	8.0

Layout for Imtest1.sas7bdat (variables used by PRED only)

Variable	Description	Format
Cell	Cell Number (1 - 1,344)	8.0
Clust	Cluster	\$char6.
D1	# values corresponding to marg56_#, where # \in (1-56)	8.0
D2	# values corresponding to marg24_#, where # \in (1-24)	8.0
Fit	raked outmover counts produced by DSSD	8.0
Rake_wt	ratio of the raked outmover count divided by the unraked outmover count.	8.0

Layout for Rakemap.sas7bdat (variables used by PRED only)

Variable	Description	Format
DANUM	# values corresponding to marg56_#, where # \in (1-56)	8.0
PFULLSHORT	A.C.E. Revision II p-sample post-stratum short code, where: <ul style="list-style-type: none"> - digits 1-2 are the post-stratum group number in Attachment A, Table 2 (64 categories) - digit 3 is the age category (10 categories) - digit 4 is the sex category (3 categories) see Haines (2002) for details on the coding.	\$char4.
TMNUM	# values corresponding to marg24_#, where # \in (1-24)	8.0

Layout for Covpc.sas7bdat (variables used by PRED only)¹

Variable	Description	Format
Clust	cluster	\$char6.
PGRP- same as PFULLSHORT	A.C.E. Revision II p-sample post-stratum short code, where: - digits 1-2 are the post-stratum group number in Attachment A, Table 2 (64 categories) - digit 3 is the age category (10 categories) - digit 4 is the sex category (3 categories) see Haines (2002) for details on the coding.	\$char4.
YPA4	unraked outmover count	8.0
YPA5	raked outmover count	8.0

¹ Same layout as covpb.sas7bdat - see section IV.

Layout for Repfile.sas7bdat (variables used by PRED only)

Variable	Description	Format
Clust	Cluster	\$char6.
Rep1_flag - Rep32_flag	Replicate sample indicators for replicates 1-32, where: 0 = cluster not in replicate 1 = cluster in replicate	8.0

Layout for Moverct.sas7bdat (variables used by PRED only)

Variable	Description	Format
Clust	Cluster	\$char6.
PFULLSHORT	A.C.E. Revision II p-sample post-stratum short code, where: <ul style="list-style-type: none"> - digits 1-2 are the post-stratum group number in Attachment A, Table 2 (64 categories) - digit 3 is the age category (10 categories) - digit 4 is the sex category (3 categories) see Haines (2002) for details on the coding.	\$char4.
Valuc2	Unraked outmover count	8.0

Replicates.sas

```
options nodate nocentre;

libname x 'dssd_se_sam:[d_olson.zuw]';
libname z '[]'; /* pred_cover1:[pred.raking] */

/*****/

/*      *****/
/*      INITIALIZE INPUT FILES      */
/*      *****/

/*****/

/* <sql margs> merges DSSD's inmover marginal counts,
   by cluster (x.imargct), with PRED's replicate flag file (z.repid) */

proc sql;
  create table margs(drop=clust) as
  select *
  from x.imargct as m, z.repid as r
  where m.clust = r.cluster;

/*****/

/* <sql cells> merges DSSD's unraked outmover
   cell counts for each raking cell, by cluster (x.ocellct),
   with PRED's replicate flag file (z.repid). */

proc sql;
  create table cells(drop=clust) as
  select *
  from x.ocellct as c, z.repid as r
  where c.clust = r.cluster;

/*****/

/* The two following data steps create rep_flag = 1 for
   both work. margs and work.cells - this represents
   the full-sample flag. */

data margs; set margs;
  rep_flag = 1;

data cells; set cells;
  rep_flag = 1;

/*****/

/*      *****/
/*      END INITIALIZE INPUT FILES      */
/*      *****/
```

Replicates.sas

```

/*****/

                /*****/

/*      *****/
/*      RAKING & POST-STRATUM DISTRIBUTION MACRO      */
/*      *****/

                /*****/

%macro reps(repvar, rep); /* repvar = replicate flag (rep_flag, rep#_flag, where # ∈ (1-32) ) - a 1 indicates
                           that the cluster is part of the replicate, a 0 indicates that it isn't.
                           rep      = rep number corresponding to the repvar input file number - 0 indicates
                           the full sample */

/***** */

/*      *****/
/*      SUM COUNTS ACROSS CLUSTERS      */
/*      *****/

/*****/

/* <data marg2> reads in work.margs from above. It sums the inmover marginals
   across the clusters for the applicable replicate. */

data marg2;
set marg2 end=eof;

where &repvar = 1;

keep m24_1-m24_24 m56_1-m56_56;

array marg24{24} marg24_1 - marg24_24;
array marg56{56} marg56_1 - marg56_56;
array m24{24} m24_1 - m24_24 (24*0);
array m56{56} m56_1 - m56_56 (56*0);

retain m24_1-m24_24 m56_1-m56_56;

do i = 1 to 24;
  m24{i} = m24{i} + marg24{i}; /* sums the tenure-by-msa/tea-by-return rate */
end; /* inmover counts across the replicate-sample clusters */

do j = 1 to 56;
  m56{j} = m56{j} + marg56{j}; /* sums the domain-by-age/sex inmover counts */
end; /* across the replicate-sample clusters */

if eof then output; /* a one-record output file */

/*****/

```

Replicates.sas

```
/* <data cells2> reads in work.cells from above. It sums the unraked outmover
   raking-cell counts across the clusters for the applicable replicate. */

data cells2;
  set cells end=eof;

  where &repvar = 1;

  drop cell1-cell1344;

  array cell{1344} cell1 - cell1344;
  array c{1344} c1 - c1344 (1344*0);
  array unr{1344} unr1-unr1344;

  retain c1-c1344;

  do i = 1 to 1344;
    c{i} = c{i} + cell{i}; /* the summation of unraked outmover counts, */
  end;                    /* by raking cell, across the replicate-sample clusters */

  if eof then do;
    do j = 1 to 1344;
      unr{j} = c{j};      /* keeping the unraked counts */
    end;
    output;              /* a one-record output file */
  end;

  /*****

  /* *****
  /* END SUM COUNTS ACROSS CLUSTERS
  /* *****
  /*

  /*****/

  /*****

  /* *****
  /* RAKING
  /* *****
  /*

  /*****/

  /* <date rake1> is the same as in VERIF2.SAS - see Attachment B
     for comments for this data step. */

data rake1; merge margs2 cells2;

  array out56(56) out56_1 - out56_56 (56*0);
  array out24{24} out24_1 - out24_24 (24*0);
  array m56{56} m56_1 - m56_56;
```


Replicates.sas

```
array m24{24} m24_1 - m24_24;
array fact56{56} fact56_1 - fact56_56;
array fact24{24} fact24_1 - fact24_24;
array unr{1344} unr1-unr1344;
array c{1344} c1 - c1344;

do u = 1 to 25;

  do i = 1 to 56;
    out56{i} = 0;
    do j = 1 to 24;
      out56{i} = out56{i} + c{j+((i-1)*24)};
    end;
  end;

  do k = 1 to 56;
    fact56{k} = m56{k}/out56{k};
  end;

  do i = 1 to 56;
    do j = 1 to 24;
      c{j+((i-1)*24)} = c{j+((i-1)*24)} * fact56{i};
    end;
  end;

  do i = 1 to 24;
    out24{i} = 0;
    do j = 0 to 1320 by 24;
      out24{i} = out24{i} + c{i+j};
    end;
  end;

  do k = 1 to 24;
    fact24{k} = m24{k}/out24{k};
  end;

  do i = 1 to 24;
    do j = 0 to 1320 by 24;
      c{i+j} = c{i+j} * fact24{i};
    end;
  end;

end;

/***** */

/*          ***** */
/*          END OF RAKING */
/*          ***** */

/***** */
```

Replicates.sas

```
/****** */

/****** */
/* RAKED COUNT POST-STRATUM DISTRIBUTION */
/****** */

/****** */

/* <data rake2> transposes the raked data to a 1,344
   record-level file, one record per outmove cell,
   with the unraked and raked counts for each cell. */

data rake2(keep=cellno marg56 marg24 unraked raked);
set rake1;

array unr{1344} unr1-unr1344;
array c{1344} c1-c1344;
array m56{56} m56_1-m56_56;
array m24{24} m24_1-m24_24;

retain cellno 0;

do a = 1 to 56;
  do b = 1 to 24;

    marg56 = a; marg24 = b;
    unraked = unr{b+((a-1)*24)};
    raked = c{b+((a-1)*24)};
    cellno + 1;

    output rake2;

  end;
end;

/****** */

/* <sql.rake3> merges work.rake2 with x.rakemap to obtain
   each outmove cell's full p-sample post-stratum code (pfullshort) */

proc sql;
create table rake3 as
select *
from work.rake2 as r full join x.rakemap as m
on r.marg56 = m.danum and
   r.marg24 = m.tmnum
order by pfullshort, marg56, marg24;

/****** */

/* <sql.unrakedo> merges x.moverct with work.margs to obtain
```

Replicates.sas

```
each full p-sample post-stratum's unraked outmover count */

proc sql;
  create table unrakedo as
  select *
  from x.moverct as m right join work.margs as g
  on m.clust = g.cluster
  order by pfullshort;

/*****/

/* <data counts> reads in work.rakedo - it sums the
   unraked mover counts for each full p-sample post-stratum */

data counts; set unrakedo;
  where &repvar = 1; /* subset on the appropriate replicate sample */
  by pfullshort;

  retain outmovct 0;

  outmovct + valuc2;

  if last.pfullshort then do;
    output; outmovct=0;
  end;

/*****/

/* <data all, collapsed> distributes/apportions the raked outmover
   counts to full p-sample post-strata. This is the same as <data
   all, collapsed> in VERIF2.SAS - see Attachment B for the
   data step's comments */

data all collapsed; merge rake3(in=r) counts(in=c);
  by pfullshort;
  if r or c;

  retain out rake unrake 0;

  if pfullshort in('0671','0672','0871','0872',
    '1371','1372','2171','2172') then do;
    rake = rake + raked;
    unrake = unrake + unraked;
  if last.pfullshort then do;
    if unrake = 0 then factor = 0;
    else factor = rake / unrake;
    outmover = factor * outmovct;
    out = outmover;
  end;
end;
else if marg56 ge 49 and marg24 le 12 then do;
  if unraked = 0 then factor = 0;
```

Replicates.sas

```
else factor = raked / unraked;
outmover = factor * outmovct;
out = out + outmover;
end;
else do;
  factor = 1;
  outmover = raked;
  out = out + outmover;
end;

output all;

if last.pfullshort then do;
  output collapsed;
  out = 0; rake = 0; unrake = 0;
end;

/*****/

/* The proc sort prepares work.collapsed for writing to
   an ascii file - <data _null_> writes the ascii file */

proc sort data = collapsed;
  by pfullshort;

data _null_;
  set collapsed;
  rep=&rep;
  if pfullshort ne ' ' then do;
    file "rake_in&rep.dat";
    put @1 rep 2.
        @5 pfullshort $char4.
        @10 out 26.15;
  end;
/*****/

/*****/
/* END RAKED COUNT POST-STRATUM DISTRIBUTION **/
/*****/

/*****/

%mend reps;

/*****/

/* ***** */
/* END RAKING & POST-STRATUM DISTRIBUTION MACRO */
/* ***** */

/*****/
```

Replicates.sas

```
%reps(rep_flag, 0) %reps(rep1_flag, 1) %reps(rep2_flag, 2)
%reps(rep3_flag, 3) %reps(rep4_flag, 4) %reps(rep5_flag, 5)
%reps(rep6_flag, 6) %reps(rep7_flag, 7) %reps(rep8_flag, 8)
%reps(rep9_flag, 9) %reps(rep10_flag, 10) %reps(rep11_flag, 11)
%reps(rep12_flag, 12) %reps(rep13_flag, 13) %reps(rep14_flag, 14)
%reps(rep15_flag, 15) %reps(rep16_flag, 16) %reps(rep17_flag, 17)
%reps(rep18_flag, 18) %reps(rep19_flag, 19) %reps(rep20_flag, 20)
%reps(rep21_flag, 21) %reps(rep22_flag, 22) %reps(rep23_flag, 23)
%reps(rep24_flag, 24) %reps(rep25_flag, 25) %reps(rep26_flag, 26)
%reps(rep27_flag, 27) %reps(rep28_flag, 28) %reps(rep29_flag, 29)
%reps(rep30_flag, 30) %reps(rep31_flag, 31) %reps(rep32_flag, 32)
```