Alternative Undercount Adjustments for 1990

This document describes the accompanying database files and supplements information in the report *Fixing the Census Until It Breaks,* published by the Michigan Information Center, Department of Management and Budget, November 2000.

   [That report is available via the Internet at http://www.state.mi.us/dmb/mic/]

Overview. The Census Bureau released three alternative sets of undercount adjustments based on the 1990 coverage survey: the "preliminary" adjustments, the "proposed" adjustments, and the "revised" adjustments.

   The "preliminary" adjustments are especially relevant to any analysis of how well the survey and the matching process succeed in measuring net undercount, since they reflect a fairly straightforward application of the adjustment formulas to data derived from the survey, without smoothing and without extensive modification of survey results or of pre-specified procedures. The "proposed" adjustments have special relevance to the undercount debate because they would have been the basis for all official 1990 population data if the Department of Commerce had allowed the Census Bureau to adjust the census. The "revised" adjustments that were released in July 1992 are important because they are widely thought to be the Bureau's best assessment of undercount in 1990.

   Because actual undercount rates are unknown, it is generally impossible to make definitive judgements about the magnitude of error for any given adjustment factor. Nevertheless, it is possible to demonstrate the presence of large errors by comparing alternative adjustments for the same segments of the population.

   The essential fact to keep in mind while comparing these alternative adjustments is that they all pertain to the same census. The actual undercount rate for any given segment of the population is therefore held constant. One of the alternatives might perhaps be correct—or the actual rate of undercount may lie somewhere between or beyond the figures being compared—but they cannot all be correct. *The differences among them are differences in the amount of adjustment error.*

   Because just the *differences* in the amount of adjustment error tend to be quite large, it follows that the errors themselves must be large as well. We still do not know the true undercount rate for any given segment of the population, but we can conclude that the survey upon which all of the adjustments are based has failed to produce accurate and unambiguous measurements of undercount.

Findings. Even though both the preliminary and the proposed adjustments indicate a national undercount of only 2.1 percent, the individual proposed factors differ from the corresponding preliminary factors *by an average of 3.3 percentage points.* The direction of 21 percent of the adjustment factors is reversed, with apparent net undercounts transformed into overcounts or overcounts transformed into undercounts. Another 11 percent of the factors are more than doubled, and 19 percent are reduced to less than half

the size of the preliminary factors. The most extreme example is the poststratum consisting of Asian/Pacific females age 65 and over in rented housing in New York city and White Plains. The preliminary adjustments *decrease* the number of people in this population segment by 34.2 percent, but the proposed adjustments *increase* it by 8.1 percent—a difference of 42.3 percentage points.

The revised adjustment factors differ from the corresponding proposed factors *by an average of 4.2 percentage points.* The direction of the proposed factors is reversed in 32 percent of the cases; the factors are more than doubled in another 22 percent of the cases, and they are reduced to less than half the size of the proposed factors in 16 percent of the cases. For 851 out of 10,192 population segments, the discrepancy between the proposed adjustment factor and the revised adjustment factor is ten percentage points or more. If the Census Bureau had been permitted to implement its proposed adjustments in 1990, for example, the number of black males age 20 to 29 in owner-occupied housing would have been *increased* by 25.0 percent for suburban and outlying portions of certain metropolitan counties in California, Oregon, Washington, Alaska, and Hawaii. But the revised adjustments for 1990 *decrease* this segment of the population by 7.7 percent—a difference of 32.7 percentage points.

Conclusions:

(1) The current approach to adjusting the census for undercount is subject to high levels of error.

(2) The adjustments are sensitive to relatively minor variations in the methodology for deriving adjustments from the underlying survey data.

(3) The alternative adjustments are inconsistent with the assumption that undercount rates are uniform (homogeneous) within each of the designated population segments (poststrata).[1]

The Databases. It is easy to compare the preliminary adjustments with the proposed adjustments, since they both divide the population into the same 1,392 designated segments ("poststrata"). However, the revised adjustments divide the population into only 357 poststrata. Each of the original poststrata overlaps several of the new poststrata, and each of the new poststrata overlaps several of the original poststrata. It is therefore necessary to identify each of the overlaps, resulting in 10,192 smaller population segments with unique combinations of alternative adjustment factors.

The first database ("**DetailedAdj.dbf**") is based upon adjustments for individual poststrata. Each record of the database represents the intersection between an original poststratum and a revised poststratum. Each record includes the preliminary, proposed, and revised adjustment factors that apply to that particular intersection.

---

[1] If the assumption were true, and if the coverage survey measured undercount accurately, then all of the preliminary or smoothed adjustment factors for poststrata that intersect with a given revised poststratum should be similar. But instead they vary quite widely.

The particular population segment to which the record pertains is described in the fields "Region," "Geog," "Tenure," "Race," "Age," and "Sex." The broader population segment to which the preliminary adjustment and proposed adjustment apply are described in the fields "OrigReg," "OrigGeo," "OrigTen," "OrigRace," "OrigAge," and "OrigSex." The broader population segment to which the revised adjustment applies is described in the fields "RevReg," "RevGeo," "RevTen," "RevRace," "RevAge," and "RevSex."

The second database ("**BroadAdj.dbf**") is based on estimated undercount rates for broader groupings of poststrata in which all age/sex categories are combined.[2] Each record of the database represents the intersection between a grouping of original poststrata and a grouping of revised poststrata. Its structure is identical to that of "Detailed.dbf," except that it does not include poststratum codes or fields for age and sex.

List of Fields in Databases

Code357: Code numbers for the revised poststratification system. [DetailedAdj.dbf only]

Code1392: Code numbers for the original poststratification system. [DetailedAdj.dbf only]

CodeReg, CodeGeo, CodeRT, and CodeAS: Portions of Code1392 representing multi-state regions, geographic distinctions within regions, race & tenure, and age & sex. [CodeAS appears in DetailedAdj.dbf only.]

Region and OrigReg (Original Region): These fields are identical. OrigReg indicates the multi-state geographic regions in the original poststratification system. The same designations apply to the intersections of the original and revised poststratification systems, as reflected in the field "Region."

NE	New England
MA	Mid Atlantic
ENC	East North Central
WNC	West North Central
SA	South Atlantic
ESC	East South Central
WSC	West South Central
MT	Mountain
P	Pacific
R	Indian Reservations

---

[2] The broad undercount estimates are taken from: Howard R. Hogan, *Journal of the American Statistical Association*, vol. 88, no. 423, September 1993, pp. 1054, 1058-59.

RevReg (Revised Region): This field indicates the multi-state geographic regions in the revised poststratification system.

NE   Northeast
MW   Midwest
S   South
W   West
ALL   All regions

Geog: This field is a composite of the fields "OrigGeo" and "RevGeo." Each record in the database reflects the intersection of the geographic area indicated in OrigGeo with the geographic area indicated in RevGeo.

OrigGeo (Original Geography): This field indicates the categories into which communities were divided for the original poststratification system.

BCC   Central cities in the New York City, Houston, Dallas, Fort Worth,
            Chicago, Detroit, or Los Angeles-Long Beach PMSA's
BNCC   Communities in above PMSA's outside central cities.
OLCC   Other central cities in MSA's or PMSA's with a city over 250,000 pop.
OLNCC   Communities in above MSA's and PMSA's outside central cities.
LCC   All central cities in MSA's or PMSA's with a city over 250,000 pop.
LNCC   Communities in above MSA's and PMSA's outside central cities.
SCC   Central cities in MSA's or PMSA's without a city over 250,000 pop.
SNCC   Communities in above MSA's and PMSA's outside central cities.
CC   Central cities regardless of size.
NCC   Metropolitan communities outside central cities regardless of size.
MSA   Communities in metropolitan areas regardless of size or type.
OPT   Places with 10,000 or more persons outside metropolitan areas.
OA   Other communities not in metropolitan areas.
NON   All non-metropolitan communities regardless of size.
RES   Indian Reservations.

RevGeo (Revised Geography): This field indicates the categories into which communities were divided for the revised poststratification system.

LU   Urbanized areas >250,000 pop.
SU   Other urbanized areas and other urban communities.
NU   Non-urban communities and rural portions of extended cities.
A   All areas regardless of urban status.

Tenure and RevTen (Revised Tenure): These fields are identical. RevTen indicates the tenure classifications used in the revised poststratification system. The same designations apply to the intersections of the original and revised poststratification systems, as reflected in the field "Tenure."

O	Owner
R	Renter
A	Owner and renter

OrigTen (Original Tenure): This field indicates the tenure classifications used in the original poststratification system. It has the same codes as Tenure and RevTen.

Race and RevRace (Revised Race): These fields are identical. RevRace indicates the race classifications used in the revised poststratification system. The same designations apply to the intersections of the original and revised poststratification systems, as reflected in the field "Race."

AIOR	American Indians on Reservations
API	Asians/Pacific Islanders
B	Blacks
NBH	Non-black Hispanics
NHWO	Non-Hispanic Whites and Non-Hispanic "Other Race"

OrigRace (Original Race): This field indicates the race classifications used in the original poststratification system.

A	Asians/Pacific Islanders
B	Blacks
BH	Blacks and Hispanics
H	Hispanics
HxA	Hispanics except Asians/Pacific Islanders
N	Native Americans
WNAO	Whites, Native Americans, Asians/Pacific Islanders, and "Other Race" (non-Hispanic)
WNO	Whites, Native Americans, and "Other Race" (non-Hispanic)

Age, OrigAge, and RevAge: Codes for these fields are self-explanatory. "Age" indicates the age range for the record in question, while OrigAge and RevAge indicate the broader age ranges to which the original and revised adjustment factors apply. [DetailedAdj.dbf only]

Sex and OrigSex: These fields are identical. OrigSex indicates the sex classifications used in the original poststratification system. The same designations apply to the intersections of the original and revised poststratification systems, as reflected in the field "Sex." [DetailedAdj.dbf only]

RevSex: This field indicates the sex classifications used in the revised poststratification system. [DetailedAdj.dbf only]

M	Male
F	Female
B	Both

Prelim: This field contains the preliminary adjustment factor that applies to the record in question (as well as to other records with the same value for Code1392). The preliminary adjustments reflect fairly straightforward application of the adjustment formulas to data derived from the coverage survey, and they were not subjected to a "smoothing" procedure.

Proposed: This field contains the proposed (or "smoothed") adjustment factor that applies to the record in question (as well as to other records with the same value for Code1392). The proposed adjustments would have been the basis for all official 1990 population data if the Department of Commerce had allowed the Census Bureau to adjust the census. They reflect application of a smoothing procedure to the preliminary adjustments.

Revised: This field contains the revised adjustment factor that applies to the record in question (as well as to other records with the same value for Code357). The revised adjustments were not subjected to a smoothing procedure. They are based upon a somewhat different set of designated population segments, and they reflect modifications of survey results and procedures.

Flag: The Flag field indicates cases where one of the original poststrata applied to two census divisions; the database contains two records for such poststrata, since they sometimes involve slightly different geographic levels for each division.

A value of "dup" indicates that the combination of preliminary, proposed, and revised undercount estimates duplicates another record of the database; records with a flag value of "dup" should be left out of any analysis based on the number of population segments formed from the intersection of the original poststrata with the revised poststrata.

A value of "BOC error" indicates a yet-unresolved contradiction between a poststratum code and the JASA table.[3] The "detailed" database is based on the poststratum codes and "broad" database is based on the JASA table; minor corrections in labeling will be needed when this issue is resolved.

Memo: The "Memo" field contains information about poststrata that apply to two divisions and other comments about individual records.

DiffPS: Absolute value of difference between Preliminary and Proposed ("Smoothed") undercount adjustment.

DiffSR: Absolute value of difference between Proposed ("Smoothed") and Revised undercount adjustment.

---

[3] Poststrata beginning with 627050 would be expected to include non-central cities in small MSA's, since "7" normally denotes non-central cities in all MSA's regardless of size. But the table on page 1059 of Hogan, *op. cit.,* indicates that central cities in small MSA's of the East North Central division are grouped with "other places 10,000+" and "other areas."

DiffPR: Absolute value of difference between Preliminary and Revised undercount adjustment.

MaxDiff: The largest of the three differences listed above.

List of Files

AltAdj90.pdf   Description of the database files.

DetailedAdj.dbf Database with alternative adjustments for 1990 by geography, owner/renter status, race, age, and sex.
DetailedAdj.fpt This file should be placed in the same directory as "DetailedAdj.dbf" in order for information in the "Memo" field to be available.

BroadAdj.dbf   Database with alternative adjustments for 1990 by geography, owner/renter status, and race. (All age and sex categories are combined.)
BroadAdj.fpt   This file should be placed in the same directory as "BroadAdj.dbf" in order for information in the "Memo" field to be available.

For further discussion of alternative adjustments for undercount, see:

Kenneth Darga, *Fixing the Census Until It Breaks,* published by the Michigan Information Center, Department of Management and Budget, November 2000.

Available via the Internet at http://www.state.mi.us/dmb/mic/

Kenneth Darga, State Demographer
Michigan Department of Management and Budget
PO Box 30026
Lansing MI 48909

DargaK@state.mi.us

November 20, 2000