THE RESEARCH AND METHODOLOGY DIRECTORATE

Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing

Laura McKenna

November 2018

United States Census Bureau

census.gov

Disclosure Avoidance Techniques Used for the 1970 through 2010 Decennial Censuses of Population and Housing

Laura McKenna[1]

October 2018

The U.S. Census Bureau conducts the decennial censuses under Title 13 of the U. S. Code with the Section 9 mandate to not "use the information furnished under the provisions of this title for any purpose other than the statistical purposes for which it is supplied; or make any publication whereby the data furnished by any particular establishment or individual under this title can be identified; or permit anyone other than the sworn officers and employees of the Department or bureau or agency thereof to examine the individual reports (13 U.S.C. § 9 (2007))." The Census Bureau applies *disclosure avoidance* techniques to its publicly released statistical products in order to protect the confidentiality of its respondents and their data.

Foreword

John M. Abowd
Chief Scientist and Associate Director for Research and Methodology

Laura McKenna is the former Chief of the Center for Disclosure Avoidance Research and former Chair of the Disclosure Review Board. I asked her to write this overview of the disclosure avoidance methods used in the last five decennial censuses in order to guide contemporary readers through that history in single document and with a coherent vocabulary. In September 2017, the Census Bureau announced that it would undertake a comprehensive disclosure avoidance modernization program beginning with the 2020 Census of Population and Housing. The 2020 census will be protected by modern formal privacy methods—specifically, differential privacy, continuing the long history of innovation in confidentiality protection documented in this review.

[1] This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the author and not necessarily those of the U.S. Census Bureau. Thanks to Connie Citro, Cynthia Clark, Jerry Gates, Nancy Gordon, Michele Hedrick, Bud Pautler, and Sara Sullivan for background, and other assistance in preparing this report.

Contents

1 Introduction

The U.S. Census Bureau's disclosure avoidance (DA) methods have evolved over the past 50 years. The 2020 Census will be the first census protected by a formally private disclosure avoidance system based on differentially privacy methods. This paper summarizes the historical methods the agency used from the 1970 to the 2010 censuses, leading up to the adoption of the modernized disclosure avoidance methods.

This history discusses only publicly available information about the confidentiality protection methods as noted in official documentation of the relevant decennial censuses. All of the information in this summary was taken from historical public sources, except as noted. None of the information in this paper is confidential.

There is no public documentation of the disclosure avoidance methods used in the 1970 Census. This paper relies on an internal Census Bureau planning paper, now cleared for release, that provided a brief description of 1970 methods while highlighting options for disclosure avoidance for the 1980 Census (Zeisset, 1978). There is no information about 1970 methods in the 1970 Technical Documentation nor the 1970 Data User's Guide. Likewise, no documentation of disclosure avoidance was found in public or internal papers for pre-1970 censuses.

The first documented discussion of disclosure avoidance techniques for Group Quarters (GQ) data was for the 2010 Census. There is no discussion of disclosure avoidance for GQ data in public or internal documents for the 1980, 1990, and 2000 censuses.[2]

This paper is focused on decennial census tabular data. A separate paper will outline the history of disclosure avoidance methodology for the Public Use Microdata Samples (PUMS) files. The American Community Survey (ACS) is out of scope for both papers.

This history gleans procedures from various types of publications (Public Law 94-171, Summary Files 1-4) and for different tabulation populations—people in households, people in Group Quarters, 100% ("short form") data, and sample ("long form") data. Complete enumeration (100% data) is used for Public Law (PL) 94-171 (data for redistricting purposes), Summary File (SF) 1, and SF2. Through the 2000 Census, sample data were published in SF3 and SF4. The 2010 Census was the first recent census not to include long form data; the ongoing ACS replaced that data source starting in 2005. All publications were based on both people in households and people in Group Quarters. Tables in SF2 were similar to tables in SF1, but they were iterated by race and Hispanic origin. Tables in SF4 were similar to tables in SF3, but they were iterated by race and Hispanic origin.

[2] Group Quarters data include information about people living in nursing homes, prisons, college dormitories, military barracks, etc. (somewhere other than a household).

The Census Bureau did not publish long form sample data at the lowest level of geography (blocks). As is still the practice with the long form's successor, the American Community Survey, the smallest published geography is the block *group* level.

Rules for special tabulations from the 2000 and 2010 decennial censuses (Appendix A) added another layer of confidentiality protection by restricting releasable special tabulation details.

Notes on Confidentiality in the Technical Documentation of the 1980 through 2010 censuses (Appendices B, C, D, and E) provided high-level information about confidentiality protection in the decennial censuses. Census Bureau researchers published additional details about methods through working papers and symposia and continue to do so. Today, data users can request information or ask questions by contacting disclosure avoidance subject matter experts at DRB_CHAIR@census.gov.

2 1970 Census of Population and Housing

The Census Bureau relied on whole table suppression — not individual cell suppression —as the primary disclosure avoidance method for the 1970 Census. Table suppression was based on the number of people or households in a given area. The method was problematic for several reasons:

1. fewer tables were available for data users;
2. the agency did not provide guidance on how to account for the suppressed data when analyzing the published data;
3. the protections brought by the suppressed whole tables were diminished by the fact that very few complementary tables were suppressed; and
4. cells within an original table could still show an original estimate of 1 or 2.

To limit disclosure risk, the lowest geographic level for which sample data were published was (and still is) the census block group. Census 100% data were published for the lowest possible geographic level: census blocks.

All disclosure avoidance information from the 1970 Census was obtained from an internal document (Zeisset, 1978).

3 1980 Census of Population and Housing

3.1 Why change the methods from the 1970 Census?

Data user dissatisfaction with whole table suppression, along with concerns about the lack of complementary table suppression, lead the Census Bureau to explore new disclosure avoidance methods for the 1980 Census. Researchers discussed options that included random rounding, ordinary rounding, combining areas, and table redesign (Zeisset, 1978). Ultimately, the Census Bureau chose to continue using table suppression, but added additional suppression of complementary tables.

3.2 100% Data (PL 94-171, Summary File (SF) 1, and SF2)

The agency used table-level data suppression for 1980 census tabular data products (Griffin et al., 1989). As in 1970, some tables with cell estimates of 1 or 2 were published. In this case, the counts were replaced with 0s and a flag designating that the cell was suppressed for disclosure, but complementary suppressions were not applied (see Appendix B).

The following univariate (one-variable) counts were not suppressed at any geographic level, the smallest being the block level (for 100% data):

- Population counts by race or Hispanic origin.
- Housing unit counts by vacancy status.
- Occupied housing unit counts by race or Hispanic origin of the householder.

The following rules were applied to data for blocks and above (larger geographical areas) (100% data) and for block groups and above (sample data). A suppression universe is defined as one variable or the cross tabulation of a very small set of variables for which many tables are iterated, such as was the case in SF2 and SF4 (which iterate SF1 and SF3, respectively, across multiple race and Hispanic origin categories).

- **Race or Hispanic origin of householder:**
    - **1 to 14 people:** Detailed characteristics collected for total population, or any suppression universe defined by race or Hispanic origin of the householder, were suppressed if there were 1 to 14 people in the specified suppression universe (for example Black female householders in a given geographic area).
    - **1 to 4 occupied housing units:** Detailed characteristics for people in households for suppression universes defined by the race or Hispanic origin of the householder were suppressed if there were 1 to 4 occupied housing units in the specified group (for example White male householders who rent in a given geographic area).
- **Vacancy status:**
    - **1 to 4 vacant and or occupied housing units:** Detailed housing characteristics for suppression universes defined by vacancy status were suppressed if there were 1

to 4 housing units in the relevant universe (for example Occupied housing units with running water in a given geographic area).

- **Complementary suppression:**
  - **Race and tenure:** Complementary table suppression was applied to protect the additive relationships for race groups that added to a total and for tenure (owners + renters = total) in non-univariate iterated tables. Pre-established rules governed the sequence of choosing complementary table suppressions, for example, suppressing smallest to largest populated tables in a given area.
  - **Cross-geographic areas:** A shortcoming of the 1980 methods was that complementary table suppression was not applied across geographic areas (Griffin et al., 1989). So, for example, if data for one of the three Delaware counties was suppressed, someone could uncover the suppressed tables by subtracting the data for the other two counties from data for the whole state.

3.3 Sample Data (SF3 and SF4)

See Section 3.2 which describes the method for 100% data and was also used for sample data.

3.4 Household Data

See Section 3.2 which describes the method for 100% data including households.

4 1990 Census of Population and Housing

4.1 Why change the methods from the 1980 Census?

Census Bureau researchers developed new disclosure avoidance methods to address three primary shortcomings of 1980 methods:

- dissatisfaction with the reduction in data tables caused by whole table suppression;
- the lack of guidance for data users using the published data in the presence of suppression;
- the disclosure risk issues caused by the lack of complementary suppression across geographic areas (Griffin et al., 1989).

4.2 100% Data (PL 94-171, SF1, and SF2)

Data were published at all geographic levels, including the smallest level, blocks.

The Census Bureau replaced whole table suppression with a new disclosure avoidance technique for the 1990 Census. The new "Confidentiality Edit" used rules-based "data swapping" at the microdata (individual record) level (known then as the "data interchange" method) for 100% data, and the "Blank and Impute" technique for sample data (see Section 4.3).

For 100% data it kept the following unchanged:

- population counts by total, race, Hispanic origin, and people of age 18 and above;
- housing unit counts by total, tenure, and rent/value categories.

To apply the Confidentiality Edit, agency data staff:

1. Selected a small sample of households from the internal census data files, with a higher sampling rate for small blocks.
2. Paired the sampled records according to a set of well-defined matching rules to other records on the file in different geographic locations.
3. Maintained a 1-to-1 matching basis for key variables between each sampled household and its paired household in the other geographic location for the following variables:
    - household size;
    - householder race;
    - householder Hispanic origin;
    - number of people age (18+);
    - tenure (own/rent); and
    - rent/value category.
4. "Interchanged" the paired household records according to a well-defined data interchange (data swapping) operation. The "interchanged" file (swapped file) became the official version of the internal detail file and was used to prepare all subsequent

census data products. A brief discussion of the evaluation of this method is available (Griffin et al., 1989).

4.3 Sample Data (SF3 and SF4)

For all published areas except small block groups, the fact that data were data from a sample was judged to provide adequate disclosure protection.

For small block groups, Census researchers developed what became known as the “Blank and Impute” technique. It involved “blanking” (removing) a sample of the data values (population and housing items) for one of the sample housing units in each small block group and imputing those values using the 1990 Census imputation methodology.

The resulting sample data file (to which disclosure avoidance had been applied) was used to prepare all subsequent census sample data products.

Primarily because of the relatively small increase in imputation rates, the Blank and Impute technique added very little to the level of error of the estimates (Griffin et al., 1989).

4.4 Household Data

The techniques described in Sections 4.2 and 4.3 were used for household data.

5 2000 Census of Population and Housing

5.1 Why change the methods from the 1990 Census?

For the 1990 Census, Census Bureau researchers applied new DA techniques targeted to one of the riskiest potential disclosure categories: small blocks and block groups. For the 2000 Census, staff sought to extend these types of protections beyond small geographies to other increased-risk categories, particularly those at greater risk due to unique cross-tabulations and key variables.

The 2000 Census was the first to allow respondents to choose multiple race categories. The additional detail brought with it a new total of 63 possible race "alone" or "combined" answers. This posed a significant disclosure avoidance challenge and prompted the Census to apply additional protections.

After the 1990 Census the science of disclosure avoidance continued to evolve, and the Census Bureau extended swapping-based protections to the 2000 Census. Swapping replaced Blank and Impute as the primary disclosure protection method for sample data. Swapping had the advantage of removing any absolute assurance that a given record belonged to a given household. It also retained relationships among the variables for each household.

5.2 100% Data (PL 94-171, SF1, and SF2)

Census researchers expanded the swapping techniques inaugurated in 1990 to additional higher-risk categories for the 2000 Census as follows:

- The probability of swapping increased for cross-tabulations of key variables, smaller blocks, and for households that contained members of a race category not found in other households in that block.
- The probability of swapping decreased for blocks already protected with high imputation rates. Records that were entirely imputed were not swapped.
- Every record not totally imputed had a small chance of being swapped.
- Pairs of households that were swapped matched on a second set of key demographic variables. All data products were created from the swapped file.
- For the SF2 dataset, a minimum of 100 people of a race or Hispanic origin group (Hispanic/Non-Hispanic) were required in a geographic area to publish a table iterated by that group for that area. (Zayatz, 2003; Zayatz, 2007). No complementary suppression was applied in order to preserve data quality and save paper.

5.3 Sample Data (SF3 and SF4)

The same disclosure avoidance methods were applied to the sample data at the block group level, with the following differences:

- In addition to decreased swapping rates for block groups with higher imputation rates, rates also decreased in block groups with lower sampling rates.
- For the SF4 dataset, a minimum of 50 people of a race or Hispanic origin group were required in a geographic area to publish a table iterated by that group for that area. (Zayatz, 2003; Zayatz, 2007). No complementary suppression was applied in order to preserve data quality and save paper.
- Sample data required a third list of variables to be held fixed (unswapped). For example, some variables between paired households weren't swapped, such as a householder's American Indian tribe. All three of the lists of variables are confidential.

5.4 Household Data

The household data were protected using data swapping as described in Sections 5.2 and 5.3.

6 2010 Census of Population and Housing

6.1 Why change the methods from the 2000 Census?

The 2010 Census was the first "short form-only" census in recent history. The former sample data long form was replaced by the ongoing "American Community Survey."

6.2 100% Data (PL 94-171, SF1, and SF2)

See Sections 6.4 and 6.5 below.

6.3 Sample Data (SF3 and SF4)

The 2010 Census did not include a long form. The questions previously asked on the long form were transferred to the new American Community Survey.

6.4 Household Data

The swapping procedures for household data were essentially the same as those used for Census 2000 with some refinements to the key variables used to identify unique records and the key variables used to find swapping partners (Zayatz et al., 2010).

6.5 Group Quarters Data

The Census Bureau developed Partially Synthetic Data models to protect Group Quarters (GQ) data. The process involved:

- Blanking some values in at-risk respondent records and using synthetic data techniques to impute those values.
- Using key variable cross tabulation to locate unique records in each tract.
- Blanking unique variable values within each record (compared to other records in the tract).
- Replacing the blanked values with predicted values developed from two types of generalized linear models developed for each county: polytomous regression models and generalized additive models. Variable values were processed in a specific order. Once a value was synthesized, it was used as a predictor for synthesizing other variables.
- Geography and type of GQ were never altered, and age groups <18, 18+ were held fixed.

7 Conclusion

The Census Bureau's disclosure avoidance techniques have evolved over the decades. In 1970 and 1980, the agency used table suppression. Beginning with the 1990 Census, the agency used newer methods, applied at the microdata (individual record) level. In 1990, the "Confidentiality Edit" applied data swapping for 100% (short form) data and blanking and imputation for sample (long form) data.

Beginning in 2000, the Census Bureau extended data swapping to the sample data. While the actual swapping rate and its impact on overall accuracy is confidential, a confidential research study found that the impact in terms of introducing error into the estimates was much smaller than errors from sampling, non-response, editing and imputation.

In 2010, the agency generated partially synthetic data to protect Group Quarters data.

Throughout the decades, the agency published 100% data at the block level and above, and sample data at the block group level and above.

		Table Suppression	Swapping	Blank and Impute	Partially Synthetic Data
1970					
	100% Data	X			
	Sample Data	X			
	Households	X			
1980					
	100% Data	X			
	Sample Data	X			
	Households	X			
1990					
	100% Data		X		
	Sample Data			X	
	Households		X	X	
2000					
	100% Data		X		
	Sample Data		X		
	Households		X		
2010					
	100% Data		X		X
	Households		X		
	Group Quarters				X

8 References

Lauger, A., Wisniewski, W., and McKenna, L. (2015), "Disclosure Avoidance Techniques at the U.S. Census Bureau: Current Practices and Research," Proceedings of the Section on Government Statistics, American Statistical Association, pp. 3630-3642.

Griffin, R., Navarro, F., and Flores-Baez, L. (1989), "Disclosure Avoidance for the 1990 Census," Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 516-521.

United States, Congress, House. United States Code. United States Government Printing Office, 30 Mar. 2010, https://www.gpo.gov/fdsys/pkg/USCODE-2007-title13/pdf/USCODE-2007-title13.pdf.

Zayatz, L. (2002), "SDC in the 2000 U.S. Decennial Census," In: Domingo-Ferrer, J. (eds) Inference Control in Statistical Databases. Lecture Notes in Computer Science, vol 2316, Springer, Berlin, Heidelberg.

Zayatz, L. (2003), "Disclosure Limitation for Census 2000 Tabular Data," Working Paper #15, ECE/Eurostat workshop on statistical data confidentiality, http://www.unece.org/stats/documents/2003/04/confidentiality/wp.15.e.pdf.

Zayatz, L. (2007), "Disclosure Avoidance Practices and Research at the US Census Bureau: An Update," Journal of Official Statistics 23 no. 2: pp. 253-265.

Zayatz, L., Lucero, J., Massell, P., Ramanayake, A. (2010), "Disclosure Avoidance for Census 2010 and American Community Survey Five-year Tabular Data Products," Section on Survey Research Methods, American Statistical Association, pp. 2279-2288.

Zeisset, P. (1978), "Suppression vs. Random Rounding: Disclosure Avoidance Alternatives for the 1980 Census," [https://www.census.gov/library/working-papers/1978/adrm/cdar1978-01.html]

9 Appendix A

Rules for special tabulations from the 2000 and the 2010 Decennial Censuses

1. All Decennial Census special tabulations must be reviewed by the Disclosure Review Board.
2. All cells in any special tabulation must be rounded. The rounding schematic is:

   0 remains 0

   1-7 rounds to 4

   8 or greater rounds to nearest multiple of 5 (i.e., 864 rounds to 865, 982 rounds to 980)

   Any number that already ends in 5 or 0 stays as is.

   This rounding applies to all special tabulations that pertain to the population in households or the population in group quarters -- those done under reimbursable agreement, those done for working papers, tables, professional papers, etc.

   Any totals or subtotals needed should be constructed before rounding. This assures that universes remain the same from table to table, and it is recognized that cells in a table will no longer be additive after rounding.
3. Medians or other quantiles may be calculated as
   A. an interpolation from a frequency distribution of unrounded data (these are not subject to additional rounding), or
   B. as a point quantile. These must be rounded to two significant digits: 12,345 would round to 12,000; 167,452 would round to 170,000. There must be at least 5 cases on either side of the quantile point. It is recognized that the interpolated quantile may indeed be some individual's response, but it is coincidental, not by design.
4. Tables for sample data are only published after weights have been applied to the data, but sometimes both weighted and unweighted counts are used when applying disclosure avoidance rules. Thresholds on universes will normally be applied to avoid showing data for very small geographic areas or for very small population groups, often 50 unweighted cases for sample data. Tables may normally not have more than 3 or 4 dimensions, and mean cell size lower limits may also be required. For example, the mean cell size of each table must be at least 3 cases for 100% data, or 20 weighted cases for sample data).
5. Percents, rates, etc., should be calculated after rounding, but the DRB has granted exceptions to this rule when the numerator and/or denominator of the percent or rate is not shown.
6. Means and aggregates must be based on at least 3 values.
7. The finest level of detail shown for Group Quarters data will be Institutional/ Noninstitutional.

8. For Demographic Profiles from user-defined geographic areas (neighborhoods), all areas must have at least 300 people in them. Using a computer program, the user-defined areas will be compared with standard Census Bureau areas to make sure users cannot obtain data from very small geographic areas by subtraction. If such small areas are found, the boundaries of the user-defined areas must be changed.

10 Appendix B

APPENDIX B
1980, part 1

1980 Census of Population and Housing

Summary Tape File 2.

Technical Documentation

pp. 25-31

prepared by

the U.S. Census Bureau, 1982

SUPPRESSION

To maintain the confidentiality promised respondents and required by law, the Census Bureau takes precautions to make sure that its public data, in print or on tape, do not disclose information about particular individuals or housing units; therefore, the Bureau suppresses tabulations of characteristics for very small groups of people or housing units. On summary tapes, zeroes are entered in suppressed cells. Flag fields which indicate suppression are shown on each record. However, a zero in a cell does not automatically mean suppression. Only by checking the suppression flag can it be determined if the zero in a specific table is suppressed data or an actual count of zero.

This discussion outlines the rules for suppression, how its occurrence can be identified by the user, and how to handle it.

No Suppression

The following counts will never be suppressed:

Total population
Total housing units
Seasonal/migratory housing units
Year-round housing units
Occupied housing units
Vacant year-round housing units
Count of persons and households in each race and Spanish origin group

Primary Suppression

Suppression of Population Characteristics. Characteristics of persons other than race or Spanish origin (e.g., age, relationship) are shown only if there are 15 or more persons in the geographic area. For example, on a record for a census tract with a population of 1 to 14 persons, population characteristics such as age and relationship are suppressed. Only counts for total population and the number of persons within specific race or Spanish origin groups are provided.

However, when the geographic area being summarized has 15 or more persons, no suppression of population characteristics will occur--except possibly when tables are cross-classified by race or Spanish origin. The rules for this type of suppression are outlined below in Suppression of Tables Cross-Classified by Race or Spanish Origin.

Suppression of Year-round Housing Characteristics. Characteristics of year round housing units which are not classified by occupancy status (e.g., number of rooms, plumbing facilities, etc.) are suppressed only when there are fewer than five year-round housing units in the geographic area being tabulated regardless of the number of occupied housing units or the number of persons.

Suppression of Family, Household, or Occupied Housing Characteristics. Characteristics of families, households, or occupied housing units are shown if there are at least five occupied housing units within the geographic area tabulated.

Suppression of Owner or Rental Characteristics. Distributions of data for owners or renters are shown only when the number of owners is at least five and the number of renters is also at least five.

Suppression of Tables Cross-Classified by Race or Spanish Origin. Population and housing characteristics cross-classified by race or Spanish origin are subject to an additional level of scrutiny. On this level the 15 person or five household criteria stated above are also applied to each race or Spanish origin category.

Individual cells of data for specific race or Spanish origin groups are not suppressed when there are 15 or more persons of that group in a geographic area.

The population and housing suppression criteria are applied independently of one another. For example, if there are 16 Spanish origin persons but only four households with Spanish origin householders, the person characteristics will be shown but the family, household, and housing characteristics will be suppressed.

Complementary Suppression

In some cases complementary suppression is applied to prevent the derivation of suppressed data by subtraction. For instance, when a table shows the number of persons in unit for all households and also for renters, there must be at least five owners and five renters for the renter data to be shown; otherwise the characteristics of the owners could be derived by subtracting renter data from data for all households.

Examples of Suppression--A Record

The following example shows two A record tables from the STF 2 Data Dictionary. The first Table (A13) is never suppressed since it is a basic count. The second table (A22) will only be suppressed if there are fewer than 5 year-round housing units in the geographic area tabulated.

Example:

Table A13. SPANISH ORIGIN (2) BY RACE (5)

This table has no suppression

Universe: Persons

This table has no suppression because a count of persons by race is never suppressed.

Spanish origin:
- White
- Black
- American Indian, Eskimo, and Aleut
- Asian and Pacific Islander
- Other

Not of Spanish origin:
- White
- Black
- American Indian, Eskimo, and Aleut
- Asian and Pacific Islander
- Other

Table A22. ROOMS (6)

SUPFLG02 applies to all cells

Universe: Condominium Housing Units

This table will be suppressed when there are 1-4 year-round housing units in the area.

- 1 room
- 2 rooms
- 3 rooms
- 4 rooms
- 5 rooms
- 6 or more rooms

Examples of Suppression--B Record

The following examples show two B record tables from the STF 2 Data Dictionary. The first table (B8) will be suppressed if there are fewer than 15 persons of a particular race or Spanish origin category in the geographic area being summarized. The second table (B28) will have one portion of the table suppressed if there are fewer than five occupied housing units of a particular race or Spanish origin category in the geographic area, and the other portion suppressed if there are fewer than 5 owner and/or renter housing units of the same race or Spanish origin category in the geographic area.

Example:

Table B8. SEX (2) BY AGE (103)

SUPFLGB1 applies to all cells

Universe: Persons

This table will be suppressed when

Total:
- Under 1 year

there are 1-14 persons in the area.

1 year
2 year
3 years
"
"
"
to
"
"
"
99 years
100 to 104 years
105 to 109 years
110 years and over

Female:
Under 1 year
1 year
2 years
3 years
"
"
"
to
"
"
"
99 years
100 to 104 years
105 to 109 years
110 years and over

Table B28. TENURE (2) BY UNITS AT ADDRESS (6)

SUPFLGB3 applies to cells 1-6

SUPFLGB4 applies to cells 7-12

Universe: Occupied Housing Units

The portion of this table indicating total will be suppressed when there are 1-4 occupied housing units in the area.

Total:
1
2
3 and 4
5 to 9
10 or more
Mobile home or trailer

The portion of this table indicating renter occupied will be sup-

Renter occupied:
1
2

pressed if there are 1-4 owner and/or renter occupied housing units in the area.	3 and 4 5 to 9 10 or more Mobile home or trailer

How Suppression Affects the B Record in STF 2--Total suppression of the B record for a specified race or Spanish origin category occurs if the population in that particular category is less than 15 and the number of occupied housing units is less than 5.

B records may be partially suppressed because population and housing suppression criteria are applied independently of each other. For example, in the Asian and Pacific Islander category, if there are 16 persons but only 4 housing units for an area, the person characteristics will be shown but the family, household, and housing characteristics will be suppressed.

Programming with Suppression

Suppressed data cells contain zeroes. To distinguish between zeroes as suppression and zeroes as valid data, occurrences of suppression are identified by a series of flag fields in the geographic identification portion of each logical record. Programmers developing software should include procedures to check these fields for the presence of suppression and, if necessary, to flag the output of any cumulation which includes one or more suppressed fields.

In reviewing the data dictionary, the programmer can determine which suppression flags indicate suppression for particular tables by checking either the table description or the flag description. An example of each follows.

Example: The boxed illustration below is the table description as it appears in the data dictionary. The next portion illustrates the suppression flag to which the table description refers.

TABLE A32 CONTRACT RENT (26)

SUPFLG06 applies to all cells

SUPFLG06	Renter Occupied Housing Unit Suppression Flag A 1 in this field indicates suppression because there are fewer than five renter housing units in the geographic area being summarized or complementary suppression is applied. It will affect the

following tables:
A29
A32
A33

0 No suppression
1 Suppression

Figure 10 below, lists each suppression flag, its location within the record, and the tables or cells within tables which are affected when suppression is applied. The suppression flag field which applies to each table or portion of a table is also identified in the table description in the data dictionary. The flags are located in the geographic identification section of each record in positions 205-210.

Figure 10. Suppression Flags

Name	Begin	Table
Record Type A		
SUPFLG01	205	A9. A10. A18. A21. A56 (cells 1-26), A60 (cells 1-18), A61 (cells 1-4), A62 (cells 1-10)
SUPFLG02	206	A22, A34, A35 (cells 1-9), A36 (cell 1), A37 (cell 1), A38 (cell 1), A39 (cells 1-4)
SUPFLG03	207	A2, A6 (cells 3-4), A14, A15 (cells 1-77), A16 (cells 1-14), A17 (cells 1-10), A19 (cells 1-42), A20, A23-25, A26 (cells 1-7), A35 (cells 19-45), A36 (cell 2), A36 (cells 5-8), A37 (cells 3-5), A38 (cells 3-5), A39 (cells 9-20)
SUPFLG04	208	A15 (cells 78-154), A16 (cells 15-28), A17 (cells 11-20), A19 (cells 43-84), A26 (cells 8-14), A35 (cells 10-18), A36 (cells 3-4), A37 (cell 2), A38 (cell 2), A39 (cells 5-8)
SUPFLG05	209	A27, A28, A30, A31
SUPFLG06	210	A29, A32, A33
Record Type B		
SUPFLGB1	205	B4-B13, B16, B17, B21

SUPFLGB2	206	-
SUPFLGB3	207	B2, B14, B15, B18, B19 (cell 1), B20 (cells 1-9), B22 (cell 1), B23 (cells 1-9), B24 (cell 1), B25 (cell 1), B26 (cells 1-4), B27 (cells 1-14), B28 (cells 1-6), B29 (cells 1-2), B30 (cells 1-2), B31 (cells 1-4), B32 (cells 1-4, B33 (cells 1-3), B34 (cells 1-3), B35 (cell 1), B36 (cell 1)
SUPFLGB4	208	B20 (cells 10-18), B22 (cells 2-3), B23 (cells 10-18), B24 B24 (cells 2-3), B25 (cell 2), B26 (cells 5-8), B27 (cells 15-28), B28 (cells 7-12), B29 (cells 3-4), B30 (cells 3-4), B31 (cells 5-8), B32 (cells 5-8), B33 (cells 4-6), B34 (cells 4-6), B35 (cell 2), B36 (cell 2)
SUPFLGB5	209	B19 (cell 4), B37-B39
SUPFLGB6	210	B19 (cells 2, 3, 5, 6), B40-B42

Evaluating the Effect of Suppression

In most cases, suppressed data values are small (fewer than 5 or 15) except where a large population is affected by complementary suppression. Therefore, in certain noncritical applications, users may simplify programming operations by ignoring suppression and treating suppressed cells as zero cells. However, when geographic entities are being summed to higher levels or new geographic areas are being created, suppression will usually result in a downward bias in the totals.

APPENDIX B
1980, part 2

1980 Census of Housing and Population

User's Guide

PART A. TEXT

pp. 103-106

prepared by

the U.S. Census Bureau, 1982

SUPPRESSION

In order to maintain the confidentiality promised respondents and required by law, the Census Bureau withholds or "suppresses" tabulations of characteristics of very small groups of people or housing units.

In printed and microfiche reports, each suppressed data item is replaced by three dots (...), as illustrated in figure 6-11. On summary tape files, special flags denote suppressed data.

The suppression of certain data may inconvenience data users, especially when they are aggregating data for groups of blocks or tracts. The inconvenience can be lessened if one understands the rules the Census Bureau followed in its disclosure analysis.

Basic Principles Governing Suppression

The Bureau never suppresses certain basic counts, even if an area has a count of only one. These basic counts are as follows:

- Total population
- Total housing units
- Year-round housing units
- Occupied units
- Vacant year-round housing units
- Counts of persons and households in each race and Spanish-origin category

All other data may be suppressed under certain conditions (discussed in detail below), primarily where the size of the population being characterized is less than a specified threshold. The suppression criteria differ for population data and household data. Also, the thresholds are higher for sample estimates than for complete counts. The application of these thresholds results in what is known as "primary suppression." In addition, the Bureau applies "complementary suppression" to avoid the possibility of disclosure by subtraction.

Suppression of Person Characteristics Derived from the Complete Count. Complete counts of person characteristics other than race or Spanish origin (e.g., age or relationship) are shown only if there are 15 or more persons in the geographic area. For example, in data for a block with a population of 1 to 14 persons, population characteristics such as age and relationship are suppressed; tabulations show only counts for total population and the numbers of persons in specific race or Spanish-origin groups.

FIGURE 6-10 **Allocation/Substitution Table Outlines from PC80-1-B**

Table B-1. **Characteristics of Persons Before and After Substitution and Allocation: 1980**—Con.

(For meaning of symbols, see introduction. For definitions of terms, see appendixes A and B)

The State	Persons			Percent		
	After substitution and allocation	After substitution	Before substitution and allocation	After substitution and allocation	After substitution	Before substitution and allocation
HOUSEHOLD RELATIONSHIP						
Total persons						
In households						
Householder						
Spouse						
Child						
Brother or sister						
Parent						
Other relative						
Nonrelatives						
In group quarters						
Inmate of institution						
Other						
SEX						

Table B-4. **Percent of Substitution and Allocation: 1980**—Con.

(For meaning of symbols, see introduction. For definitions of terms, see appendixes A and B)

The State Urban and Rural and Size of Place Inside and Outside SMSA's SCSA's SMSA's Urbanized Areas Places of 1,000 or More Counties	Total persons (number)	Persons substituted for—		Persons with one or more items allocated	Persons with allocated—					Marital status—Persons 15 years and over
		Mechanical failure	Noninterview		Relationship	Sex	Age	Race	Origin	
The State										
URBAN AND RURAL AND SIZE OF PLACE										
Urban										
Inside urbanized areas										
Central cities										
Urban fringe										
Outside urbanized areas										
Places of 10,000 or more										
Places of 2,500 to 10,000										
Rural										
Places of 1,000 to 2,500										
Other rural										
INSIDE AND OUTSIDE SMSA's										

FIGURE 6-11 **Report Table Illustrating Suppression[1]**

Table 2. **Characteristics of Population and Housing Units, by Blocks: 1980**—Con. **Bonneville County, Idaho**

[For meaning of symbols, see introduction. For definitions of terms, see appendixes A and B]

Blocks Within Census Tracts or Block Numbering Areas (BNA's)	Persons						Year-round housing units				Occupied housing units									
											Owner		Renter							
	Total	Black	Asian and Pacific Islander	Spanish origin	Under 18 years	65 years and over	Total	One unit attached	10 or more units attached	Mean rooms	Total	Mean value (dollars), specified owner	Total	1.01 or more persons per room	Lacking complete plumbing for exclusive use	Mean contract rent (dollars), specified renter	1.01 or more persons per room	Lacking complete plumbing for exclusive use	Persons per unit	One-person households	Family householder, no spouse present, persons under 18
Idaho Falls city:																				
BNA 9901*	7827	30	100	85	3243	257	2496	2196	187	6.4	1834	59700	474	17		231	62	3	3.4	269	147
201	114	3	1		46	4	39	36	–	5.4	23	39900	12	3	–	229	3	–	3.3	5	5
202	50	–	–	–	22	–	16	16	–	6.5	12	47200	3	...	...	...	–	–	3.3	1	1
203	48	–	–	–	13	2	17	17	–	6.8	14	32400	1	...	...	...	–	–	3.2	–	–
204	114	4	5	4	49	2	38	38	–	5.3	23	31500	10	–	–	240	2	–	2.5	3	5
205	77	–	–	5	25	–	28	28	–	4.8	19	30400	7	1	–	210	1	–	3.8	5	1
206	56	–	–	–	22	3	16	16	–	6.4	13	48000	2	...	...	...	1	–	3.7	–	–
207	60	–	–	–	29	1	16	16	–	7.1	14	53300	1	...	...	...	1	–	4.0	1	–
208	97	–	1	–	45	3	25	25	–	6.7	22	49500	2	...	...	...	–	–	4.0	1	2
209	58	–	–	–	27	–	16	16	–	6.3	14	46800	1	...	...	...	1	–	3.9	–	2
210	45	–	3	–	20	1	13	12	–	5.3	8	35900	5	2	–	199	2	–	3.5	2	–
211	99	–	1	–	36	3	43	34	–	5.2	20	51000	19	–	–	235	–	–	2.5	13	5
301	201	–	3	–	66	6	39	33	–	6.7	33	39700	1	...	...	...	5	–	2.6	4	2
302	122	–	–	–	51	7	35	35	–	6.8	30	41400	4	...	...	...	–	–	2.6	1	2
303	132	–	–	–	55	1	37	37	–	7.2	32	43800	4	...	...	...	–	–	2.7	3	4
304	111	–	5	–	52	3	29	29	–	6.5	24	41200	3	...	...	...	2	–	4.1	3	1
305	59	–	–	–	29	–	18	16	–	5.8	15	45000	1	...	...	...	–	–	2.7	1	1
306	225	–	6	1	88	5	69	69	–	6.8	62	59700	4	...	...	...	1	–	3.4	5	4
307	72	–	–	2	35	2	19	19	–	7.3	19	69700	–	–	–	–	–	–	3.8	1	–
308	103	–	–	1	43	2	29	27	–	7.1	26	75500	1	...	...	...	–	–	3.8	–	–
309	47	–	1	1	12	–	18	12	–	5.7	9	62500	9	–	–	294	–	–	2.6	3	–
310	54	–	5	5	22	1	17	17	–	8.0	16	70900	–	–	–	–	–	–	3.4	1	1
311	46	–	3	–	17	2	13	13	–	7.2	13	77500	–	–	–	–	–	–	3.5	–	1
312	79	4	–	–	35	4	29	23	–	7.3	21	69900	1	...	...	...	–	–	3.6	1	1
313	109	6	–	4	50	2	27	27	–	6.7	22	61000	5	2	–	254	3	–	4.0	1	–
314	91	–	1	–	40	–	27	27	–	7.1	25	66800	2	...	...	...	–	–	3.4	3	1
315	132	–	–	–	63	1	35	35	–	6.7	33	61300	1	...	...	...	3	–	3.9	2	3
316	132	–	4	3	63	2	34	34	–	7.1	32	56200	2	...	...	...	–	–	3.9	3	2
317	82	–	5	1	44	2	19	19	–	7.9	19	73800	–	–	–	–	1	–	4.3	–	2
318	71	–	2	–	40	–	15	15	–	8.5	13	82900	1	...	...	...	–	–	5.1	1	–
319	72	–	–	2	49	–	16	16	–	8.1	16	81300	–	–	–	–	–	–	4.5	–	–
320	59	–	–	–	28	1	17	17	–	7.6	15	80700	1	...	...	...	–	–	3.7	2	–
321	46	–	–	–	23	1	10	10	–	7.8	8	81400	1	...	...	...	–	–	5.1	–	1
322	76	–	–	–	34	1	22	22	–	6.5	19	45700	3	...	...	...	–	–	3.6	1	1
323	50	–	–	–	26	–	13	13	–	6.0	9	47500	3	...	...	...	–	–	4.2	1	1
324	58	–	2	2	24	2	16	16	–	7.3	15	68800	1	...	...	...	–	–	3.6	1	1
401	69	–	–	–	40	–	16	16	–	5.4	14	48300	1	...	...	...	4	–	4.6	–	1
402	129	1	1	–	60	3	32	32	–	6.3	27	48600	5	–	–	277	5	–	4.0	2	3
403	68	–	–	–	18	2	16	15	–	4.6	9	43200	6	–	–	192	1	–	2.7	3	–
404*	139	–	–	5	66	3	37	36	–	6.1	30	65300	7	–	1	275	2	1	3.8	5	1
407*	30	–	–	–	12	–	10	10	–	6.6	8	112500	–	–	–	–	–	–	3.8	–	–
408	143	–	–	1	61	2	40	40	–	6.6	31	50700	7	1	–	312	2	–	3.8	1	3
409	112	–	1	–	43	1	35	35	–	6.3	29	44800	4	...	...	...	–	–	3.4	2	1
410	125	–	–	–	52	2	32	32	–	6.2	25	43100	7	1	–	296	3	–	3.9	1	–
411	101	–	–	1	49	1	27	27	–	6.2	19	45100	6	1	–	277	1	–	4.0	2	–
412	81	–	–	1	36	–	24	24	–	6.4	22	47600	2	...	...	...	1	–	3.4	2	1
413	91	–	–	–	45	2	24	24	–	5.8	23	49900	1	...	...	...	2	–	3.8	2	1
414*	578	4	2	3	239	24	288	117	130	5.0	81	52800	122	1	1	154	3	1	2.8	38	40
415	67	–	–	–	28	–	18	17	–	6.3	17	50500	1	...	...	...	–	–	3.7	1	2
416	92	–	–	–	39	1	27	27	–	6.6	25	49500	1	...	...	...	–	–	3.5	3	3
417	41	–	–	–	17	2	13	13	–	5.5	12	47600	1	...	...	...	–	–	3.9	2	–
418	58	–	–	–	23	–	21	9	–	5.0	10	53200	7	–	–	226	1	–	3.4	–	1
419	16	–	–	–	7	–	8	8	–	6.8	...	...	...	...	...	...	...	...	...	...	...
502	54	–	–	–	17	1	20	19	–	5.9	17	42600	3	...	...	...	...	...	2.9	3	2
503	67	–	–	–	25	–	20	20	–	6.3	18	49400	2	...	...	...	1	–	3.4	2	1
504	66	–	–	–	28	1	20	20	–	7.0	20	51800	–	–	–	–	–	–	3.3	3	2
505	77	–	–	3	32	1	24	24	–	7.0	22	49300	1	...	...	...	–	–	3.3	3	–
506	93	–	–	–	33	0	26	26	–	7.2	25	54500	–	–	–	–	1	–	3.6	–	–
507	75	2	–	6	33	2	24	24	–	6.9	22	55700	–	–	–	–	–	–	3.4	1	2
508	87	–	–	1	30	7	26	26	–	7.1	25	48100	1	...	...	...	–	–	3.5	–	1
509	40	1	–	–	22	2	13	12	–	6.7	13	47900	–	–	–	–	1	–	3.0	1	–
601	231	–	4	1	68	21	108	106	1	5.6	75	54800	23	–	–	268	–	–	2.4	25	3
602	183	–	4	4	40	29	77	76	–	5.2	51	57200	25	–	–	330	–	–	2.4	23	4
603	122	–	–	1	59	2	31	31	–	7.1	27	59400	3	...	...	...	2	–	4.1	1	–
604	79	–	–	–	26	5	24	22	–	6.5	21	52900	3	...	...	...	–	–	3.3	2	–
605	108	–	1	–	50	3	28	27	–	6.9	26	49400	1	...	...	...	2	–	3.9	3	3
606	52	–	–	–	5	22	28	28	–	7.3	26	–	–	–	–	–	–	–	2.0	3	–
607*	250	2	5	3	64	8	123	68	53	4.6	27	54900	82	2	–	210	3	–	2.3	37	7
608	42	3	–	–	5	1	25	20	1	4.3	1	...	21	...	...	214	–	1	1.9	8	2
701	33	–	–	–	11	2	9	9	–	8.8	9	90900	–	–	–	–	–	–	3.4	–	–
702	65	–	–	–	22	11	36	13	2	6.3	13	77700	12	–	–	342	–	–	2.8	7	1
703	50	–	1	–	25	–	11	11	–	9.2	10	79300	1	...	...	...	–	–	4.5	–	1
704	68	–	2	–	29	4	19	19	–	8.9	19	80400	–	–	–	–	–	–	3.6	–	–
705	134	–	7	–	49	3	37	37	–	8.5	36	90100	–	–	–	–	–	–	3.7	–	–
706	72	–	–	–	24	2	21	21	–	7.9	19	94000	2	...	...	...	–	–	3.4	–	–
707	35	–	2	–	15	–	10	10	–	7.6	10	107000	–	–	–	–	–	–	3.5	–	1
708	86	–	–	6	27	4	23	23	–	7.9	23	98300	–	–	–	–	1	–	3.7	–	1
709	60	–	–	–	28	–	14	13	–	9.2	13	92500	1	...	...	...	–	–	3.1	–	1
710*	599	–	24	6	289	12	132	132	–	8.1	142	88400	4	...	...	...	2	–	4.1	4	3
711	37	–	–	–	20	–	8	8	–	8.4	8	116900	–	–	–	–	–	–	4.6	–	–
801*	3	–	–	–	...	...	2	...	...	...	...	...	...	...	...	...	...	...	...	...	...
BNA 9902*	4968	15	49	150	1449	537	1978	1423	152	4.9	1080	37300	774	43	8	185	85	10	2.6	446	181
101	99	–	–	1	28	2	46	4	–	4.5	37	115000	–	–	–	–	2	–	2.7	3	1
104	201	–	–	2	76	6	65	60	–	5.6	45	31300	18	2	–	228	2	–	3.2	11	10
105*	291	–	2	5	66	126	102	21	62	4.2	4	...	95	...	...	125	1	–	2.8	49	23
106	125	–	1	9	32	7	43	43	–	4.9	24	29100	16	1	–	202	4	–	3.1	1	4
107	101	–	2	–	34	4	39	39	–	4.7	18	28600	19	–	–	211	1	–	2.7	8	8

Idaho—14 **SELECTED AREAS IN IDAHO** **BLOCK STATISTICS**

[1] Three dots "..." mean not applicable, or that the data are being withheld to avoid disclosure of information for individuals. A dash "—" represents zero or a percent which rounds to less than 0.1.

The 15-person criterion applies only to the applicable "critical universe," in this case, total persons. These rules would not prevent the display of data showing, for example, that there are 2 persons 65 years old or over, as long as the area includes 15 or more persons in total.

Suppression of Family, Household or Housing Unit Characteristics Derived from the Complete Count. The threshold for family, household or housing unit data is 5, not 15. Characteristics of year-round housing units are shown if the area includes five or more year-round units. Characteristics of occupied housing units, households, or families are shown if the area includes five or more occupied units. Similar thresholds govern characteristics of owners and renters, except that the Bureau must also avoid complementary disclosure. For example, if an area includes 10 occupied housing units, 8 rented and 2 owner-occupied, any data provided for the total and for renters would be derivable for the 2 owners by subtraction. Therefore, most characteristics of owners and renters are shown only if the area includes at least five owners and five renters.

The suppression criteria for population and housing are applied independently of each other. For example, if an area includes 16 persons but only 4 housing units, the person characteristics will be shown but family, household and housing characteristics will be suppressed.

Suppression of Complete-Count Tables Cross-Classified by Race or Spanish Origin. Population and housing characteristics cross-classified by race or Spanish-origin are subject to an additional level of scrutiny. The 15-person or 5-household criteria stated above for complete-count data also apply to each race or Spanish-origin category. For example, a table of race by age for a geographic area that has 200 persons—124 White; 14 Black; 10 American Indian, Eskimo and Aleut; and 52 Asian and Pacific Islanders—shows actual age data for Whites and the Asian and Pacific Islander group, but not for the 2 groups with fewer than 15 persons.

On the other hand, if only one of the race categories in the foregoing example had more than zero but fewer than 15 persons, the Bureau would have employed complementary suppression to avoid the derivation of data about that one race by subtraction. Figure 6-12 illustrates the fact that a second race group would have been complementarily suppressed, generally the "other race" category (or "race, n.e.c." in sample data); but, since that group has no population in the example, the next smallest race group is targeted for complementary suppression. (Complementary suppression is not always obvious since most published tables omit the "other race" category, thereby requiring that it be derived by the subtraction of data for specified races from the total. If one of the specified races is suppressed, characteristics for the "other race" category can no longer be derived.)

Suppression in Sample Data. Thresholds applied to sample estimates are double those applied to complete counts, i.e., 30 persons/10 households instead of 15 persons/5 households. Otherwise, the rules are analogous. The size of the sample in the area (50 percent or 16 2/3 percent) does not affect the thresholds. Note that it is a sample estimate that is tested relative to the threshold; for example, an area with 30 persons in the complete count but only 25 persons estimated in the sample would have its sample characteristics suppressed.

Suppression of sample data normally should be of less concern than complete-count suppression, since any sample number small enough to be suppressed would have been unreliable anyway.

Illustrations of Suppression

Users occasionally misunderstand census suppression rules since they expect suppression to be on a cell-by-cell basis (e.g., every number less than 15 suppressed) rather than a critical universe basis (e.g., category cells suppressed only if the total population is less than 15 persons).

FIGURE 6-12 **Hypothetical Table Illustrating Suppression in Complete-Count Data**

Race by Age	Data Before Suppression	Data As Made Public	
Total	200	200	
Under 5 years	10	10	
5 to 17 years	20	20	
18 to 64 years	140	140	
65 years and over	30	30	
White	124	124	
Under 5 years	7	7	
5 to 17 years	11	11	
18 to 64 years	90	90	
65 years and over	16	16	
Black	14	14	
Under 5 years	1	Suppressed	Primary Suppression
5 to 17 years	1	Suppressed	
18 to 64 years	10	Suppressed	
65 years and over	2	Suppressed	
American Indian, Eskimo, and Aleut	62	62	
Under 5 years	2	Suppressed	Complementary Suppression
5 to 17 years	8	Suppressed	
18 to 64 years	40	Suppressed	
65 years and over	12	Suppressed	
Asian and Pacific Islander	0	0	
Under 5 years	0	0	
5 to 17 years	0	0	
18 to 64 years	0	0	
65 years and over	0	0	
Other	0	0	
Under 5 years	0	0	
5 to 17 years	0	0	
18 to 64 years	0	0	
65 years and over	0	0	

Several aspects of these suppression criteria are illustrated by the following examples. The number of owners and the number of renters are critical universes for certain tabulations. For complete-count housing value data to be shown, an area must include at least five owners, and for rent data, at least five renters. For plumbing facilities data to be shown for renters, an area must include both five owners and five renters, since owner data would be derivable by subtracting renter data on plumbing facilities from corresponding data for all occupied units. On the other hand, if a table shows only the number of owners and renters (no characteristics), the only requirement is that there be at least five occupied units. A table cross-tabulating plumbing facilities and persons per room is also subject only to the five occupied unit threshold, since plumbing facilities and persons per room are not involved in the definition of critical universes.

The user need not memorize these criteria, only understand the general principles. More explicit detail about which criteria apply to which data cells is provided in summary tape technical documentation.

Certain reports show data for race or Spanish-origin groups only if the race or origin group in the given area meets a certain threshold (usually 400 or 1,000 persons). The purpose is not to avoid disclosure, but merely to reduce publication costs. Complete-count data are available on STF 2 for race or Spanish-origin groups with 15 or more persons and 5 or more households, and sample estimates are available on STF 4 for similar groups with 30 or more persons and 10 or more households—each in at least as much detail as is available for larger groups in print.

Programming with Suppression

Suppression Indicators. Suppressed data cells on summary tape files contain zeroes. To distinguish between zeroes as suppression and zeroes as valid data, occurrences of suppression are identified by a series of flag fields in the geographic identification portion of each data record. Programmers developing software should include procedures to check these fields for the presence of suppression and, if necessary, to flag the output of any cumulation which includes one or more suppressed fields.

Technical documentation for each STF defines the relationship between data tables and the suppression flags in two ways. First, the description of each flag, in the identification section of each record, lists each table or part which is governed by that suppression flag. Second, each table description indicates which suppression flag applies.

Consequences of Ignoring Suppression. In most cases, suppressed data values are small (less than 30 in any case). A sizable percentage of individual suppressed data cells were actual zeroes before suppression, although a large population may be affected by complementary suppression. Therefore, in certain applications that are not critical, users may simplify programming operations by ignoring suppression and treating suppressed cells as zero cells.

However, if the user is adding up blocks or enumeration districts to derive tables for specially defined areas, ignoring suppression will result in a downward bias in the totals. A user can gauge the impact of the downward bias if the universe of the tabulation is one that is never suppressed, as the following example illustrates. An age distribution for all persons may be suppressed, but the total number of persons is never suppressed. Therefore, if an age distribution is cumulated for a user-defined group of blocks, the total population should also be cumulated. If the sum of persons in all age categories for the group of blocks is 425 and the total population is 460, one can conclude that there were 35 people in blocks where the age distribution was suppressed.

APPENDIX B
1980, part 3

1980 Census of Population

General Population Characteristics

UNITED STATES SUMMARY

Introduction, pp. IV-V

prepared by

the U.S. Census Bureau, 1983

SYMBOLS AND GEOGRAPHIC ABBREVIATIONS

The following symbols and geographic abbreviations are used in the tables:

- A dash "—" represents zero or a percent which rounds to less than 0.1.
- Three dots "..." mean not applicable, or that the data are being withheld to avoid disclosure of information for individuals or housing units. (For further information on disclosure, see the section below on "Suppression of Data for Confidentiality.")
- CDP is census designated place.
- SCSA is standard consolidated statistical area.
- SMSA is standard metropolitan statistical area.

SUPPRESSION OF DATA FOR CONFIDENTIALITY

To maintain the confidentiality promised respondents and required by law, the Census Bureau takes precautions that its published data do not disclose information about specific individuals and housing units. To accomplish this, the Bureau suppresses data for characteristics which are based on a small number of persons and/or housing units in the geographic area. Under certain conditions, both primary and complementary suppression, as defined below, may take place.

The general rules of primary suppression of sample data are as follows: estimates of total population by race and Spanish origin are never suppressed; other characteristics for persons are shown only if there are 30 or more persons in the geographic area; estimates of total housing units, vacant housing units, year-round housing units, and occupied housing units are never suppressed; characteristics of year-round housing units which are not classified by occupancy status are shown only when there are 10 or more year-round housing units in the geographic area; characteristics of families, households, or occupied housing units are shown only if there are at least 10 occupied housing units within the geographic area; and distributions of data for owners or renters are shown only where the number of owners is at least 10 and the number of renters is also at least 10. These primary suppression criteria are applied independently of one another. The comparable figures for complete count (100 percent) data are 15 or more persons and 5 or more housing units of the specified type.

Population and occupied housing unit characteristics cross-classified by race or Spanish origin (of the householder in the case of occupied housing units) are subject to an additional level of examination. This requires the 30 person or 10 housing unit criterion stated above be applied individually to each race or Spanish origin category.

Finally, complementary suppression is applied to prevent the derivation of primary suppressed data by subtraction. For example, housing unit data shown by tenure may require complementary suppression when the number of owner-occupied or renter-occupied housing units is less than 10.

11 Appendix C

APPENDIX C

1990

1990 Census of Population

General Population Characteristics

Delaware

Appendix C, page 1

prepared by

the U.S. Census Bureau, 1992

Note that the same documentation appeared in all 1990 Summary Files.

APPENDIX C.
Accuracy of the Data

CONTENTS

CONFIDENTIALITY OF THE DATA

To maintain confidentiality required by law (Title 13, United States Code), the Bureau of the Census applies a confidentiality edit to assure published data do not disclose information about specific individuals, households, and housing units. The result is that a small amount of uncertainty is introduced into some of the census characteristics to prevent identification of specific individuals, households, or housing units. The edit is controlled so that the counts of total persons, totals by race and American Indian tribe, Hispanic origin, and age 18 years and over are *not affected* by the confidentiality edit and are published as collected. In addition, total counts for housing units by tenure are not affected by this edit.

The confidentiality edit is conducted by selecting a sample of census households from the 100-percent data internal census files and interchanging its data with other households that have identical characteristics on a set of selected key variables but are in different geographic locations within the same State. To provide more protection for "small areas," a higher sampling rate was used for these areas. The net result of this procedure is that the data user's ability to obtain census data, particularly for small areas and subpopulation groups, has been significantly enhanced.

12 Appendix D

APPENDIX D

2000

Census 2000 Summary File 3 Technical Documentation Chapter 8, pp. 8-4 and 8-5 prepared by the U.S. Census Bureau, 2002

Note that the same documentation appeared in all 2000 Summary Files.

CONFIDENTIALITY OF THE DATA

The Census Bureau has modified or suppressed some data in this data release to protect confidentiality. Title 13 United States Code, Section 9, prohibits the Census Bureau from publishing results in which an individual can be identified. The Census Bureau's internal Disclosure Review Board sets the confidentiality rules for all data releases. A checklist approach is used to ensure that all potential risks to the confidentiality of the data are considered and addressed.

Title 13, United States Code. Title 13 of the United States Code authorizes the Census Bureau to conduct censuses and surveys. Section 9 of the same Title requires that any information collected from the public under the authority of Title 13 be maintained as confidential. Section 214 of Title 13 and Sections 3559 and 3571 of Title 18 of the United States Code provide for the imposition of penalties of up to 5 years in prison and up to $250,000 in fines for wrongful disclosure of confidential census information.

Disclosure limitation. Disclosure limitation is the process for protecting the confidentiality of data. A disclosure of data occurs when someone can use published statistical information to identify an individual who provided information under a pledge of confidentiality. Using disclosure limitation procedures, the Census Bureau modifies or removes the characteristics that put confidential information at risk for disclosure. Although it may appear that a table shows information about a specific individual, the Census Bureau has taken steps to disguise the original data while making sure the results are still useful. The techniques used by the Census Bureau to protect confidentiality in tabulations vary, depending on the type of data.

Data swapping. Data swapping is a method of disclosure limitation designed to protect confidentiality in tables of frequency data (the number or percentage of the population with certain characteristics). Data swapping is done by editing the source data or exchanging records for a sample of cases when creating a table. A sample of households is selected and matched on a set of selected key variables with households in neighboring geographic areas that have similar characteristics (such as the same number of adults and the same number of children). Because the swap often occurs within a neighboring area, there is no effect on the marginal totals for the area or for totals that include data from multiple areas. Because of data swapping, users should not assume that tables with cells having a value of 1 or 2 reveal information about specific individuals. Data swapping procedures were first used in the 1990 census and were also used for Census 2000.

13 Appendix E

APPENDIX E

2010

2010 Census Advance Group Quarters Summary File-- Technical Documentation Chapter 5, page 6 prepared by the U.S. Census Bureau, 2011.

Note that the same documentation appeared in all 2010 Summary Files.

CONFIDENTIALITY OF THE DATA

The Census Bureau has modified some data in this data release to protect confidentiality. Title 13 U.S. Code, Section 9, prohibits the Census Bureau from publishing results in which an individual's data can be identified.

The Census Bureau's internal Disclosure Review Board monitors the disclosure review process and sets the confidentiality rules for all data releases. A checklist approach is used to ensure that all potential risks are considered and addressed. A list of possible concerns is created and the Disclosure Review Board makes sure that the appropriate steps are taken to assure the confidentiality of the data.

Title 13 U.S. Code

Title 13 of the U.S. Code authorizes the Census Bureau to conduct surveys and censuses and mandates that any information obtained from private individuals and establishments remains confidential. Section 9 of Title 13 prohibits the Census Bureau from releasing "any publication whereby the data furnished by any particular establishment or individual under this title can be identified." Section 214 of Title 13, as modified by the Federal Sentencing Reform Act, imposes a fine of not more than $250,000 and/or imprisonment of not more than 5 years for publication or communication in violation of Section 9.

Disclosure Avoidance

Disclosure avoidance is the process of disguising data to protect confidentiality. A disclosure of data occurs when someone can use published statistical information to identify an individual who provided information under a pledge of confidentiality. Using disclosure avoidance, the Census Bureau modifies or removes all of the characteristics that put confidential information at risk for disclosure. Although it may appear that a table shows information about a specific individual, the Census Bureau has taken steps (such as data swapping) to disguise the original data while making sure the results are useful.

Data Swapping

Data swapping is a method of disclosure avoidance designed to protect confidentiality in tables of frequency data (the number or percentage of the population with certain characteristics). Data swapping is done by editing the source data or exchanging records for a sample of cases. A sample of households is selected and matched on a set of selected key variables with households in neighboring geographic areas (geographic areas with a small population) that have similar characteristics (same number of adults, same number of children, etc.). Because the swap often occurs within a geographic area with a small population, there is no effect on the marginal totals for the geographic area with a small population or for totals that include data from multiple geographic areas with small populations. Because of data swapping, users should not assume that tables with cells having a value of one or two reveal information about specific individuals.