Detailed Demographic and Housing Characteristics File A (Detailed DHC-A) Proof of Concept

U.S. Census Bureau

Released January 31, 2023

The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release.

DRB Approval Numbers: CBDRB-FY22-DSEP-005 & CBDRB-FY23-0158

Contents

About the Detailed Demographic and Housing Characteristics File A (Detailed DHC-A) Proof of Concept 4

What this product includes 4

What this product does not include 4

Introduction 5

About the 2020 Census Detailed DHC-A 5

About the Detailed DHC-A Proof of Concept 5

Technical Background 6

Disclosure Avoidance 6

Differential Privacy 6

Differences between the TopDown and SafeTab-P Differentially Private Algorithms 6

Table 1. Differences between the TopDown and SafeTab-P Differentially Private Algorithms 6

Privacy-Loss Budget 7

Data Used in the Proof of Concept 8

Race and Hispanic Origin 8

Sex and Age 9

Table 2. Categories for Sex by Age Tables 9

Geography 10

Minimum Population Counts 10

Table 3. Minimum Population Counts for Detailed Groups in 2010 Published Data Compared to 2010 Noise Infused Data in the Proof of Concept 11

Detailed DHC-A Product Design 11

Adaptive Design 11

Figure 1. Adaptive Design 12

Data Accuracy and Margins of Error 12

Table 4. Examples of Margin of Error of 50 for Groups of Varying Sizes 13

Table 5. Detailed DHC-A Minimum Noise Infused Population Counts and Margins of Error (MOE) by Geography 13

Table 6. Example of Aggregated MOEs for Sex by Age Tables 14

Table 7. Aggregated Margins of Error by Geography and Sex by Age Data 14

Example Tables 15

Limitations 15

Table 8. Limitation 1a 16

Table 9. Limitation 1b. 16
Table 10. Limitation 2a. 17
Table 11. Limitation 2b. 17
Table 12. Limitation 2c. 17
Table 13. Limitation 3a. 18
Table 14. Limitation 3b. 18
Table 15. Limitation 3c. 19
Table 16. Limitation 4a. 19
Future Steps 19
How to Provide Feedback 19
Detailed DHC-B 20
Table 17. Household Type 20
Table 18. Tenure 20
Table 19. Tenure 21

About the Detailed Demographic and Housing Characteristics File A (Detailed DHC-A) Proof of Concept

The U.S. Census Bureau (Census Bureau) operates under Title 13 of the U.S. Code, which prohibits the publication of any results in which an individual's data can be identified. The Census Bureau has been, and continues to be, a leader in the science of protecting respondent data. To ensure protection of respondent data in present and future data releases, the Census Bureau has launched a major modernization of disclosure avoidance techniques by deploying differential privacy, a type of formal privacy, for the 2020 Census products. Differential privacy is designed to withstand modern re-identification threats to protect 2020 Census data products, while maximizing the availability and utility of published census data. For more information, please visit 2020 Census Disclosure Avoidance Modernization.

This Proof of Concept uses 2010 Census enumerated values – that is, the values obtained during the 2010 Census data collection effort – and applies one of the new 2020 Disclosure Avoidance System (DAS) differentially private algorithms, helping to ensure the confidentiality of respondent data. The Census Bureau created this Proof of Concept to help data users understand how differential privacy may impact the 2020 Detailed DHC-A. This Proof of Concept includes proposed content, which has not been finalized by the Census Bureau, for the final 2020 Detailed DHC-A. The information and examples provided in the Proof of Concept reflect the proposed disclosure avoidance settings, and therefore this documentation is not intended to be used with the final 2020 Detailed DHC-A. Updated documentation will be released with that product.

The Census Bureau is also seeking positive, negative, and neutral public comment on this Proof of Concept. For more information on submitting comments, see How to Provide Feedback.

What this product includes

This product includes descriptions of what data users can expect to see in the 2020 Census Detailed DHC-A, including table content and geographic content; information on the proposed minimum population counts (or thresholds) and accuracy targets; metrics; and information about inconsistencies and limitations of the data. The data in this product are based on 2010 Census data and represent the tables planned for publication in the 2020 Census Detailed DHC-A.

This product also includes a brief description of the 2020 Census Detailed DHC-B. For more information, see Detailed DHC-B.

What this product does not include

Unlike the demonstration products previously released for other 2020 Census data products, this product does not include summary files or a complete dataset. The 2020 Census had improvements in the design, processing, and coding of the 2020 Census Hispanic origin and race questions, allowing for the release of data for almost five times as many detailed race and ethnicity groups than was possible in 2010. The increase in the amount of detailed race and ethnicity data that will be available in the 2020 Census Detailed DHC-A is not reflected in the Proof of Concept because this product is based on 2010 Census data.

Introduction

About the 2020 Census Detailed DHC-A

The 2020 Census Detailed DHC-A (in combination with the Detailed DHC-B) is the successor to the 2010 Census Summary File 2 and the 2010 Census American Indian and Alaska Native Summary File. The 2020 Detailed DHC-A provides population counts and sex and age statistics for approximately 370 detailed racial and ethnic groups and 1,200 American Indian and Alaska Native tribes and villages. Data will be available for the nation, states, counties, tracts, places, and American Indian/Alaska Native/Native Hawaiian (AIANNH) areas.

In developing the Detailed DHC-A, the primary goal was to produce accurate data for our nation's myriad detailed racial and ethnic groups in an equitable way across all major race and ethnicity categories. The goal was also to provide as many population counts as confidentiality protections will allow at various geography levels, for as many detailed racial and ethnic groups and American Indian and Alaska Native tribes and villages as possible. This is an improvement from 2010, where Asian, Native Hawaiian and Other Pacific Islander, and Hispanic origin groups needed a minimum population count of 100 to have published counts at any given geography level, and detailed data were not elicited or provided for White and Black or African American populations.

In previous decennial censuses, detailed data on race and ethnicity from the census were published only for a limited set of detailed Hispanic groups, detailed Asian groups, detailed Native Hawaiian and Other Pacific Islander groups, and American Indian and Alaska Native tribes and villages. For the 2020 Census, the Census Bureau improved the race and ethnicity questions, coding, and processing to enable the collection and tabulation of all racial and ethnic groups in the United States, including new statistics for detailed European groups, detailed Middle Eastern or North African groups, detailed Sub-Saharan African groups, detailed Afro-Caribbean groups, and other detailed population groups for the first time. These improvements allow for the tabulation of detailed data for all race, ethnicity, and tribal populations in the 2020 Census Detailed DHC-A.

About the Detailed DHC-A Proof of Concept

The data in the Proof of Concept are based on 2010 Census data and represent the tables planned for publication in the 2020 Census Detailed DHC-A. Because 2010 Census data are used in this product, not all 2020 Census Detailed DHC-A race and ethnicity groups are represented. The 2010 Census did not collect data for detailed White and Black or African American responses. Additionally, numerous improvements were made to the coding and processing of race and ethnicity data for the 2020 Census that are not reflected in the Proof of Concept because it utilizes 2010 Census data.

Readers will see three types of data from 2010 in the Proof of Concept. The first type is the published 2010 counts. These data have swapping – the type of disclosure avoidance used in 2010 – already applied to them. The second type of data is the 2010 noise infused counts that have differential privacy – the new type of disclosure avoidance used in 2020 – applied to them. Both types of data started with the values obtained during the 2010 Census collection -- the enumerated count. Differences between the two types of counts therefore reflect the two different privacy methods used. Finally, the third type of data is summary metrics that compare the enumerated 2010 Census data to the 2010 noise infused counts, providing users with an overall assessment of the accuracy of the data in the Proof of Concept.

Technical Background

Disclosure Avoidance

Disclosure avoidance is the processing of data to protect respondent confidentiality. A disclosure of data occurs when someone can use data elements such as a combination of sex, race, date of birth, geographic indicators, or other descriptors to identify an individual respondent. Using disclosure avoidance, the Census Bureau modifies or removes respondent characteristics that put confidential information at risk for disclosure.

Differential Privacy

The Detailed DHC-A uses the differential privacy framework with implementing mechanisms that infuse the data produced with "statistical noise" in a way that protects each respondent's identity and responses. Differential privacy mechanisms work by adding small, random additions and subtractions to the statistics produced from the data using mathematical formulas. These are called "noise infused" or "noisy" counts. Differential privacy mechanisms allow the Census Bureau to ensure that enough noise is added to protect confidentiality but not so much as to damage the statistical validity of published data.

Differences between the TopDown and SafeTab-P Differentially Private Algorithms

The Redistricting Data (P.L. 94-171) Summary File, Demographic and Housing Characteristics File (DHC), and Demographic Profile use the TopDown algorithm developed by the Census Bureau. In contrast, the Detailed DHC-A uses an algorithm called SafeTab-P, which was developed for the Census Bureau by Tumult Labs. SafeTab-P is an algorithm that produces noise infused total population counts as well as noise infused sex by age statistics for detailed racial and ethnic groups and American Indian and Alaska Native tribes and villages. The requirements for the Detailed DHC-A differed substantially from those for the Redistricting Data (P.L. 94-171) Summary File and the DHC. In particular, the number of categories of detailed race, ethnicity, and American Indian and Alaska Native tribes and villages far exceeded the 128 categories used in the Redistricting Data (P.L. 94-171) Summary File and DHC. The Census Bureau determined that the quality of the published data would be greatly improved if the same differential privacy framework (zero-Concentrated Differential Privacy) and implementing mechanism (discrete Gaussian) used in the TopDown algorithm were engineered into a new algorithm that was custom designed for the Detailed DHC-A. Table 1 below describes the differences between these two algorithms in more detail.

Table 1. Differences between the TopDown and SafeTab-P Differentially Private Algorithms

TopDown (Redistricting Data (P.L. 94-171) Summary File, DHC, and Demographic Profile)	**SafeTab-P (Detailed DHC-A)**
Algorithm produces privacy-protected microdata and tabulations are built from those microdata.	Algorithm directly produces privacy-protected tabulations for each geography.
Counts are produced for the nation-level and counts for lower-level geographies are controlled to the national counts, so all geographies aggregate as expected (e.g., state counts sum to national count).	Counts are produced for each geography independently, and there is no requirement that geographies aggregate as expected (e.g., states will not add up to the national count).
When aggregating data, the variability of the statistical noise is controlled so that the statistics become more accurate for larger-population geographies. For example, the Census	When aggregating data, they generally become more variable the more you aggregate. *Data users should use the published statistic they are interested in, when available, rather than*

Bureau recommends data users aggregate block level data to meaningful geographies to get more accurate data.	*aggregating data. Users should use caution when aggregating data for custom geographies.*
The Redistricting Data (P.L. 94-171) Summary File is used as input when processing the DHC and Demographic Profile, so these three data products are mutually consistent.	The Detailed DHC-A will not be consistent with other 2020 Census data products.
Overall accuracy and accuracy for characteristics and geographies can be targeted but the exact levels of accuracy cannot be known in advance, except in the case of the Redistricting Data (P.L. 94-171) Summary File, where the algorithm was tuned to meet pre-specified accuracy targets for total population, largest race/ethnicity group as a proportion of total population, and tract-level race and ethnicity counts.	All margins of error are determined in advance and are met 95% of the time.
Does not use adaptive design.	Uses adaptive design to determine the amount of data provided based on each racial and ethnic group's population size for a given geography.

For more information on the TopDown algorithm, please see Confidentiality Protection in the 2020 US Census of Population and Housing and The 2020 Census Disclosure Avoidance System TopDown Algorithm.

For more information on the SafeTab-P algorithm, please see SafeTab: DP Algorithms for 2020 Census Detailed DHC Race & Ethnicity.

Privacy-Loss Budget

The level of noise introduced to protect against disclosure is guided by a "privacy-loss budget" (PLB) — the budget controls the distribution of the statistical noise and the associated probability of confidentiality breach for data with the schema of the Detailed DHC-A. See Bayesian and Frequentist Semantics for Common Variations of Differential Privacy: Applications to the 2020 Census for a detailed explanation of how the differential privacy framework used for all 2020 Census data products, including the Detailed-DHC-A, protects confidentiality. The PLB, reflected in the parameter *rho*, can be set higher or lower, acting like a dial that tunes the amount of noise that is added to the data. This PLB can be set anywhere on a spectrum from "no accuracy but high protection" to "high accuracy but no protection." Choosing the PLB is a policy decision based on a desired balance between accuracy and confidentiality protection, and the decision must be simultaneously informed by the Census Bureau's legal obligations and feedback on data utility from stakeholders. The lower the budget, the higher the protection and the less precise each data point will be. As the PLB rises, noise decreases (a greater share of the random noise numbers drawn are at or close to zero), meaning the data will be more accurate, but the risk of disclosure rises.

The Detailed DHC-A Proof of Concept uses a PLB of *rho*=19.776. The full confidentiality protection is reflected in *rho*, not the conversion to a single point on the epsilon, delta continuum that it summarizes. It is important to note that the PLB for this Proof of Concept may not be the final PLB that will be used

for the 2020 Census production run of the Detailed DHC-A. The Census Bureau's Data Stewardship Executive Policy (DSEP) committee will make final PLB decisions for the 2020 Census Detailed DHC-A this spring. To see more information on the *rho* used in the Proof of Concept, please see the Privacy Loss Budget Allocation.

Data Used in the Proof of Concept

The Detailed DHC-A Proof of Concept tables use data from the 2010 Census. These data contain detailed race, Hispanic origin, sex, and age data by various geography levels. They do not contain any data on households. For more information on the Hispanic origin and race categories included, the number of detailed groups within each Hispanic origin and race category, and the average size of each category, see the Summary Metrics.

Race and Hispanic Origin

The Detailed DHC-A Proof of Concept includes examples of detailed racial and ethnic groups using data from the 2010 Census. Detailed groups include disaggregated groups such as Puerto Rican, Chinese, the Navajo Nation, Samoan, etc.

Following the 1997 Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity, only a single Hispanic origin response was tabulated in response to the Hispanic origin question.

The race question tabulates multiple race responses when reported and, as a result, detailed race groups come in two forms: "alone" and "alone or in any combination."

- The concept of "detailed race alone" includes people who reported a single entry (e.g., Korean) and no other detailed race group.
- The concept of "detailed race alone or in any combination" includes people who reported a single entry (e.g., Korean) and people who reported that entry with one or more other race(s) (e.g., Korean *and* Thai, or Korean *and* Native Hawaiian). The "detailed race alone or in any combination" concept therefore represents the maximum number of people who reported as that detailed race group, either alone or in any combination with one or more additional detailed race group(s).

The Detailed DHC-A also tabulates regional groups for racial and ethnic groups. Regional groups include groups such as Caribbean, East Asian, American Indian, Polynesian, etc.

Due to tabulating only one Hispanic origin response from the Hispanic origin question, regional ethnic groups only come in one form. Regional race groups come in two forms: "alone" and "alone or in any combination."

- The concept of "regional group alone" includes people who reported one or more detailed race group(s) that aggregate into the same regional group header. For example, respondents who reported Beninese, as well as those who reported Nigerian *and* Ghanaian, are part of the larger "Sub-Saharan African alone" regional group.
- The concept of "regional group alone or in any combination" includes people who reported one or more detailed race group(s) that aggregate into the same regional group header, as well as

people who reported detailed race groups that aggregate into different regional group headers. For example, respondents who reported Navajo Nation as well as those who reported Hopi *and* Brazilian are part of the larger "American Indian alone or in any combination" regional group.

Regional groups were not published in 2010, meaning these tabulations are new to the 2020 Detailed DHC-A. In the Proof of Concept, regional group totals will be demonstrated using 2010 noise infused counts only.

Sex and Age

Sex by age data will not be available for every group. When detailed sex by age data are available, it will show one of three age category types. The sex by age tables available are a four category table, a nine category table, and a 23 category table, shown below in Table 2. See Adaptive Design for information on how the availability of sex by age data is determined.

Table 2. Categories for Sex by Age Tables

4 age categories	**9 age categories**	**23 age categories**
Male	**Male**	**Male**
Under 18 years	Under 5 years	Under 5 years
18 to 44 years	5 to 17 years	5 to 9 years
45 to 64 years	18 to 24 years	10 to 14 years
65 years and over	25 to 34 years	15 to 17 years
Female	35 to 44 years	18 to 19 years
Under 18	45 to 54 years	20 years
...	55 to 64 years	21 years
	65 to 74 years	22 to 24 years
	Female	30 to 34 years
	Under 5 years	35 to 39 years
	...	40 to 44 years
		45 to 49 years
		50 to 54 years
		55 to 59 years
		60 to 61 years
		62 to 64 years
		65 to 66 years
		67 to 69 years
		70 to 74 years
		75 to 79 years
		80 to 84 years
		85 years and over
		Female
		Under 5 years
		…

Geography

Examples of data tables and data limitations will be provided for the following six geographic summary levels:

- 010 – United States
- 040 – State
- 050 – State-County
- 140 – State-County-Census Tract
- 160 – State-Place
- 250 – American Indian Area/Alaska Native Area/Hawaiian Home Land (AIANNH)[1]

The term "sub-state" is used throughout this document and refers to county, tract, and place levels.

Minimum Population Counts

Minimum population counts are used to protect respondent confidentiality in both the 2010 published data and the 2010 noise infused data. However, due to the different types of disclosure avoidance applied to the two types of data, the minimum population counts vary between the 2010 published counts and the 2010 noise infused counts.

Table 3 shows the minimum population counts used in the 2010 published data compared to the minimum population counts used in the 2010 noise infused data for detailed groups in the Proof of Concept.

Because regional groups did not exist in 2010, they are not shown in Table 3. Regional groups have a minimum population count of 94 at sub-state geographies to receive total population counts and a minimum population count of 5,000 to receive sex by age data at any geography level.

[1] This includes legal entities (federally recognized American Indian reservations and off-reservation trust land areas, the tribal subdivisions that can divide these entities, state-recognized American Indian reservations, Alaska Native Regional Corporations, and Hawaiian home lands) and statistical entities (Alaska Native village statistical areas, Oklahoma tribal statistical areas, tribal designated statistical areas, and state designated tribal statistical areas).

Table 3. Minimum Population Counts for Detailed Groups in 2010 Published Data Compared to 2010 Noise Infused Data in the Proof of Concept

	2010 Published counts	2010 Noise infused counts
Total population		
Nation and state	Asian, Native Hawaiian and Other Pacific Islander, and Hispanic origin groups: • Minimum population count of 100 American Indian and Alaska Native tribes and villages: • No minimum population count	All race and ethnicity groups and American Indian and Alaska Native tribes and villages: • No minimum population counts but data may be suppressed (See Data Suppression for more information)
Sub-state geographies and AIANNH areas	Minimum population count of 100	Minimum population count of 22
Sex by age		
Nation and state	Minimum population count of 100	Minimum population count of 500
Sub-state geographies and AIANNH areas	Minimum population count of 100	Minimum population count of 1,000

Detailed DHC-A Product Design

Adaptive Design

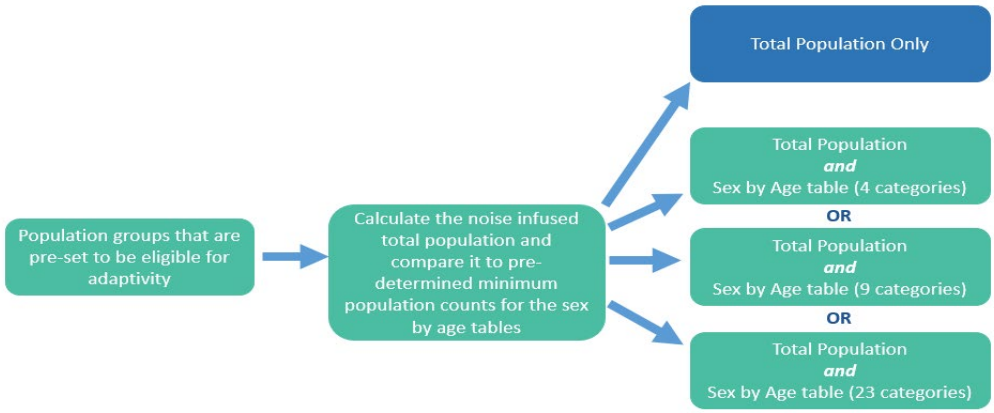
The Detailed DHC-A uses an adaptive design to determine the amount of data detailed and regional racial and ethnic groups and American Indian and Alaska Native tribes and villages receive based on population size and geography level. At the nation and state levels, detailed and regional groups will receive a total population count table and groups meeting a minimum population count will also receive a sex by age table. For sub-state geographies and AIANNH areas, the adaptive design will use minimum population counts to determine eligibility for a total population count table and a sex by age table.

Based on data from the 2010 Census, we anticipate that some groups will be too small to receive a sex by age table in 2020. Detailed groups with a national population less than 50 in the 2010 Census are pre-set to only receive nation and state level total population counts. This is determined in advance, which allows us to adjust the noise infusion and produce more accurate counts for these small populations. For these groups, a noise infused total population count will always be calculated, but no sex by age statistics will be available.

All other population groups are eligible to go through the adaptive design, meaning groups with a national population of at least 50 in the 2010 Census or that we did not collect data for in the 2010 Census.

The noise infused population count is calculated and compared to minimum population counts used to determine which table(s) the group is eligible to receive. This process, shown in Figure 1, is repeated for each group for each geography.

Figure 1. Adaptive Design



The following groups will receive total population counts only:

- Detailed groups with noise infused counts of less than 500 at the nation and state levels
- Detailed groups with noise infused counts of 22-999 in sub-state geographies and AIANNH areas
- Regional groups with noise infused counts of less than 5,000 at the nation and state levels
- Regional groups with noise infused counts of 94-4,999 in sub-state geographies[2]

Groups not limited to total population counts only are eligible for sex by age data. The number of age categories a group is eligible for is dependent on their population at a given geography, as shown in Table 5.[3]

The adaptive design used for the Detailed DHC-A means that the tables produced for a detailed or regional group will vary based on the specific geography a data user selects. The type of table a group qualifies for at a given geography is independent of the table(s) a group is qualified for at other geographic levels. Similarly, there will be differences within a given geographic level: a group that qualifies for sex by age tables in some states may only receive a total count in others due to the size of that specific group in any given state.

Data Accuracy and Margins of Error

One of the key measures we use to ensure quality data is the margin of error (MOE) with a 95% confidence interval. The MOE is how close we can expect the noise infused counts to be to the enumerated counts approximately 95% of the time. In other words, if we were to run the disclosure avoidance system or noise infusion 100 times, we would expect the enumerated counts to be within the MOE of the noise infused counts in about 95 of those 100 runs. For a noise infused count of 20 with an MOE of 3, we are 95% confident that the enumerated count is somewhere between 17 and 23. The smaller the MOE the more accurate the data are; however, given the same MOE, the relative impact is

[2] Regional groups are not available for AIANNH areas.
[3] Groups will be assigned the sex by age table that corresponds to their total population count 99.9% of the time.

less for large groups, compared to small groups. Table 4, shown below, illustrates this by showing small and large population counts with the same MOE.

Table 4. Examples of Margin of Error of 50 for Groups of Varying Sizes

	Noise infused count of 50	**Noise infused count of 75**	**Noise infused count of 1,000**	**Noise infused count of 50,000**
Enumerated count between:	0 to 100	25 to 125	950 to 1,050	49,950 to 50,050

Based on our review, we find that the Detailed DHC-A SafeTab-P algorithm is functioning as specified and produces accurate data. The metrics developed for this Proof of Concept show the following at the national level:

- Over 95% of the noise infused counts for each racial and ethnic category are within their expected MOE.
- The average difference between the enumerated count and the noise infused count for racial and ethnic groups and American Indian and Alaska Native tribes and villages at the national level ranges from +/- 0.5 to +/-1.35.
- Less than 1% of noise infused counts are outliers for each racial and ethnic category, meaning they are outside their MOE by more than twice the value of the MOE. For example, given an MOE of 3, fewer than 1% of noise infused counts are +/-6 from their enumerated counts.

For more information on how many noise infused counts fall within their expected MOE, how many outliers there are, and more about the quality of these data, please see the Summary Metrics.

The MOE for detailed groups in the Detailed DHC-A varies by geography. When only the total count table is produced for a detailed group at the nation and state level, the MOE is 3. When only a total count table is produced at sub-state geographies and AIANNH areas, the MOE is 11. As regional groups are larger than detailed groups, their MOE is 50 when only the total count is produced at all levels of geography.

Table 5. Detailed DHC-A Minimum Noise Infused Population Counts and Margins of Error (MOE) by Geography

	Detailed groups		**Regional groups**	
Most comprehensive table type produced	**Nation & state (MOE=3)**	**Sub-state & AIANNH (MOE=11)**	**Nation & state (MOE=50)**	**Sub-state (MOE=50)**
Total count only	0-499	22-999	0-4,999	94-4,999
Sex by age – 4 categories	500-999	1,000-4,999	5,000-19,999	5,000-19,999
Sex by age – 9 categories	1,000-6,999	5,000-19,999	20,000-149,999	20,000-149,999
Sex by age – 23 categories	7,000+	20,000+	150,000+	150,000+

When a sex by age table is produced for a group, the MOE shown in Table 5 applies to the individual sex by age counts. This means we expect these noise infused sex by age counts to vary from their enumerated count by no more than plus or minus their MOE 95% of the time. As illustrated in Table 6, in

a sex by age table with four age categories, eight counts are noise infused and therefore have an MOE of 3 (each of the four age categories, twice: once for male and once for female).

Table 6. Example of Aggregated MOEs for Sex by Age Tables

Sex by age – 4 categories	**Nation & state (MOE=3)**
Total count	Aggregated MOE=8.5
Male	Aggregated MOE=6
Under 18 years	Noise Infused with MOE=3
18 to 44 years	Noise Infused with MOE=3
45 to 64 years	Noise Infused with MOE=3
65 years and over	Noise Infused with MOE=3
Female	Aggregated MOE=6
Under 18 years	Noise Infused with MOE=3
18 to 44 years	Noise Infused with MOE=3
45 to 64 years	Noise Infused with MOE=3
65 years and over	Noise Infused with MOE=3

However, the total count for groups with sex by age data is aggregated from the noise infused sex by age counts. This will lead to a higher MOE for the total population count because each individual sex by age count is noise infused. As a result, totals created from these counts combine the noise from each individual count, leading to noisier totals. As shown in Table 6, the aggregated male and female totals have larger MOEs than the individual sex by age counts, as does the total count for the whole population group.

Table 7 shows the MOE for the total counts for population groups based on the sex by age data they receive. While the MOEs do increase, it is important to remember that the larger MOEs shown below correspond to larger population groups. We expect these aggregated noise infused counts to vary from their enumerated count by no more than plus or minus their MOE 95% of the time.

To learn more about how to calculate aggregated MOEs, see the tab "Calculations for Aggregated MOE" in the Examples documentation.

Table 7. Aggregated Margins of Error by Geography and Sex by Age Data

Most comprehensive table type produced	**Nation & state detailed groups**	**Sub-state & AIANNH detailed groups**	**Regional groups**
Total count only	3	11	50
Total count aggregated from sex by age – 4 category table	8.5	31.1	141.4
Total count aggregated from sex by age – 9 category table	12.7	46.7	212.1
Total count aggregated from sex by age – 23 category table	20.3	74.6	339.1

Example Tables

For examples of 2010 published counts, 2010 noise infused counts, and margins of error for various detailed and regional race and ethnicity groups, see their respective "Nation & State," "Sub-State," and "AIANNH" tabs in the Examples documentation.

Limitations

Data users will encounter data inconsistencies and unpublished data in the Detailed DHC-A. Below, we describe the reasons for this and give examples of how they appear in the data.

Due to the limitations explored below, we caution data users against aggregating data. This includes creating new regional groups from detailed groups, combining lower-levels of geography to create higher-levels of geography, combining sex by age data, etc. We strongly recommend using the published noise infused counts associated with your geographic level of interest when available.

Improbable data

Improbable data in the Detailed DHC-A stem from two sources. One is the noise necessary to protect respondent confidentiality. The other is the lack of postprocessing (or consistency) requirements compared to the Redistricting Data (P.L. 94-171) Summary File and the DHC. As a result, data users will observe inconsistencies, including some implausible results, in the Detailed DHC-A. Although these inconsistencies can seem conceptually illogical, they help preserve random noise and unbiased data in the product, as well as preserving the MOEs of the noise infused data with 95% confidence.

Below, we show examples of inconsistencies in the data that stem from these two sources.

1. **Inconsistencies from the noise necessary to protect confidentiality**

Due to the way noise is applied to data to protect respondent confidentiality, aggregating counts can result in noisier or less accurate data. *As a result, we caution data users from creating their own aggregations, and we recommend using published noise infused counts when available.*

a. **Total population counts for groups with sex by age statistics will have larger margins of error** – Each sex by age count in a sex by age table has noise applied to it, which makes the total count noisier than if only the total population count had been published. This is because the population totals for groups with sex by age data are the aggregation of those individual noise infused sex by age counts. The resulting total population count will have a larger margin of error due to the aggregation of noise infused counts used to create it. Below, we show the aggregated MOEs. For instructions on how to calculate aggregated MOEs, see the Examples documentation tab "Calculations for Aggregated MOE."

Table 8. Limitation 1a.

Universe: Arctic Slope Corporation alone		
United States	**2010 Noise infused count**	**MOE**
Total count	695	Aggregated MOE = 8.5
Male:	334	Aggregated MOE = 6
Under 18 years	158	MOE= 3
18 to 44 years	131	MOE= 3
45 to 64 years	40	MOE= 3
65 years and over	5	MOE= 3
Female:	361	Aggregated MOE = 6
Under 18 years	161	MOE= 3
18 to 44 years	152	MOE= 3
45 to 64 years	39	MOE= 3
65 years and over	9	MOE= 3

b. **Detailed groups may not sum to their corresponding regional group** – If a data user adds up all detailed groups in a regional group, they may get a different total than the noise infused total count for the regional group. Noise is applied once to the regional group total whereas the aggregation of individual detailed groups compounds the noise applied to each detailed group's individual data, resulting in a noisier total. *We recommend using the published noise infused counts when available.* To see the data for all Caribbean detailed groups, see tab "Limitation 1b" in the Examples documentation.

Table 9. Limitation 1b.

Universe: Caribbean regional group			
United States	**Caribbean regional group total**	**Sum of detailed Caribbean groups**	**Difference between regional group total and sum of detailed groups**
2010 Noise infused count	7,828,451	7,828,335	116

2. Inconsistent data due to minimal post-processing consistency requirements

Noise is applied independently to each race and ethnicity group at every geography level. This means that geographic hierarchies were not taken into account when applying noise to the data. In data products using the TopDown algorithm, post-processing procedures were applied to ensure geographic consistency. This was not a requirement for the Detailed DHC-A, and as a result, data may be geographically inconsistent.

a. **Geographic equivalents may have different counts** – A location with the same boundaries at varying geography levels may have different counts for the same

racial, ethnic, or American Indian or Alaska Native population because the algorithm outputs a noise infused count for each geography level independent of other geographies. *We recommend using the count associated with your geographic level of interest.*

Table 10. Limitation 2a.

Universe: Samoan alone or in any combination			
Philadelphia, PA	**County**	**Place**	**Difference between county and place**
2010 Published count	198	198	0
2010 Noise infused count	205	197	8

b. **Values from lower-level geographies may not sum to values from higher-level geographies** – For example, if a data user adds up all county data for a specific group in a state, they may get a different count than the published noise infused count for that state. *We recommend using the published noise infused counts when available.* To see the data for all states, see tab "Limitation 2b" in the Examples documentation.

Table 11. Limitation 2b.

Universe: Native Hawaiian alone or in any combination			
	United States	**Sum of states**	**Difference between United States total and sum of states**
2010 Published count	527,077	527,077	0
2010 Noise infused count	527,209	527,108	101

c. **Lower-level geographies may have larger counts than higher-level geographies** – For example, the state count for a population group may be larger than the count for the United States.

Table 12. Limitation 2c.

Universe: Village of Red Devil alone			
	United States	**Alaska**	**Difference between United States and Alaska**
2010 Published count	3	3	0
2010 Noise infused count	2	3	-1

Unpublished counts

Data users may encounter unpublished data due to two reasons: implausible or small counts that have been suppressed and counts that are not included in the product. *As a result, we caution data users from creating their own aggregations, and we recommend using published counts when available.*

3. Data Suppression

Data are suppressed, meaning they are not published, in three situations: (1) the noise infused counts are negative, (2) the "race alone" count is suppressed when it is larger than its equivalent "race alone or in any combination" count, and (3) there are small, statistically unreliable, noise infused counts in sub-state geographies and AIANNH areas. Suppressed counts will be denoted with an "X" in tables.

a) **Noise infused counts below zero will not be published** – Due to the way noise is applied to data to protect confidentiality, the noise infused counts can be negative, meaning they have counts of less than zero. Because negative counts are not demographically reasonable, these counts are not published.

Table 13. Limitation 3a.

Pre-suppression noise infused count	Post-suppression count
-2	X

b) **Detailed and regional "race alone" counts are not published when larger than the corresponding "race alone or in any combination" count** – To be demographically reasonable, the "race alone or in any combination" population should always be equal to or greater than the "race alone" population for a group.

Table 14. Limitation 3b.

Universe: Detailed or regional race group		
	Alone	Alone or in any combination
Pre-suppression noise infused count	139	137
Post-suppression count	X	137

c) **Detailed groups with populations of less than 22 and regional groups with populations of less than 94 at any given sub-state geography and AIANNH area will not have data published for that geography** – For example, a county with a noise infused count of less than 22 for a given detailed group will not have data published for that group but will have data available for all groups of at least 22. These minimum population counts ensure that enumerated zeros are not published 99.99% of the time. This is an improvement from 2010, where Asian, Native Hawaiian and Other Pacific Islander, and Hispanic origin groups needed a minimum population count of 100 to have published counts at any given geography level.

Table 15. Limitation 3c.

Universe: Pakistani alone				
Delaware	**State total**	**Kent County**	**New Castle County**	**Sussex County**
2010 Published count	1,090	106	895	X
2010 Noise infused count	1,095	121	886	97
Difference	-5	-15	9	X

4. Counts that are not included in the product

The lack of certain race data in the product will only be noticeable if data users create their own aggregated groups. *As a result, we caution data users from creating their own aggregations, and we recommend using published noise infused counts when available.*

a) **Totals for regional "race alone" groups and aggregated totals of detailed "race alone" groups will not equal each other** – The provided regional group totals include people who report they are two or more detailed race groups within the same regional grouping (e.g., Korean *and* Mongolian within East Asian). There is no equivalent category in the detailed "race alone" data, as was also the case in the 2010 data products. In the example below, the difference between the regional group total and the sum of detailed groups is a combination of noise aggregation from the individual detailed groups and the respondents who are two or more East Asian groups alone. *We recommend using the published noise infused counts when available.* To see the data for all East Asian detailed groups, see tab "Limitation 3" in the Examples documentation.

Table 16. Limitation 4a.

Universe: East Asian alone regional group			
United States	**East Asian alone regional group total**	**Sum of detailed East Asian groups**	**Difference between regional group total and sum of detailed groups**
2010 Noise infused count	5,858,438	5,786,768	71,670

Future Steps

How to Provide Feedback

Your feedback will be instrumental in making final decisions regarding the 2020 Census Detailed DHC-A. Data users are invited to send feedback on the Detailed DHC-A Proof of Concept to 2020DAS@census.gov with the subject "2020 Census Detailed DHC-A."

To meet our 2020 Census Detailed DHC-A production release date of August 2023, all feedback must be received by **March 2, 2023.** If you do not want us to publish your feedback or if you want us to remove your identifying information, please indicate that in the email.

We would like to hear from you about how you or your work will be positively, negatively, or not impacted if we release the final product using the current specifications outlined in the Proof of Concept. Please provide the level of geography, description of the use case(s), and implications should the data be released as reflected in this Proof of Concept.

Detailed DHC-B

The Detailed DHC-A and Detailed DHC-B are successors to the 2010 Census Summary File 2 and American Indian and Alaska Native Summary File. Based on public feedback, the Census Bureau prioritized the release of Detailed DHC-A. As a result, the Detailed DHC-A is further along in development with a release date of August 2023. At the same time, the Census Bureau is developing the Detailed DHC-B to be as similar to the Detailed DHC-A as possible.

The Detailed DHC-B differs from the Detailed DHC-A because it provides household and housing unit counts, not population counts. The Detailed DHC-B is comprised of two tables – Household Type and Tenure – by race of householder or Hispanic origin of householder. Household Type includes information about family and nonfamily households, and Tenure includes information about owner- or renter-occupied housing units. Both tables are provided for approximately 370 detailed racial and ethnic groups and 1,200 American Indian and Alaska Native tribes and villages. The table shells are shown in Table 17 and Table 18.

Table 17. Household Type

Universe: Households
Total:
Family households:
Married couple family
Other family:
Male householder, no spouse present
Female householder, no spouse present
Nonfamily households:
Householder living alone
Householder not living alone

Table 18. Tenure

Universe: Occupied housing units
Total:
Owned with a mortgage or a loan
Owned free and clear
Renter occupied

We are trying to keep the Detailed DHC-B as similar to the Detailed DHC-A as possible. We are proposing the same geographies: nation, states, counties, tracts, places, and AIANNH areas. The same detailed and regional groups will be eligible to receive data. We are proposing the use of an adaptive design that works similarly to the Detailed DHC-A where detailed groups will receive a full or partial table depending on the size of their population. For example, depending on the detailed group's size, they could receive

the full tenure table shown above or a partial household table shown in Table 19. Despite these tables being tabulated at the household/householder level, adaptability will also be based on the count of people in the racial or ethnic group, not householders. We are proposing the same margins of error, and the Detailed DHC-B may include the same limitations, as the Detailed DHC-A.

Table 19. Tenure

Universe: Occupied housing units
Total:

As we continue to develop the Detailed DHC-B, we will keep data users informed of updates to the proposal and solicit feedback from them.