Disclosure Avoidance Methods for the Detailed Demographic and Housing Characteristics File A (Detailed DHC-A): How SafeTab-P Works

2020 Census Briefs

By the Population Reference Bureau and the U.S. Census Bureau's 2020 Census Data Products and Dissemination Team

C2020BR-05

October 2023

INTRODUCTION

This is the fourth in a series of briefs describing disclosure avoidance methods used to protect 2020 Census data products and the implications of those methods for data users. This brief describes how differential privacy works and how it is applied to the 2020 Census Detailed Demographic and Housing Characteristics File A (Detailed DHC-A). The methodology used to protect the data in the Detailed DHC-A is different than the methodology used in other census data products. This brief also explains those differences and provides guidance for data users.

At the U.S. Census Bureau, disclosure avoidance is defined as a process used to protect the confidentiality of respondents' personal information. The Census Bureau has applied disclosure avoidance methods for decades to keep respondents' information confidential and maintain public trust in the data.

Over time, the Census Bureau has published more detailed data while advances in data science, more powerful computers, and externally accessible data have increased the risk of identifying individuals from published statistics. With ever-advancing technology, the threats to disclosure are expected to continue growing with time. To reduce this risk, the Census Bureau is implementing new disclosure avoidance methods for the 2020 Census based on a framework known as differential privacy.

What Is Differential Privacy?

Differential privacy is a scientific framework for processing data to protect the identities and personal information of the people in the data. It works by adding statistical noise—small, random additions or subtractions—to every published statistic so that no one can reidentify a specific person or household with any certainty using any combination of the published data.

Differential privacy forms the foundation of the Disclosure Avoidance System used to adjust the data to protect 2020 Census respondent confidentiality.

WHAT IS THE DETAILED DHC-A?

The Detailed DHC-A is a successor (in combination with the forthcoming Detailed DHC-B) to the 2010 Census Summary File 2 and the 2010 Census American Indian and Alaska Native Summary File. In developing the Detailed DHC-A, the primary goal was to produce accurate data for the nation's detailed race and ethnicity groups and American Indian and Alaska Native tribes and villages.

The Detailed DHC-A provides population counts for more detailed groups than ever before because of improved race and ethnicity questions, coding, and processing. Detailed groups include:

- 30 detailed Hispanic or Latino origin groups such as Mexican, Salvadoran, and more.
- 270 detailed race groups such as Japanese, Native Hawaiian, Irish, Lebanese, Haitian, Brazilian, and more.
- 1,187 American Indian and Alaska Native tribes and villages such as the Navajo Nation, the Akiak Native Community, and more.

United States® Census Bureau

U.S. Department of Commerce
U.S. CENSUS BUREAU
census.gov

Detailed groups reported in the race question have "alone" and "alone or in any combination" counts. The alone count represents the minimum number of people who identified as that detailed group. The alone count includes respondents with only one response such as Hungarian. The alone or in any combination count represents the maximum number of people who identified as that detailed group. It includes respondents with only one response, such as Hungarian, and those with multiple responses such as Hungarian and Romanian or Hungarian and Black or African American.

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

This is the first time data for regional groups will be released, including:

- 4 regional Hispanic or Latino origin groups (Central American, South American, and more).
- 24 regional race groups, including Middle Eastern and North African, Sub-Saharan African, and more.

Regional groups aggregated from detailed race responses also have alone and alone or in any combination counts. The alone count includes respondents with one or more detailed race group(s) that aggregate into the same regional group. For example, respondents who reported only Hungarian, as well as those who reported both Lithuanian and Romanian, are part of the larger "European alone" regional group. The alone or in any combination count includes people who reported one or more detailed race group(s) that aggregate into the same regional group, as well as people who reported detailed race groups that aggregate into different regional groups. 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.

Given their relatively small numbers, publishing statistics for many of these detailed groups while maintaining confidentiality is a challenge. To overcome this challenge and address feedback from data users, the Detailed DHC-A disclosure avoidance framework focused on providing as many accurate population counts as confidentiality protections would allow at various geography levels. The tradeoff was some population groups have less granular sex by age detail or may have total population counts only.

How Adaptive Design Works

Adaptive design is a data-driven framework for choosing which statistics to publish. The Detailed DHC-A uses an adaptive design that adjusts the amount of age data published for a population group based on a combination of predetermined thresholds and level of geography. Adaptive design allows the Census Bureau to provide more detailed statistics for racial and ethnic groups with larger populations, while ensuring confidentiality protections.

As shown in Table 1, all groups receiving data will have a total population count table. In addition, some will receive one of three levels of age data. Table types include the following:

- Total population only.
- A table with 4 age categories.
- A table with 9 age categories.
- A table with 23 age categories.

Table 1.

Detailed Demographic and Housing Characteristics File A Minimum Population Thresholds by Geography

Most comprehensive table type produced	Detailed groups		Regional groups	
	Nation and state	Substate and AIANNH	Nation and state	Substate
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+

Note: AIANNH is American Indian/Alaska Native/Native Hawaiian areas. Substate includes county, place, and census tract.
Source: U.S. Census Bureau.

Working With the Data

As explained in this brief, the Detailed DHC-A data product is different than some other 2020 Census data products. Data users should keep the following practices in mind when working with the data:

- Do use the published counts for detailed and regional groups and geographies whenever possible, rather than creating custom aggregations, which will result in less accurate data.
- Do use caution if aggregation of published counts is necessary to produce statistics for custom groups or geographies. Unlike in the Redistricting Data (P.L. 94-171) Summary File and Demographic and Housing Characteristics File (DHC) data products, aggregations of data from the Detailed DHC-A get noisier in proportion to the number of counts that are included in the aggregation.
- Do aggregate the smallest number of items possible. For example, if an area could be calculated as the sum of 10 tracts or a county minus 2 tracts, calculate the count based on county minus 2 tracts rather than sum of 10 tracts.
- Do refer to the "Guidance for Working With the Detailed DHC-A" section of this brief for information about calculating percentages, understanding suppressed counts, comparing to other sources, and aggregating.
- Don't expect consistency across tables within the Detailed DHC-A. For example, counties may not sum to state totals, and detailed groups may not sum to regional groups.
- Don't expect consistency with other 2020 Census data products. The Detailed DHC-A was not produced to match population counts published in other products, including the Redistricting Data (P.L. 94-171) Summary File and the DHC data.
- Don't infer suppressed totals by adding or subtracting published data. These data were suppressed for demographic reasonableness or being statistically unreliable. Inferred totals may be unreliable (such as negative numbers).

Detailed groups with a national population smaller than 50 in the 2010 Census are preset to receive total population counts at the nation and state levels only. For these groups, total population counts are available at the nation and state levels, but no other data are produced. All other population groups (those that had a national population of 50 or higher in the 2010 Census and groups for which information was not collected in the 2010 Census) go through the adaptive design.

First, the adaptive design algorithm calculates a noise-infused total population count for each population group in the 2020 Census. Second, that noise-infused population count determines if the population group meets the threshold to produce a total population count only or one of the three possible sex by age tables. This process is repeated for each combination of geography and race or ethnicity group.

A similar process is used to publish Detailed DHC-A data for regional groups, but the population thresholds for these groups are higher (Table 1).

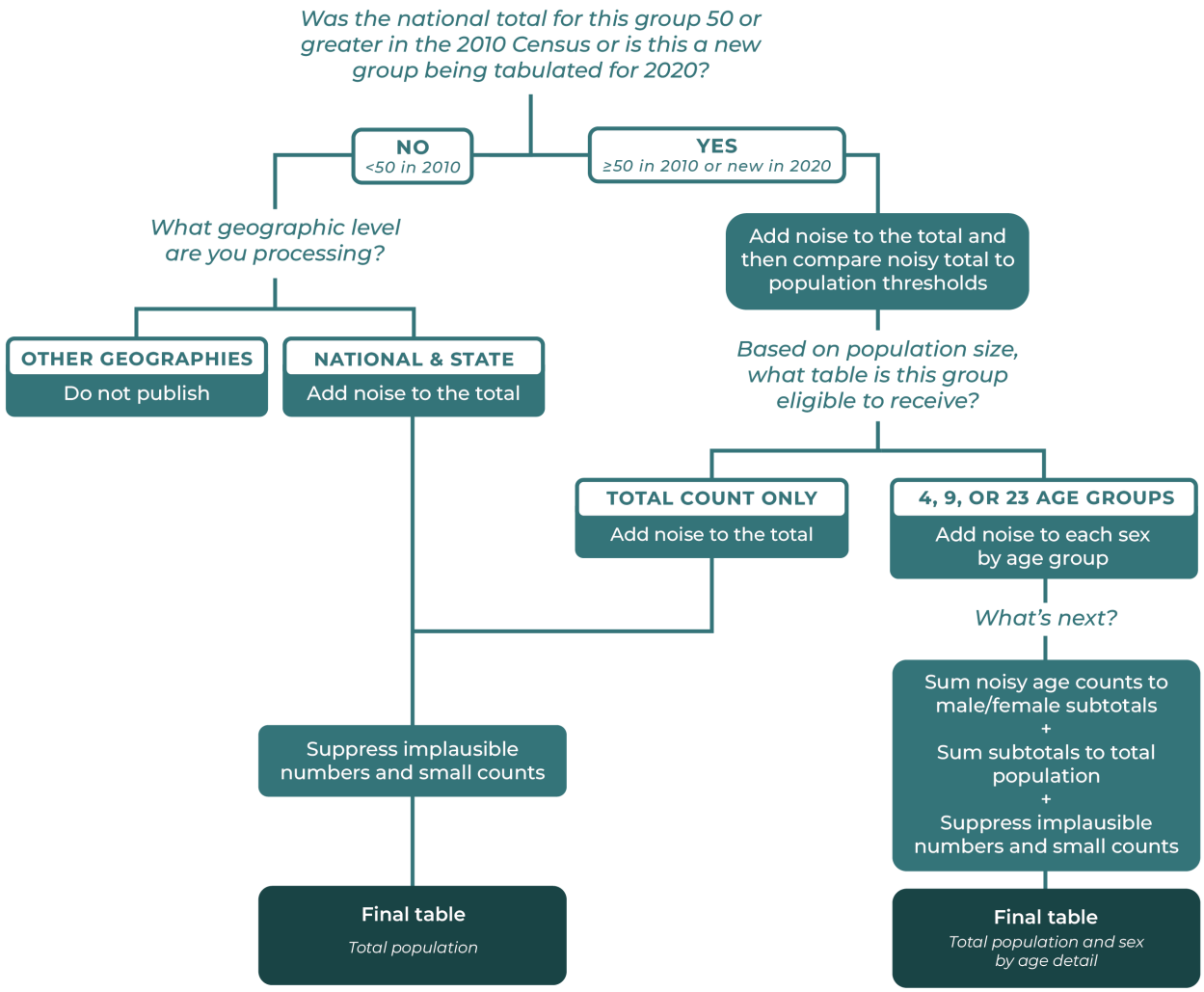
HOW THE SAFETAB-P ALGORITHM WORKS

Underlying this adaptive design is an algorithm called SafeTab-P, a series of mathematical procedures that first determines whether and how much sex by age detail is published for each racial and ethnic group (refer to the "How Adaptive Design Works" section in this brief). Then, the algorithm adds noise to the results. Finally, a postprocessing routine addresses certain demographic reasonableness requirements.

First, SafeTab-P checks to see if a given group is among those predetermined to only receive total population counts (Figure 1). If it is, noise is added to the total, and no sex by age detail is processed. If it is not, SafeTab-P moves on to the next step of determining how much sex by age detail, if any, is published.

Because any published statistic—including total population size—represents a piece of the total disclosure risk, the decision about how much detail to publish is calculated based on a noise-infused total rather than the enumerated total. SafeTab-P adds noise to the

Figure 1.
Adaptive Design for Each Detailed or Regional Group



Source: Population Reference Bureau.

enumerated total and uses a new noisy value that is never published to determine the level of sex by age detail to produce. Because the published count is not used to determine the amount of sex by age data, groups close to threshold boundaries may receive a different amount of sex by age data than indicated by the thresholds. Groups will only receive less data than the thresholds indicate in approximately 0.01 percent of cases.

Let's use the total U.S. Singaporean population from the Detailed DHC-A Proof of Concept, which uses 2010 Census data, as our example of how this works for a group that is eligible to receive sex by age data.

Let's say the enumerated Singaporean alone or in any combination population is 5,350 in 2010; SafeTab-P would first add noise to produce a noisy national total for that detailed group that is never published. In our example this unpublished noisy total is 5,355 (Table 2, Step 1). At the national level, the Singaporean alone or in any combination table would receive nine age categories because the unpublished noisy population size is between 1,000 and 6,999 (Table 1). Then, SafeTab-P tabulates the values for the sex by age categories and adds noise to each of the sex by age counts (Table 2, Step 2).

Table 2.

Example of Noise Added to the U.S. Table for a Detailed Race Group

(Singaporean alone or in any combination)

Sex and age	Step 1: Determine level of age detail to be published			Step 2: Add noise		
	Hypothetical enumerated population	Noise added to check total (not published)	Noisy total for sex by age detail decision	Hypothetical enumerated population	Noise added	Published data
Total count (step 1)	**5,350**	**5**	**5,355**	**X**	**X**	**X**
Total count (step 2)	**X**	**X**	**X**	**5,350**	**22**	**5,372**
Male	**X**	**X**	**X**	**2,178**	**9**	**2,187**
Under 5	X	X	X	146	0	146
5 to 17	X	X	X	318	1	319
18 to 24	X	X	X	316	6	322
25 to 34	X	X	X	472	2	474
35 to 44	X	X	X	410	–3	407
45 to 54	X	X	X	326	0	326
55 to 64	X	X	X	141	1	142
65 to 74	X	X	X	31	1	32
75 and over	X	X	X	18	1	19
Female	**X**	**X**	**X**	**3,172**	**13**	**3,185**
Under 5	X	X	X	147	–3	144
5 to 17	X	X	X	315	2	317
18 to 24	X	X	X	413	5	418
25 to 34	X	X	X	567	4	571
35 to 44	X	X	X	730	2	732
45 to 54	X	X	X	647	–3	644
55 to 64	X	X	X	264	2	266
65 to 74	X	X	X	67	3	70
75 and over	X	X	X	22	1	23

X Not applicable.

Note: Based on the hypothetical noisy count of 5,355 falling between 1,000–6,999, the U.S. table for Singaporean alone or in any combination would have nine age categories.

Source: U.S. Census Bureau, Detailed Demographic and Housing Characteristics File A (Detailed DHC-A) Proof of Concept.

For groups with published sex by age data, the noisy data for each age category is summed to get the published total for the male population and the published total for the female population. The data are also summed across age and sex categories to get the published total population.

Because SafeTab-P repeats this noise infusion process independently across geographies and population groups, data may be inconsistent across tables. For this reason, the Census Bureau encourages data users to access the published counts for detailed and regional groups whenever possible, rather than creating custom aggregations of Detailed DHC-A counts, which will result in less accurate data.

How Are Data Protection and Accuracy Balanced?

The risk of disclosure is referred to as "privacy loss." Tracking the balance between accuracy and protection occurs through a privacy-loss budget. If the system is set to prioritize accuracy for a certain set of results (such as state-level detailed groups), those groups require a larger share of the overall privacy-loss budget. "Balancing the budget" then requires adding more noise (a smaller share of the privacy-loss budget) to certain other results that may not require as much accuracy.

The total amount of noise can be set on a spectrum from "high accuracy but no protection" to "no accuracy but high protection." "No accuracy but high protection" means that large values are added or subtracted—completely distorting the data. "High accuracy but low protection" means that most of the noise added is at or close to zero.

The privacy-loss budget is set by the Census Bureau's Data Stewardship Executive Policy Committee to ensure fitness for use of the 2020 Census data, while effectively protecting confidentiality. Details of the privacy-loss budget for Detailed DHC-A are described in the 2020 Census Demographic and Housing Characteristics File A (Detailed DHC-A) Technical Documentation.

How Much Noise Is There in the Detailed DHC-A?

The Census Bureau set accuracy targets for each count that is produced in the Detailed DHC-A. Those targets ensure the noise-infused counts are within (plus or minus) a target margin of error (MOE) of their enumerated count about 95 percent of the time.[1]

The target MOEs for detailed groups in the Detailed DHC-A vary by geography (Table 3). For detailed groups that get total population counts only, the MOE for the total population is ±3 at the nation and state levels. The MOE for an individual sex by age count for detailed groups at the national and state level is ±3. For population counts produced at substate geographies and American Indian, Alaska Native, and Native Hawaiian (AIANNH) areas, the MOE for a detailed age group is ±11. The MOE for regional groups, which are generally larger than detailed groups, is ±50 at all levels of geography. For example, in any state table for any regional group with 23 age categories, the female population aged 40 to 44 would be within ±50 of the enumerated count 95 percent of the time.

In addition, it is important to remember that subtotals (e.g., male and female populations) and total populations represent the sum of the age group data for groups where sex by age data are published. As a result, noise is aggregated from age details to sex totals and then to the total population. In other words, the sums are noisier than the parts.

For example, in a state table with four age categories, the MOE for each age is ±3, but the MOE for female population is ±6, and the MOE for total population is ±8.5. Table 3 provides the expected MOE for each table type in Detailed DHC-A using the four-category sex by age table as an expanded example.

For more guidance on how to aggregate data, refer to the "Aggregating Data" section in this brief.

[1] The Disclosure Avoidance System is not the only source of uncertainty in 2020 Census data. Noise introduced by disclosure avoidance may compound underlying errors or may offset those errors. Examples of these types of errors are available in the 2020 Census Post-Enumeration Survey, <www.census.gov/programs-surveys/decennial-census/about/coverage-measurement/pes.html>.

Table 3.
Target Margins of Error (±) by Level of Geography and Demographic Detail

Sex by age	Detailed groups		Regional groups
	Nation and state	Substate and AIANNH	All geographies
Total Population Only			
Preset total only	3	X	*X
Total only	3	11	50
Sex by Age—4 Age Categories Total	**8.5**	**31.1**	**141.4**
Female	6	22	100
Under 18	3	11	50
18 to 44	3	11	50
45 to 64	3	11	50
65 and over	3	11	50
Male	6	22	100
Under 18	3	11	50
18 to 44	3	11	50
45 to 64	3	11	50
65 and over	3	11	50
Sex by Age—9 Age Categories Total	**12.7**	**46.7**	**212.1**
Female	9	33	150
Each age group	3	11	50
Male	9	33	150
Each age group	3	11	50
Sex by Age—23 Age Categories Total	**20.3**	**74.6**	**339.1**
Female	14.4	52.8	239.8
Each age group	3	11	50
Male	14.4	52.8	239.8
Each age group	3	11	50

* Regional groups preset to total only receive data at nation and state levels.
X Not applicable.
Note: AIANNH is American Indian/Alaska Native/Native Hawaiian areas. Substate includes county, place, and census tract.
Source: U.S. Census Bureau.

How Does Disclosure Avoidance for Detailed DHC-A Compare With Other 2020 Census Products?

The 2020 Census Redistricting Data (P.L. 94-171) Summary File and the DHC data are protected through a disclosure avoidance system called the TopDown Algorithm. However, the data requirements for the Detailed DHC-A are different than those for the Redistricting Data (P.L. 94-171) Summary File and the DHC data. In particular, the number of categories of detailed race, ethnicity, and American Indian and Alaska Native tribes and villages far exceed the 128 categories used in the Redistricting Data (P.L. 94-171) Summary File and the DHC data. These differences led the Census Bureau to create a new differential privacy algorithm that was tailored to improve the quality of Detailed DHC-A published data.

As with Redistricting Data (P.L. 94-171) Summary File and the DHC data, SafeTab-P infuses noise into the 2020 Census data to protect respondent confidentiality. However, SafeTab-P differs from the TopDown Algorithm in several important ways (Table 4).

As Table 4 describes, in the TopDown Algorithm—and, thus, in the Redistricting Data (P.L. 94-171) Summary File, the DHC, and the Demographic Profile data—counts are produced first for the United States, and then counts for lower-level geographies are controlled to the U.S. counts so that all counts within the same universe (e.g., person tables or housing unit tables) are consistent. For example, the algorithm ensures that state totals, when added together, are consistent with the national data. Data users should be aware that data for very small geographic areas, such as census blocks, may be very noisy and should be aggregated into larger geographic areas before use.

In contrast, in the SafeTab-P Algorithm—and, thus, in the Detailed DHC-A—the data are not required to aggregate, or sum, as expected. For example, the state counts for detailed groups do not need to sum to the national counts for those groups. In fact, the more data you sum or combine, the less accurate the totals will be since you are also aggregating the noise.

For a detailed comparison of the differences between the TopDown and SafeTab-P differentially private algorithms, refer to Detailed Demographic and Housing Characteristics File A (Detailed DHC-A) Technical Documentation.

Data Suppression and Postprocessing

Because of the way SafeTab-P adds noise to the data, implausible results—like negative counts—are possible. To avoid misinterpretation, some Detailed DHC-A data are not available because they have been suppressed for data reasonableness concerns rather than for confidentiality protection. Reasons for suppression include:

- **Negative counts**. Because of the way noise is applied to data to protect confidentiality, some counts may be negative. These negative counts are suppressed and are denoted with an "X" in tables.

Table 4.
Differences Between the TopDown and SafeTab-P Differential Privacy Algorithms

Algorithm	Privacy	Geography	Accuracy When Aggregating Data	Consistency Across Products	Accuracy Targets	Design
TopDown (Redistricting Data, DHC, Demographic Profile)	Algorithm produces privacy-protected microdata.	All geographies can be aggregated as expected.	When aggregating data, the statistical noise generally cancels out, and the statistics become more accurate.	Consistent across data products.	Overall, accuracy can be targeted but the exact levels of accuracy cannot be known in advance.	Does not use adaptive design.
SafeTab-P (Detailed DHC-A)	Algorithm directly produces privacy-protected tabulations.	There is no requirement that geographies can be aggregated as expected.	When aggregating data, it generally becomes more variable the more you aggregate.	Not consistent with other 2020 Census data products.	All margins of error are determined in advance and met 95 percent of the time.	Uses adaptive design to determine the amount of data provided.

Note: DHC is Demographic and Housing Characteristics.
Source: U.S. Census Bureau.

- **Implausible counts for race groups**. The alone counts are suppressed when they are larger than their corresponding alone or in any combination counts. These suppressed counts are denoted with an "X" in tables.

In addition, the Census Bureau does not publish small, statistically unreliable counts for substate geographies and AIANNH areas. Detailed groups with populations of less than 22 and regional groups with populations of less than 94 will not have data published for that geography. These minimum population counts ensure that published population counts reflect positive enumerated counts 99.99 percent of the time.

Such minimum population counts in the 2020 Census are an improvement over the 2010 Census, which required a minimum population count of 100 for Asian, Native Hawaiian and Other Pacific Islander, and Hispanic origin groups to publish counts for any given geography level.

Geographic Equivalents

Although SafeTab-P processes each iteration independently by geography, a postprocessing routine ensures that equivalent geographies (i.e., are statistically equivalent) will have the same counts. For example, Washington, DC—which has the same boundaries whether it is listed as a state, county, or place—has the state counts regardless of which level of geography is used. Kalawao County, HI, is an example of a statistically equivalent geography where the county is comprised of two tracts, but only one of those tracts contains people. Kalawao county and the populated tract will have the same counts. As a result of this postprocessing, AIANNH areas may have counts for regional groups.

INCONSISTENT RESULTS AND OTHER LIMITATIONS

The Detailed DHC-A uses an adaptive design to meet the needs of data users who reported a critical need for accurate population counts of detailed race and ethnicity groups, including American Indian and Alaska Native tribes and villages. However, this method also has several limitations:

- Detailed groups may not sum to their corresponding regional group. Noise is added to regional groups and detailed groups separately. Because the two are processed independently, adding up the population of all detailed groups in a regional group may yield a different total than the total count for the regional group. Sums of detailed groups will also be noisier than their corresponding regional group. Similarly, a detailed group may be larger than its corresponding regional group due to noise being applied independently to detailed and regional groups.
- Values from lower-level geographies may not sum to values from higher-level geographies. For example, summing up all the county data for a specific group in a state may yield a different total than the count for that state.
- 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.
- Detailed DHC-A is not consistent with other 2020 Census data products. This includes the Redistricting Data (P.L. 94-171) Summary File, the DHC, and the Demographic Profile data.

GUIDANCE FOR WORKING WITH THE DETAILED DHC-A

When working with the Detailed DHC-A, data users should consider the following guidance.

Calculating Percentages

Data users may need or want to create several types of percentages using Detailed DHC-A data. The Census Bureau recommends using specific denominators (Table 5).

Dealing With Suppressed Counts

When an alone count has been suppressed, we recommend using the count from the equivalent alone or in any combination table, if available.

There are situations where a data user may be able to use subtraction to recreate a suppressed count such as when only one count in a sex by age table has been suppressed. We do not recommend using these counts because they may be implausible or statistically unreliable.

Comparing to Other Sources

There are some cautions users should consider when comparing Detailed DHC-A data to previous decennial censuses, the American Community Survey, and other 2020 Census data products. For detailed guidance, refer to the 2020 Census Demographic and Housing Characteristics File A (Detailed DHC-A) Technical Documentation.

Table 5.
Calculating Percentages in the Detailed DHC-A

Characteristic	Example		Source of denominator
	To calculate the percentage of . . .	Use . . . as the denominator	
Race (White, Black or African American, etc.)	The Asian alone population that is Korean alone at the national level	The total Asian alone population in the United States (19,886,049)	2020 Census Redistricting Data (P.L. 94-171) Summary File
Hispanic or Latino origin	People of Hispanic or Latino origin who are Salvadoran in the District of Columbia	The total Hispanic or Latino origin population in the District of Columbia (77,652)	2020 Census Redistricting Data (P.L. 94-171) Summary File
Geography (California; Harris County, Texas, etc.)	The population of Honolulu County, HI, that is Native Hawaiian alone or in any combination	The total population of Honolulu County, HI (1,016,508)	2020 Census Redistricting Data (P.L. 94-171) Summary File
Regional Group (European, Caribbean, etc.)	The Sub-Saharan African alone or in any combination population that is Beninese alone	The Sub-Saharan African alone or in any combination population (2,818,785)	2020 Census Detailed Demographic and Housing Characteristics File A

Source: U.S. Census Bureau.

Comparing to Previous Decennial Censuses

Comparisons between the 2020 Census and 2010 Census race data should be made with caution. Because of the improvements the Census Bureau made to the race question design, processing, and coding, users may observe differences in the data when comparing to other Census Bureau surveys or non-Census Bureau data sources.

Caution is not required when comparing detailed Hispanic origin data from the ethnicity question to the 2010 Census.

Comparing to the American Community Survey (ACS)

The 2020 Census provides the official counts (including Hispanic origin and race) of the population and housing units for the nation, states, counties, cities, and towns. The ACS provides estimates of certain additional characteristics of the population to add rich context for understanding the nation's population.

Users should not compare ACS MOEs, which summarize sampling variability, to the Detailed DHC-A MOEs referenced in this brief that describe how much noise has been applied to enumerated decennial counts. Although they share a name, they measure different concepts and are not comparable.

Comparisons using 2020 Census Detailed DHC-A data to ACS race data from 2019 or earlier should be made with caution, taking into account the improvements we have made to the Hispanic origin and race questions and the ways we code what people tell us. These improvements were applied to the ACS race and ethnicity questions and data processing starting with the 2020 ACS.

Caution is not required when comparing detailed Hispanic origin data from the ethnicity question to the ACS regardless of data year.

Comparing to the 2020 Census Redistricting Data (P.L. 94-171) Summary File, Demographic and Housing Characteristics File, and Demographic Profile

Comparisons between the 2020 Census Detailed DHC-A and the Redistricting Data (P.L. 94-171) Summary File, the DHC, and the Demographic Profile data should be made with caution. Because of the different disclosure avoidance methods used for these data products, the results may be inconsistent.

Hispanic Responses From the Ethnicity Question Compared to Hispanic Responses to the Race Question

Use the "Hispanic or Latino" responses for the official counts of detailed Hispanic origin groups (e.g., Mexican, Salvadoran, Dominican, etc.). These are the responses collected from the ethnicity question.

The Detailed DHC-A includes data on respondents who reported in the race question that they were (alone or alone or in any combination): Hispanic, Latin American, Mexican, Puerto Rican, Cuban, or another Hispanic, Latino, or Spanish response. These are not the official

counts for these Hispanic origin groups, rather, they are the official counts for people who gave these responses when asked to provide their race.

One of the best ways to tell these data apart is whether they include the terms "alone" or "alone or in any combination." If they do, such as "Puerto Rican alone," they are race data and are not the official Hispanic origin counts. If they do not, such as "Puerto Rican," they are Hispanic origin data and are the official Hispanic origin counts.

Aggregating Data

To create new aggregations of detailed groups, remove or add as few groups as possible. For example:

- To create a Central American count that includes Mexican, take the Central American total and add the Mexican count, rather than adding together all detailed Central American groups.
- To create a count of West African groups, add together the counts for the desired West African groups, such as Senegalese and Ghanaian, rather than subtracting non-West African detailed groups from the Sub-Saharan African total.

When using race data, data users should be mindful of whether they want to use the alone count or the alone or in any combination count.

To create counts for a custom geography, remove or add as few geographies as possible. For example:

- To create a count for the Pacific West states, add together the counts for Alaska, California, Hawaii, Oregon, and Washington, rather than subtracting the 46 other states and state equivalents from the national counts.
- To create a count for Arizona counties that are majority urban, remove Apache, Graham, Greenlee, Navajo and Santa Cruz counties from the Arizona state total rather than adding together the ten majority urban counties.

To create custom counts using age data, collapse as few categories as possible. For example:

- To compare groups that have 9 and 23 age categories, collapse the 23 age categories into the 9 rather than into four age categories or other custom categories when possible.

To learn more about how to calculate aggregated MOEs, reference the 2020 Census Demographic and Housing Characteristics File A (Detailed DHC-A) Technical Documentation.

WHERE CAN I LEARN MORE?

- Demonstration Data and Progress Metrics: Detailed Demographic and Housing Characteristics File A (Detailed DHC-A)

  <www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance/2020-das-development.html#detailed-dhc-a>

- 2020 Decennial Census: Processing the Count: Disclosure Avoidance Modernization

  <www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance.html>

- Disclosure Avoidance Webinar Series

  <www.census.gov/data/academy/webinars/2021/disclosure-avoidance-series.html>

For timely updates and to contact us, subscribe to the Census Bureau's 2020 Census Data Products Newsletter at <2020DAS@census.gov> or on the Decennial Census: Data Products and Operation Updates email sign-up page.

For key terms, refer to the "Glossary" in Disclosure Avoidance for the 2020 Census: An Introduction.