Calculating and Interpreting Average Household Size Ratios in the Demographic and Housing Characteristics File

The 2020 Census Supplemental Demographic and Housing Characteristics File (S-DHC) will provide official statistics on average household size for the total population and for race and ethnicity iterations, voting age, and tenure.[1] The S-DHC is planned for release in September 2024, but it will be limited to nation and state geographies only. To assist data users in estimating people per household ratios at lower geographic levels using 2020 Census data, this fact sheet provides guidance on calculating and interpreting the reliability of average household size ratios using published tables from the already released 2020 Census Demographic and Housing Characteristics File (DHC).[2] It focuses on average household size for total population and by race and ethnicity, voting age, and tenure for selected geographies: county, tract, place, and American Indian/Alaska Native/Native Hawaiian (AIANNH) areas. It provides information on multiple calculations and the limitations with each calculation (e.g., undefined ratios and proxy data).

METHODS

The U.S. Census Bureau conducted this analysis by comparing the 2010 Demonstration Data Product Suite (i.e., the Privacy Protected Microdata File [PPMF]) to the published 2010 Census data.[3] The 2010 PPMF was created by applying the 2020

disclosure avoidance system to the 2010 Census data. We used the 2010 PPMF for our analysis as it is currently publicly available, and we expect it to show similar accuracy results as the 2020 PPMF, which is planned for release later in 2024. The tables in the 2020 DHC can all be produced from the PPMF. We measure expected error introduced by the 2020 disclosure avoidance system, with the caveat that our errors are only in respect to the published 2010 Census data. The published 2010 Census data included 2010 disclosure avoidance error from data swapping, which swapped the geographical identifiers on records for certain households with the identifiers from nearby households with similar characteristics. The 2010 PPMF used census data that did not have swapping applied.

For the 2010 PPMF, the disclosure avoidance system added statistical noise—small, random additions and subtractions—to the enumerated data so that no persons or households can be reidentified with certainty using the published data. The settings for the disclosure avoidance system used for the 2010 PPMF were the same as those used for the 2020 DHC. The DHC has two underlying data files—the person file and the housing unit file. The person file contains individual-level characteristics (e.g., sex, age, race, ethnicity, and relationship to householder), and the unit file contains household- and householder-level characteristics (e.g., family or nonfamily household, family type, household size, tenure, and characteristics of the householder). The 2020 disclosure avoidance method for the DHC created independently protected person and unit files, creating inconsistencies across the two files and some implausible and impossible results.

---

[1] In addition, the S-DHC provides the count of people living in certain types of households.

[2] Note, some of these tables are also available in the American Community Survey (ACS), which is a sample-based survey. The 2020 Census is a complete count of the entire U.S. population.

[3] U.S. Census Bureau, "Factsheet on Disclosure Avoidance for the 2010 Demonstration Data Products Suite — Redistricting and Demographic and Housing Characteristics File — Production Settings (2023-04-03)," <www2.census.gov/programs-surveys/decennial/2020/program-management/data-product-planning/2010-demonstration-data-products/04-Demonstration_Data_Products_Suite/2023-04-03/2023-04-03_Factsheet.pdf>.

CALCULATIONS FOR AVERAGE HOUSEHOLD SIZE FOR THE TOTAL POPULATION

The average household size for the total population can be calculated using various combinations of the two underlying data files. We limited our analysis to the following three calculations:

Person file only:

$$\frac{\text{Count of people in households from the person file}}{\text{Count of householders from the person file}}$$

Unit file only:

$$\frac{\text{Sum of household size from the unit file}}{\text{Count of occupied housing units from the unit file}}$$

Combination:

$$\frac{\text{Count of people in households from the person file}}{\text{Count of occupied housing units from the unit file}}$$

These three ratios can be calculated from the published 2020 DHC tables using H8 and P17 for the person file only, H9 and H3 for the unit file only, and H8 and H3 for the combination calculation, or they can be calculated using the 2020 PPMF, once the file is released. Note that the person file only and unit file only calculations use proxy information to derive the ratios. For the person file only calculation, the count of householders is a proxy for the number of households. For the unit file only calculation, household size is used as a proxy for the total count of people in a household. To calculate the total number of people per household from the unit file using household size, you must sum across each household size multiplied by the number of units within each size. For example, for an area with ten households with household size equal to one

and five households with a household size equal to two, the total number of persons is (10 x 1) + (5 x 2) = 20. Household size in the unit file is top-coded at seven or more people, so this calculation results in a slight undercount of people per households. In the case of a household containing ten individuals, the household size variable would indicate seven or more people, leading to an underestimation of three people.

All three calculations can result in a ratio of 0/0, and for our analysis, we treat these results as zero instead of undefined. The person file only and combination calculations can result in a ratio where the numerator has a non-zero value but the denominator is zero, resulting in division by zero. These implausible results are a consequence of the 2020 disclosure avoidance system. For the person file only calculation, a given geography may have household population in the numerator but no householders in the denominator. For the combination calculation, the person file may have household population for a given geography without occupied housing units in the housing unit file.[4] Our analysis treated these division-by-zero cases as nulls, and they were dropped from summary calculations (such as means and percentiles).

The combination calculation could also result in extremely large averages, indicating a geography has households with an unexpectedly high number of people living in them. This can happen in small geographical areas and could be a result of noise infusion (if the random noise increased the number of persons and decreased the number of occupied housing units in the geography) or a data collection issue.

_____

[4] Housing units were held invariant down to the block level (and therefore published as enumerated), but occupied housing units were not held invariant. After disclosure avoidance is applied, an occupied housing unit could become vacant and vice versa.

CALCULATIONS FOR AVERAGE HOUSEHOLD SIZE BY RACE AND ETHNICITY ITERATION

The calculations for race and ethnicity are very similar to the calculations for total population because each person's race and ethnicity are provided on the person file, and each householder's race and ethnicity are provided on the unit file. Similar to the average household size for the total population, the average household size by race and ethnicity can be calculated using various combinations of the two underlying data files. We limited our analysis to the following three calculations:

Person file only:

$$\frac{\text{Count of people in households of a given race and ethnicity from the person file}}{\text{Count of householders of the given race and ethnicity from the person file}}$$

Unit file only:

$$\frac{\text{Sum of household size from units with a householder of the given race and ethnicity from the unit file}}{\text{Count of occupied housing units with a householder of the given race and ethnicity from the unit file}}$$

Combination:

$$\frac{\text{Count of people in households of a given race and ethnicity from the person file}}{\text{Count of occupied housing units with a householder of the given race and ethnicity from the unit file}}$$

These three ratios can be calculated from the published 2020 DHC tables using PCT13A-I and PCT17A-I for the person file only, PCT7A-I and P16A-I for the unit file only, and PCT13A-I and P16A-I for the combination calculation, or they can be calculated using the 2020 PPMF once released.

These calculations use the same proxy data as was used for the average household size for the total population. In addition, the person file only and combination calculations for race and ethnicity use an additional proxy. To obtain the true average household size by race and ethnicity, the unit file only calculation sums household size from units with a householder of a given race divided by the count of housing units with a householder of that race, which allows for multiracial households (but top codes household size at seven or more people). Unfortunately, we cannot perform that calculation when using the PPMF person file because it does not contain information linking individuals to other members of their household. Instead of using the count of people in a household with a householder of a given race, we use the count of people of a given race living in households. For the person file only and combination calculations, the proxy measure incorrectly assumes that everyone in the household is the same race and ethnicity of the householder. Therefore, multiracial households are not captured when using the person file only and combination calculations.

ACCURACY BY COUNT OF OCCUPIED HOUSING UNITS

As expected, we found the ratios tended to be less reliable for small counts. Many times, even small amounts of error resulted in large relative changes when there was a small number of households in the denominator of the ratio. To demonstrate the increasing reliability for ratios calculated with increasing numbers of households, refer to Table 1, which shows the mean absolute error (MAE) for the three ratio calculations for tracts. Results are grouped into bins based on number of occupied units. Error was calculated by subtracting the ratio calculated using the 2010 Census published counts from the ratio calculated using the 2010 DHC Demonstration Data. MAE was calculated by taking the average of the absolute errors. Our full analysis focused on county and tract level ratios, MAEs, and 90th percentiles of the absolute error for the total population and nine race and ethnicity iterations.

From our full analysis, we saw that at roughly around the 100–200 occupied housing units group level (generalizing across the geographic levels and the nine race and ethnicity iterations), implausible ratio results from the application of disclosure avoidance were less likely, and the MAEs and 90th percentiles of the absolute error became more comparable to the magnitude of error of similar statistics from the ACS. For this reason, we recommend limiting ratio calculations to cells with at least 100 households (i.e., greater than or equal to 100 in the denominator). Therefore, the remainder of our analysis is limited to those cases.

ACCURACY OF AVERAGE HOUSEHOLD SIZE FOR THE TOTAL POPULATION AND BY RACE AND ETHNICITY ITERATION

We are sharing the following tables for data users to reference when deciding which calculation method to use for their analysis and to understand the magnitude of the errors associated with each calculation.

Table 2 shows the number of times each of the three ratio calculations was the most accurate for county, tract, place, and AIANNH areas. All three

ratios and the ratio calculated from 2010 Census published counts were rounded to the hundredths place, and then the three calculation methods were compared with the 2010 Census ratio to determine which was the most accurate. When there were ties (multiple ratios had the same accuracy), the tied ratios were all counted as the most accurate. Therefore, summing across the three ratios results in a count greater than the total number of geographic units.

Table 2 shows which calculations performed the best. The person file only calculation performed the best most often at each geographic level for the total population. Out of 3,219 counties, the person file only calculation performed best for 2,922 counties, the combination calculation performed best for 2,572 counties, and the unit file only calculation performed best for 697 counties. For race and ethnicity iterations, the most accurate calculation varied between the unit file only calculation and the combination calculation, depending on the race and ethnicity iteration and geographic level. Data users should reference Table 2 when determining which calculation to use for race and ethnicity iterations.

Table 1.
Tract-Level Mean Absolute Error of People Per Household

Number of occupied units for total population or race and ethnicity iteration	Total			White			Black			Hispanic		
	Person file only	Unit file only	Combination of both	Person file only	Unit file only	Combination of both	Person file only	Unit file only	Combination of both	Person file only	Unit file only	Combination of both
0[1]	–	–	–	1.32	1.66	1.06	2.01	2.44	1.76	3.11	1.90	3.63
1–5	1.40	1.84	4.41	1.80	1.39	1.61	2.35	1.43	2.12	2.73	1.23	2.73
5–20	0.62	0.87	1.42	1.36	0.70	0.89	1.27	0.67	0.90	1.68	0.61	1.29
20–50	0.24	0.52	0.64	0.53	0.40	0.45	0.51	0.38	0.35	0.92	0.38	0.70
50–100	0.13	0.37	0.22	0.31	0.26	0.25	0.29	0.24	0.18	0.60	0.27	0.45
100–200	0.07	0.21	0.11	0.19	0.19	0.16	0.17	0.15	0.11	0.40	0.22	0.31
200–400	0.03	0.12	0.04	0.12	0.14	0.10	0.10	0.10	0.07	0.25	0.18	0.20
400–600	0.02	0.10	0.02	0.07	0.09	0.06	0.06	0.08	0.04	0.15	0.17	0.13
600–1000	0.01	0.08	0.01	0.05	0.06	0.05	0.04	0.07	0.03	0.10	0.18	0.08
1000–5000	0.01	0.05	0.01	0.04	0.03	0.03	0.03	0.06	0.02	0.05	0.11	0.04
5000+	0.00	0.02	0.00	0.03	0.01	0.03	–	–	–	–	–	–

[1] We treated ratios of 0/0 as 0, instead of undefined.
Source: U.S. Census Bureau.

Table 2.
Number of Times Each Ratio Was the Most Accurate

Total population or race and ethnicity iteration	County				Tract				Place				AIANNH			
	Number of counties	Person file only	Unit file only	Combination of both	Number of tracts	Person file only	Unit file only	Combination of both	Number of places	Person file only	Unit file only	Combination of both	Number of AIANNH areas	Person file only	Unit file only	Combination of both
Total population	3,219	2,922	697	2,572	72,858	51,225	10,074	49,273	23,756	15,138	3,539	13,468	331	222	43	169
A. White alone	3,219	978	2,293	1,040	70,151	23,616	35,042	28,057	22,695	8,861	9,063	9,340	162	28	100	53
B. Black or African American alone	1,833	591	695	946	28,883	9,062	7,885	15,518	5,769	1,987	1,756	2,782	40	10	20	15
C. American Indian and Alaska Native alone	1,088	312	485	341	1,015	281	390	390	1,302	355	509	499	258	80	82	110
D. Asian alone	982	135	786	109	11,792	3,473	4,021	5,285	3,063	773	1,632	856	24	6	15	3
E. Native Hawaiian and Other Pacific Islander alone	149	45	46	66	130	37	31	69	271	65	81	138	19	3	8	8
F. Some Other Race alone	1,526	458	655	584	15,150	4,561	3,620	7,990	3,552	1,189	1,044	1,589	41	12	17	15
G. Two or More Races	1,626	45	1,569	17	1,812	153	1,407	283	2,644	114	2,392	156	65	7	57	1
H. Hispanic or Latino	2,104	362	1,646	248	30,174	8,333	13,033	10,832	6,267	1,514	3,705	1,362	73	12	53	11
I. White alone, not Hispanic or Latino . .	3,165	634	2,514	614	66,997	16,341	44,837	17,347	22,070	7,148	10,836	7,301	156	25	102	39

Note: Analysis was restricted to geographical units with 100 or more households for a total population or by race and ethnicity iteration.
Source: U.S. Census Bureau.

Table 3 shows the MAE and 90th percentiles of the absolute error at the county level. For the total population, using the person file only calculation, the mean absolute difference across counties between published and demonstration data is 0.00; this tells us that on average, the DHC demonstration data ratio differs from the published 2010 Census ratio by ±0.00. The 90th percentile of the absolute errors is 0.01; i.e., 90 percent of counties have an absolute difference of 0.01 or less. The MAE for the race and ethnicity iterations ranged from 0.02 (for White alone, not Hispanic or Latino, using the unit file only calculation) to 1.40 (for Two or More Races, using the person file only calculation). The unit file only calculation, which did not use proxy information for race and ethnicity iterations, performed much better than the other two calculation methods for Two or More Races, with a MAE of 0.12.

Tables 4 through 6 provide the same information as Table 3, but they include results for census tracts, places, and AIANNH areas.

Table 3.
County-Level Accuracy of Average People Per Household Ratios

Total population or race and ethnicity iteration	Person file only		Unit file only		Combination	
	MAE	90th percentile	MAE	90th percentile	MAE	90th percentile
Total population	0.00	0.01	0.03	0.06	0.00	0.01
A. White alone.	0.03	0.05	0.02	0.05	0.03	0.05
B. Black or African American alone.	0.14	0.38	0.09	0.25	0.09	0.24
C. American Indian and Alaska Native alone	0.19	0.42	0.14	0.31	0.16	0.34
D. Asian alone	0.37	0.84	0.12	0.31	0.31	0.61
E. Native Hawaiian and Other Pacific Islander alone	0.26	0.59	0.22	0.44	0.17	0.35
F. Some Other Race alone	0.22	0.54	0.14	0.30	0.17	0.39
G. Two or More Races	1.40	2.42	0.12	0.28	1.38	2.32
H. Hispanic or Latino	0.47	1.03	0.14	0.34	0.42	0.83
I. White alone, not Hispanic or Latino. . .	0.04	0.07	0.02	0.05	0.04	0.07

Note: analysis was restricted to geographical units with 100 or more households for total population or by race and ethnicity iteration.
Source: U.S. Census Bureau.

Table 4.
Tract-Level Accuracy of Average People Per Household Ratios

Total population or race and ethnicity iteration	Person file only		Unit file only		Combination	
	MAE	90th percentile	MAE	90th percentile	MAE	90th percentile
Total population	0.01	0.02	0.05	0.12	0.01	0.02
A. White alone.	0.05	0.10	0.05	0.11	0.04	0.09
B. Black or African American alone.	0.10	0.25	0.11	0.23	0.07	0.16
C. American Indian and Alaska Native alone	0.21	0.46	0.18	0.35	0.18	0.42
D. Asian alone	0.15	0.34	0.14	0.31	0.11	0.24
E. Native Hawaiian and Other Pacific Islander alone	0.28	0.59	0.37	0.65	0.19	0.43
F. Some Other Race alone	0.17	0.37	0.21	0.44	0.11	0.24
G. Two or More Races	0.85	1.75	0.19	0.40	0.72	1.54
H. Hispanic or Latino	0.25	0.60	0.19	0.40	0.20	0.47
I. White alone, not Hispanic or Latino. . .	0.07	0.14	0.04	0.09	0.07	0.14

Note: Analysis was restricted to geographical units with 100 or more households for total population or by race and ethnicity iteration.
Source: U.S. Census Bureau.

Table 5.

Place-Level Accuracy of Average People Per Household Ratios

Total population or race and ethnicity iteration	Person file only		Unit file only		Combination	
	MAE	90th percentile	MAE	90th percentile	MAE	90th percentile
Total population	0.03	0.08	0.09	0.22	0.03	0.07
A. White alone.	0.06	0.12	0.08	0.20	0.05	0.11
B. Black or African American alone.	0.11	0.28	0.11	0.25	0.09	0.22
C. American Indian and Alaska Native alone	0.22	0.46	0.21	0.46	0.18	0.39
D. Asian alone	0.20	0.47	0.12	0.28	0.18	0.41
E. Native Hawaiian and Other Pacific Islander alone	0.29	0.66	0.30	0.66	0.20	0.43
F. Some Other Race alone	0.18	0.43	0.19	0.39	0.14	0.33
G. Two or More Races	1.37	2.52	0.16	0.35	1.27	2.32
H. Hispanic or Latino	0.34	0.79	0.17	0.39	0.31	0.66
I. White alone, not Hispanic or Latino. . .	0.07	0.14	0.07	0.19	0.07	0.13

Note: analysis was restricted to geographical units with 100 or more households for total population or by race and ethnicity iteration.
Source: U.S. Census Bureau.

Table 6.

AIANNH Area Tract-Level Accuracy of Average People Per Household Ratios

Total population or race and ethnicity iteration	Person file only		Unit file only		Combination	
	MAE	90th percentile	MAE	90th percentile	MAE	90th percentile
Total population	0.03	0.08	0.15	0.36	0.04	0.11
A. White alone.	0.14	0.32	0.08	0.18	0.12	0.34
B. Black or African American alone.	0.13	0.33	0.10	0.21	0.14	0.42
C. American Indian and Alaska Native alone	0.18	0.36	0.19	0.42	0.15	0.36
D. Asian alone	0.48	1.20	0.19	0.35	0.43	0.93
E. Native Hawaiian and Other Pacific Islander alone	0.47	0.70	0.35	0.66	0.33	0.62
F. Some Other Race alone	0.16	0.46	0.15	0.32	0.12	0.32
G. Two or More Races	1.14	2.01	0.21	0.51	1.10	1.76
H. Hispanic or Latino	0.66	1.15	0.17	0.38	0.60	1.15
I. White alone, not Hispanic or Latino. . .	0.14	0.27	0.07	0.16	0.13	0.35

Note: analysis was restricted to geographical units with 100 or more households for total population or by race and ethnicity iteration.
Source: U.S. Census Bureau.

CALCULATIONS FOR AVERAGE HOUSEHOLD SIZE BY VOTING AGE

To assess the accuracy of average household size by voting age, the analysis compared ratios calculated using the 2010 Census published counts to ratios calculated based on demonstration data, by race and Hispanic origin and voting age. Average household size by voting age could be calculated using the person file only or the combination of person and unit file data; however, our supplemental analysis showed the person file only calculation tended to produce similar results to the combination calculation, but in some cases, it was slightly less accurate. Therefore, we limited this fact sheet to show the results of using the combination calculation only.

Average household size for children under 18 years of age:

$$\frac{\text{Count of people under 18 years of age in households from the person file}}{\text{Count of occupied housing units from the unit file}}$$

Average household size for adults 18 years of age or older:

$$\frac{\text{Count of people 18 years of age or older in households from the person file}}{\text{Count of occupied housing units from the unit file}}$$

Average household size by voting age for the race and ethnicity iterations was calculated like the average household size combination calculation for the race and ethnicity iterations; for example, average number of children under 18 per household (White alone householder) was calculated using the number of people under 18 that were White alone from the person file divided by the number of occupied housing units with a White alone householder from the unit file.

Averages by voting age can be calculated for the total population using the published 2020 DHC tables P15 and H3 or the PPMF. Calculations by the race and ethnicity iterations require using tables PCT13A-I and P16A-I, or the PPMF. As before, implausible and impossible scenarios are possible. Division by zero can happen when there are people for a given geography but no occupied housing units, or people of a given race or ethnicity but no householders of that race or ethnicity.

The analysis was limited to use cases with 100 or more occupied housing units in the denominator. For example, we only calculated average number of children under 18 years per household (White alone householder), when there were at least 100 households with a householder who was White alone.

ACCURACY OF AVERAGE HOUSEHOLD SIZE BY VOTING AGE

For the analysis on average household size by voting age, we used one calculation method and assessed the magnitude of the errors. Table 7 presents the MAE and the 90th percentile absolute difference between the published and demonstration data ratios. For the total population, the average difference across counties between the published and demonstration data for children under 18 years per household is 0.00. The average absolute difference in adults per household is also 0.00. In comparison, results by race and ethnicity have larger errors. Two or More Races tends to perform the worst; on average, the DHC ratio differs from the 2010 Census ratio for Two or More Races by ±1.27 for children under 18 per household. For Two or More Races, 90 percent of counties have an absolute difference of 2.12 or less for average children per household.

Tables 8 through 10 provide the same information, but they include results for census tracts, places, and AIANNH areas.

Table 7.

County-Level Accuracy of Average Number of Children and Average Number of Adults Per Household by Race and Hispanic Origin

Total population or race and ethnicity iteration	Absolute difference in ratio of children < 18 years per household		Absolute difference in ratio of people ≥ 18 years per household	
	MAE	90th percentile	MAE	90th percentile
Total population	0.00	0.00	0.00	0.01
A. White alone.	0.02	0.03	0.01	0.02
B. Black or African American alone.	0.05	0.14	0.05	0.13
C. American Indian and Alaska Native alone	0.10	0.22	0.09	0.18
D. Asian alone	0.07	0.16	0.28	0.54
E. Native Hawaiian and Other Pacific Islander alone	0.11	0.21	0.18	0.38
F. Some Other Race alone	0.09	0.21	0.11	0.23
G. Two or More Races	1.27	2.12	0.14	0.29
H. Hispanic or Latino	0.30	0.61	0.12	0.25
I. White alone, not Hispanic or Latino. . .	0.03	0.05	0.01	0.02

Note: analysis was restricted to geographical units with 100 or more households for total population or by race and ethnicity iteration.
Source: U.S. Census Bureau.

Table 8.

Tract-Level Accuracy of Average Number of Children and Average Number of Adults Per Household by Race and Hispanic Origin

Total population or race and ethnicity iteration	Absolute difference in ratio of children < 18 years per household		Absolute difference in ratio of people ≥ 18 years per household	
	MAE	90th percentile	MAE	90th percentile
Total population	0.01	0.01	0.01	0.02
A. White alone.	0.03	0.06	0.02	0.04
B. Black or African American alone.	0.04	0.09	0.04	0.11
C. American Indian and Alaska Native alone	0.12	0.30	0.09	0.21
D. Asian alone	0.05	0.10	0.11	0.23
E. Native Hawaiian and Other Pacific Islander alone	0.10	0.22	0.14	0.31
F. Some Other Race alone	0.07	0.14	0.08	0.17
G. Two or More Races	0.63	1.31	0.18	0.38
H. Hispanic or Latino	0.13	0.31	0.08	0.20
I. White alone, not Hispanic or Latino . . .	0.04	0.09	0.03	0.06

Note: analysis was restricted to geographical units with 100 or more households for total population or by race and ethnicity iteration.
Source: U.S. Census Bureau.

Table 9.

Place-Level Accuracy of Average Number of Children and Average Number of Adults Per Household by Race and Hispanic Origin

Total population or race and ethnicity iteration	Absolute difference in ratio of children < 18 years per household		Absolute difference in ratio of people ≥ 18 years per household	
	MAE	90th percentile	MAE	90th percentile
Total population	0.01	0.03	0.02	0.06
A. White alone.	0.03	0.06	0.03	0.07
B. Black or African American alone.	0.05	0.11	0.06	0.14
C. American Indian and Alaska Native alone	0.10	0.23	0.11	0.24
D. Asian alone	0.06	0.13	0.18	0.37
E. Native Hawaiian and Other Pacific Islander alone	0.14	0.29	0.17	0.36
F. Some Other Race alone	0.08	0.17	0.11	0.23
G. Two or More Races	1.16	2.09	0.18	0.38
H. Hispanic or Latino	0.20	0.44	0.12	0.26
I. White alone, not Hispanic or Latino . . .	0.04	0.08	0.03	0.07

Note: analysis was restricted to geographical units with 100 or more households for total population or by race and ethnicity iteration.
Source: U.S. Census Bureau.

Table 10.

AIANNH Area Accuracy of Average Number of Children and Average Number of Adults Per Household by Race and Hispanic Origin

Total population or race and ethnicity iteration	Absolute difference in ratio of children < 18 years per household		Absolute difference in ratio of people ≥ 18 years per household	
	MAE	90th percentile	MAE	90th percentile
Total population	0.02	0.05	0.03	0.08
A. White alone.	0.09	0.20	0.08	0.21
B. Black or African American alone.	0.05	0.16	0.10	0.27
C. American Indian and Alaska Native alone	0.07	0.19	0.10	0.26
D. Asian alone	0.06	0.11	0.46	0.90
E. Native Hawaiian and Other Pacific Islander alone	0.14	0.20	0.20	0.44
F. Some Other Race alone	0.07	0.17	0.10	0.23
G. Two or More Races	0.96	1.63	0.18	0.32
H. Hispanic or Latino	0.41	0.72	0.20	0.38
I. White alone, not Hispanic or Latino . . .	0.10	0.21	0.08	0.21

Note: analysis was restricted to geographical units with 100 or more households for total population or by race and ethnicity iteration.
Source: U.S. Census Bureau.

CALCULATIONS FOR AVERAGE HOUSEHOLD SIZE BY TENURE

To assess the accuracy of average household size by tenure, we calculated average household size for owned housing units and average household size for rented housing units. Tenure is only available on the unit file, so these ratios only have one possible calculation method.

Average household size for owned, occupied housing units:

$$\frac{\text{Sum of household size in owned and occupied units from the unit file}}{\text{Count of owned and occupied housing units from the unit file}}$$

Average household size for rented, occupied housing units:

$$\frac{\text{Sum of household size in rented and occupied units from the unit file}}{\text{Count of rented and occupied housing units from the unit file}}$$

Average household size by tenure for the race and ethnicity iterations was calculated like the unit file only average household size calculation for the race and ethnicity iterations; for example, average household size in owned households (White alone householder) was calculated using the sum of household size from owned households with a White alone householder from the unit file divided by the number of occupied, owned housing units with a White alone householder from the unit file.

These tenure averages can be calculated using the published 2020 DHC tables for the total population (H12 and H4) and by race and ethnicity iterations (H12A-I and H4A-I) or using the PPMF. The average household size by tenure calculation can not result in division by zero. However, household size from the unit file is capped at seven people; therefore, there is a slight undercount of the number of people in housing units. The analysis was restricted to use cases with 100 or more occupied housing units in the denominator. For example, we only calculated ratios for average household size in owned households (White alone householder) when there were at least 100 owned, White householder housing units.

ACCURACY OF TENURE RATIOS

For the analysis on average household size by tenure, we used one calculation method and assessed the magnitude of the errors. Table 11 presents the MAE and 90th percentile of the absolute error between the published and demonstration data ratios at the county level. For the total population, the average absolute difference across counties between published and demonstration data for average household size in owned, occupied housing units is 0.02. Therefore, on average, the DHC ratio differs from the 2010 Census ratio by ±0.02. For total population in owned, occupied housing units, 90 percent of counties have an absolute difference of 0.05 or less. The average absolute difference in rented,

occupied housing units is 0.04, with 90 percent of counties having a difference of 0.10 or less. For the race and ethnicity iterations, the differences are larger. For example, for Native Hawaiian and Other Pacific Islander alone households, the average absolute difference is 0.21 for average household size in owned, occupied housing units, and 0.22 for average household size in rented, occupied housing units.

Tables 12 through 14 provide the same information as Table 11, but they include results for census tracts, places, and AIANNH areas.

Table 11.

County-Level Accuracy of Average Household Size in Owned and Rented Occupied Housing Units by Race and Hispanic Origin

Total population or race and ethnicity iteration	Absolute difference in average household size of owned housing units		Absolute difference in average household size of rented housing units	
	MAE	90th percentile	MAE	90th percentile
Total population	0.02	0.05	0.04	0.10
A. White alone.	0.02	0.04	0.04	0.09
B. Black or African American alone.	0.07	0.18	0.07	0.18
C. American Indian and Alaska Native alone	0.13	0.28	0.14	0.30
D. Asian alone	0.12	0.27	0.09	0.23
E. Native Hawaiian and Other Pacific Islander alone	0.21	0.45	0.22	0.38
F. Some Other Race alone	0.15	0.33	0.12	0.27
G. Two or More Races	0.11	0.24	0.11	0.24
H. Hispanic or Latino	0.13	0.31	0.12	0.29
I. White alone, not Hispanic or Latino . . .	0.02	0.04	0.03	0.09

Note: analysis was restricted to geographical units with 100 or more households for total population or by race and ethnicity iteration.
Source: U.S. Census Bureau.

Table 12.

Tract-Level Accuracy of Average Household Size in Owned and Rented Occupied Housing Units by Race and Hispanic Origin

Total population or race and ethnicity iteration	Absolute difference in average household size of owned housing units		Absolute difference in average household size of rented housing units	
	MAE	90th percentile	MAE	90th percentile
Total population	0.06	0.13	0.08	0.19
A. White alone.	0.05	0.11	0.08	0.17
B. Black or African American alone.	0.11	0.23	0.10	0.22
C. American Indian and Alaska Native alone	0.17	0.35	0.19	0.38
D. Asian alone	0.15	0.32	0.11	0.26
E. Native Hawaiian and Other Pacific Islander alone	0.46	0.95	0.35	0.73
F. Some Other Race alone	0.27	0.58	0.20	0.41
G. Two or More Races	0.20	0.40	0.16	0.37
H. Hispanic or Latino	0.21	0.46	0.18	0.39
I. White alone, not Hispanic or Latino . . .	0.04	0.09	0.06	0.15

Note: analysis was restricted to geographical units with 100 or more households for total population or by race and ethnicity iteration.
Source: U.S. Census Bureau.

Table 13.

Place-Level Accuracy of Average Household Size in Owned and Rented Occupied Housing Units by Race and Hispanic Origin

Total population or race and ethnicity iteration	Absolute Difference in Average Household Size of Owned Housing Units		Absolute Difference in Average Household Size of Rented Housing Units	
	MAE	90th percentile	MAE	90th percentile
Total population	0.08	0.19	0.09	0.21
A. White alone.	0.07	0.17	0.07	0.17
B. Black or African American alone.	0.10	0.23	0.09	0.21
C. American Indian and Alaska Native alone	0.20	0.42	0.18	0.39
D. Asian alone	0.13	0.29	0.10	0.23
E. Native Hawaiian and Other Pacific Islander alone	0.37	0.84	0.28	0.63
F. Some Other Race alone	0.21	0.45	0.16	0.34
G. Two or More Races	0.17	0.37	0.15	0.31
H. Hispanic or Latino	0.18	0.39	0.16	0.36
I. White alone, not Hispanic or Latino . . .	0.06	0.16	0.06	0.15

Note: analysis was restricted to geographical units with 100 or more households for total population or by race and ethnicity iteration.
Source: U.S. Census Bureau.

Table 14.

AIANNH Area-Level Accuracy of Average Household Size in Owned and Rented Occupied Housing Units by Race and Hispanic Origin

Total population or race and ethnicity iteration	Absolute difference in average household size of owned housing units		Absolute difference in average household size of rented housing units	
	MAE	90th percentile	MAE	90th percentile
Total population	0.14	0.32	0.12	0.28
A. White alone.	0.06	0.14	0.08	0.19
B. Black or African American alone.	0.10	0.18	0.09	0.23
C. American Indian and Alaska Native alone	0.18	0.35	0.17	0.31
D. Asian alone	0.21	0.44	0.13	0.23
E. Native Hawaiian and Other Pacific Islander alone	0.37	0.78	0.16	0.25
F. Some Other Race alone	0.15	0.28	0.10	0.20
G. Two or More Races	0.19	0.51	0.13	0.31
H. Hispanic or Latino	0.17	0.45	0.19	0.47
I. White alone, not Hispanic or Latino . . .	0.06	0.14	0.08	0.16

Note: analysis was restricted to geographical units with 100 or more households for total population or by race and ethnicity iteration.
Source: U.S. Census Bureau.

GUIDANCE FOR DATA USERS

Upon completion of our analysis, we have the following recommendations for calculating average household size for the total population, and by race and ethnicity iteration, voting age, and tenure.

- Use the S-DHC tables for nation and state-level ratios. The S-DHC provides the official statistics for people per household ratios.

- Limit ratio calculations to areas with a larger number of occupied housing units per cell (100 households or more) by aggregating geographies or combining small categories. This produces more reliable statistics; it prevents division by zero calculations and reduces the possibility of disclosure avoidance random noise producing large relative changes in the ratios.

- For average household size for the total population, use the person file only calculation, which performed better than the other two calculation methods.

- For average household size by race and ethnicity, reference Table 2 for the best calculation method for your use case, as results vary by race and ethnicity iteration and geography. For averages for Two or More Races, use the unit file only calculation, as it had much better accuracy than the other two calculations.

- Use this fact sheet as a resource to help determine if the data are fit for your particular use case. For voting age and tenure, some results may not be accurate enough for specific use cases.

- When presenting results, especially for small populations, clearly convey the noise-infused error associated with the statistics whenever it is possible.