Summary of U.S. Census Bureau's 2020 Disclosure Avoidance System (DAS) and the Effects of Differential Privacy on Massachusetts Census 2020 Data

Prepared by:
UMass Donahue Institute
Economic and Public Policy Research
Population Estimates Program

October 26, 2021

Background

The Census Bureau has adapted the technique of *differential privacy* as part of their new Disclosure Avoidance System (DAS) to uphold Census data confidentiality and to prevent personal re-identification threats to the 2020 Census data products. The Bureau's implemented DAS affects the underlying data by adding "noise" at every geographic level except for the total state population, which remains invariant. This "noise" changes the enumerated population of every other geographic level including counties, minor civil divisions (MCDs), census tracts, and blocks as well as the age and racial makeup of these populations for privacy purposes. Unlike population, the counts of housing units at all levels of geography are unperturbed by the new DAS.

Starting in 2018, the Census Bureau has tested various DAS applications and released results in a series of DAS Privacy Protected Microdata Files (PPMF), which are essentially the Census 2010 SF1 data with differential privacy applied, as well as an accompanying detailed summary metrics file. The Census Bureau released these PPMF files as "demonstration data' so that the researchers and interested parties can understand how perturbed the publicly released Census 2020 data might be based on how perturbed the Census 2010 data looks after differential privacy is applied to it. The difference between the original Census 2010 data and the demonstration data is referred to as the "error" when evaluating the demonstration data and is also displayed in the detailed summary metrics file.

On July 1, 2021, the U.S. Census Bureau released the detailed summary metrics for the final production settings parameters chosen by the Data Stewardship Executive Policy Making Committee (DSEP) on June 8, 2021. This is the sixth set of detailed summary metrics that the Census Bureau has released. In addition, the Census Bureau released the sixth and final set of privacy-protected microdata files (PPMFs) on August 12th, 2021 using the June 8 DAS together with the PL94-171 Census data file to which this version of the DAS is applied. This summary analyzes the June 8, 2021 DAS PPMF demonstration data released on August 12, 2021 and the corresponding summary metrics file from the Census Bureau.

The global privacy loss budget (PLB) chosen by the DSEP for the 2020 Census Redistricting data (PL-94-171) that is displayed in the new DAS PPMF File and summary metrics is 19.6. The Privacy Loss Budget is the measured tradeoff between privacy and accuracy that the Census Bureau is in control of. The PLB is represented by a numerical value epsilon, ε, and was previously tested at 4.5 and 12.2 before the Census Bureau settled on 19.6. According to the Bureau, a higher epsilon value represents greater accuracy in the data but an increased risk of privacy loss. The first four demonstration data releases,

which were published in October 2019, May 2020, September 2020, and November 2020, were algorithmically "tuned" for complete privacy where the fifth April 2021 release was "tuned" to ensure that the Census Bureau met the accuracy targets that were established for redistricting and Voting Rights Act enforcement. The final demonstration data release focused on making the data more reliable especially for "smaller block groups, Minor Civil Divisions, and places" according to the Census Bureau. The following analysis looks into the "error" and differences across various geographic levels, racial and ethnic groups, and age cohorts to see the potential effects of differential privacy on Massachusetts Census 2020 Data.

It is important to note that *the specific cities and towns and their particular levels of error found in this analysis of the DAS effects on Census 2010 are not necessarily the same places and errors that are present in the actual PL-94 data*. For one, the effects of DAS are randomized in the actual application; a city that gains population in the demonstration data may have lost population in the actual application or vice-versa. Also, the population counts in Massachusetts for many geographies are higher in 2020 than in 2010, meaning that ranges of error may also be higher in the actual DAS application to 2020 data compared to error in the 2010 demonstration data. When evaluating the potential effects of the final DAS application on any particular geography or cohort, it will be most important to bear in mind the size of the place or cohort and to understand that the error could have either added or subtracted population to the final published data, but that users will not know which.

Massachusetts Analysis

State

As previously mentioned, the State total population must remain unchanged and therefore the differences between the 2010 SF1 data and the DAS PPMF demonstration data are minuscule. After breaking down the State total population by race and ethnicity, the range of error is -31 people to 23 people with no racial and ethnic category seeing a percent error above 0.6% (Table 1). When analyzing the errors and percent errors, it is important to note that smaller, less populated racial and ethnic groups will experience larger percent errors due to their smaller denominators even if the actual error is quite small.

Table 1 Massachusetts Total Population Error and Percent Error by Race

Race/Ethnicity Groups	SF1	DP	Difference	Percent Difference
Hispanic or Latino (All Races)	627,654	627,677	23	0.00%
Not Hispanic White alone	4,984,800	4,984,798	-2	0.00%
Not Hispanic Black or African Alone	391,693	391,700	7	0.00%
Not Hispanic American Indian and Alaska Native Alone	10,778	10,761	-17	-0.16%
Not Hispanic Asian alone	347,495	347,496	1	0.00%
Not Hispanic Native Hawaiian and Other Pacific Islander Alone	1,467	1,475	8	0.55%
Not Hispanic Some other Race alone	61,547	61,558	11	0.02%
Not Hispanic Two or more races	122,195	122,164	-31	-0.03%
Massachusetts Total	**6,547,629**	**6,547,629**	**0**	**0.00%**

Furthermore, when analyzing the State level data by age cohort and racial and ethnic group, it is evident that there is little to no change is the age breakdown between voting age people and non-voting age

people. The largest shift is 0.17% in the Not Hispanic Native Hawaiian and Other Pacific Islander Alone which is subject to larger percent errors and shifts due to its smaller population (Table 2).

Table 2 Massachusetts Voting Age Population Shift by Race

Race/Ethnicity group	SF1		DP	
	Voting age	Non Voting age	Voting age	Non Voting age
Hispanic or Latino (All Races)	66.40%	33.60%	66.40%	33.60%
Not Hispanic White alone	80.83%	19.17%	80.83%	19.17%
Not Hispanic Black or African Alone	73.66%	26.34%	73.66%	26.34%
Not Hispanic American Indian and Alaska Native Alone	76.26%	23.74%	76.27%	23.73%
Not Hispanic Asian alone	77.44%	22.56%	77.44%	22.56%
Not Hispanic Native Hawaiian and Other Pacific Islander Alone	81.05%	18.95%	80.88%	19.12%
Not Hispanic Some other Race alone	74.53%	25.47%	74.54%	25.46%
Not Hispanic Two or more races	56.94%	43.06%	56.94%	43.06%
Massachusetts Total	**78.33%**	**21.67%**	**78.33%**	**21.67%**

Counties

In the June 8 2021 DAS demonstration data, the total 2010 population at the county level also remains largely unchanged. The total error, which is the difference between the 2010 SF1 Census file and the differential privacy affected data, ranges from a loss of three people to a gain of four people, which is miniscule compared to the total population of the counties (Table 4).

Table 4 Total Population Error and Percent Error, Massachusetts Counties

County	SF1	DP	Difference	Percent Difference
Barnstable County	215,888	215,889	1	0.000%
Berkshire County	131,219	131,221	2	0.002%
Bristol County	548,285	548,285	0	0.000%
Dukes County	16,535	16,537	2	0.012%
Essex County	743,159	743,159	0	0.000%
Franklin County	71,372	71,376	4	0.006%
Hampden County	463,490	463,487	-3	-0.001%
Hampshire County	158,080	158,077	-3	-0.002%
Middlesex County	1,503,085	1,503,084	-1	0.000%
Nantucket County	10,172	10,173	1	0.010%
Norfolk County	670,850	670,848	-2	0.000%
Plymouth County	494,919	494,920	1	0.000%
Suffolk County	722,023	722,022	-1	0.000%
Worcester County	798,552	798,551	-1	0.000%

Table 3 Voting Age Population Shift, Massachusetts Counties

County	SF1		DP	
	Voting age	Non Voting age	Voting age	Non Voting age
Barnstable County	82.75%	17.25%	82.75%	17.25%
Berkshire County	80.47%	19.53%	80.46%	19.54%
Bristol County	77.67%	22.33%	77.68%	22.32%
Dukes County	80.81%	19.19%	80.82%	19.18%
Essex County	76.84%	23.16%	76.84%	23.16%
Franklin County	80.29%	19.71%	80.27%	19.73%
Hampden County	76.29%	23.71%	76.29%	23.71%
Hampshire County	83.07%	16.93%	83.07%	16.93%
Middlesex County	78.68%	21.32%	78.68%	21.32%
Nantucket County	79.28%	20.72%	79.33%	20.67%
Norfolk County	77.32%	22.68%	77.32%	22.68%
Plymouth County	75.86%	24.14%	75.86%	24.14%
Suffolk County	82.51%	17.49%	82.51%	17.49%
Worcester County	76.55%	23.45%	76.56%	23.44%

Once the county populations are broken down by race, however, some counties see a loss of 149 people to a gain of 156 people of a specific racial/ethnic group. In this DAS release, for example, Plymouth County saw the decrease of 149 Hispanic or Latino people of all races, representing a decrease of 0.95% of all Hispanic or Latino people in Plymouth County and Suffolk County saw a gain of 156 Hispanic of Latino People of all races which represents an 0.11% gain.[1] While Plymouth County and Suffolk County saw the largest decrease and increases respectively, it does not necessarily mean that these counties will see the largest increases and decreases in the actual Census 2020 data. We see little to no change

---

[1] As mentioned in the report overview, which counties gain or lose population due to DAS application are randomized and will vary in the actual DAS application to Census 2020 data, and actual numeric variances will also differ.

when analyzing the racial and ethnic makeup of each county, with the largest shifts coming from Nantucket County at a 0.37% increase in the Hispanic or Latino of all races population and a 0.36% decrease in the Not Hispanic White alone population. In similar fashion to the smaller racial and ethnic groups, smaller counties and towns will see larger percent errors. County level errors and percent errors continue to be insignificant in Massachusetts after breaking the population down by racial and ethnic group as well as age cohort (Table 5).

Table 5 SF1 to DP Racial Shift by Massachusetts County

County	Hispanic or Latino (All Races)	Not Hispanic White alone	Not Hispanic Black or African Alone	Not Hispanic American Indian and Alaska Native Alone	Not Hispanic Asian alone	Not Hispanic Native Hawaiian and Other Pacific Islander Alone	Not Hispanic Some other Race alone	Not Hispanic Two or more races
Barnstable County	0.00%	-0.01%	0.00%	0.00%	0.01%	0.01%	0.00%	0.00%
Berkshire County	0.01%	-0.01%	0.00%	0.00%	0.00%	0.00%	-0.02%	0.01%
Bristol County	-0.01%	0.01%	0.01%	0.00%	0.00%	0.00%	0.00%	0.00%
Dukes County	0.16%	-0.09%	-0.06%	0.01%	-0.04%	-0.03%	0.02%	0.04%
Essex County	0.01%	-0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Franklin County	-0.04%	0.08%	-0.02%	0.01%	0.00%	0.01%	0.00%	-0.03%
Hampden County	0.01%	-0.01%	0.00%	0.00%	-0.01%	0.00%	0.00%	0.01%
Hampshire County	-0.01%	0.03%	0.01%	0.00%	0.00%	0.00%	0.00%	-0.02%
Middlesex County	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Nantucket County	0.37%	-0.36%	-0.02%	0.00%	-0.03%	-0.01%	0.04%	0.01%
Norfolk County	-0.01%	0.00%	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%
Plymouth County	-0.03%	0.02%	0.00%	0.01%	-0.01%	0.00%	0.00%	0.01%
Suffolk County	0.02%	-0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	-0.01%
Worcester County	-0.01%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

The summary metrics published by the Cenus Bureau for the most recent DAS demonstration data also show that county-level errors are miniscule across the U.S. There is only one county in the nation where the absolute percent difference is larger than 2% of the population and there are no counties where the absolute percent error is larger than 5%. The mean absolute error for each size category for counties is below 2 and only counties with a population below 1,000, which there are only 35, see a mean absolute percent error above 0.05% (Table 6).

Table 6 Detailed Summary Metrics County Level Error, Percent Error, and Outliers (U.S. Counties)

Total Population for county size categories - MAE, MAPE, and outliers						
County Level	Count of Units (N)	Mean Absolute Error	Mean Absolute Percent Error (%)	Count of geographies where the absolute percent difference is 2% to 5%	Count of geographies where the absolute percent difference exceeds 5%	Count of incorporated places where the absolute numeric difference exceeds 200
All counties	3,143	1.75	0.02	1	-	-
Counties with total population less than 1,000	35	1.31	0.31	1	-	-
Counties with total population 1,000 to 4,999	268	1.46	0.05	-	-	-
Counties with total population 5,000 to 9,999	395	1.72	0.02	-	-	-
Counties with total population 10,000 to 49,999	1,469	1.78	0.01	-	-	-
Counties with total population 50,000 to 99,999	398	1.72	-	-	-	-
Counties with total population of 100,000 or more	578	1.89	-	-	-	-

However, when the county level data is broken down by racial and ethinc groups at the national level, there is an uptick in mean absolute error and mean absolute percent error (Table 7). The highest mean absolule percent error for Massachusetts counties is 14.8% for the Not Hispanic Native Hawaiian and Other Pacific Islander and no other racial and ethnic group has a mean absolute persent error above 3.5%.

Table 7 Detailed Summary Metrics County Level Error, Percent Error, and Outliers by Race and Ethnicity (U.S. Counties)

Hispanic or Latino Origin by Race Alone for counties - MAE, MAPE, and outliers				
All Counties by Race and Hispanic Origin	Count of Units (N)	Mean Absolute Error	Mean Absolute Percent Error (%)	Count where the absolute percent difference exceeds 10%
Hispanic or Latino				
White alone	3,143	10.93	4.73	351
Black alone	3,143	4.37	46.72	1,671
AIAN alone	3,143	8.22	46.57	1,989
Asian alone	3,143	2.63	65.78	1,849
NHPI alone	3,143	2.19	67.40	1,789
SOR alone	3,143	12.37	10.42	711
Two or more races	3,143	11.90	37.66	1,561
Not Hispanic or Latino				
White alone	3,143	20.74	0.17	1
Black alone	3,143	12.77	11.62	692
AIAN alone	3,143	9.08	11.72	925
Asian alone	3,143	7.84	13.66	936
NHPI alone	3,143	3.74	59.39	1,991
SOR alone	3,143	4.08	43.44	1,746
Two or more races	3,143	13.95	9.63	620

Cities and Towns

At the Minor Civil Division level (cities and towns), total population errors range from a loss of 16 people to a gain of 35 people, while the percent errors can range from -2.67% to 1.65% of an MCD's total population (Table 8). As expected, the mean absolute error increases as the population increases and the mean absolute percent error deacreases as the population increases in size.

Table 8 MCD Total Population Error and Percent Error by MCD Size, Massachusetts

MCD Size	n	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs % Error
MCD Under 5,000	106	-9	10	2.2	-2.67%	1.65%	0.21%
MCD 5,000 to 19,999	152	-13	11	2.9	-0.12%	0.13%	0.03%
MCD 20,000 to 49,999	68	-10	18	5.0	-0.04%	0.07%	0.02%
MCD 50,000+	25	-16	35	7.8	-0.02%	0.04%	0.01%
All MCDs	351	-16	35	3.5	-2.67%	1.65%	0.09%

The same trends are seen when looking at MCDs by size and racial and ethnic group where the mean absolute percent errros tend to be smaller when the population and population of the racial and ethnic groups are larger. Futhermore, mean absolute errors tend to be larger for MCDs with larger populations and for racial and ethnic groups with larger populations in Massachusetts (Tables 9 and 10).

Table 9 Mean Absolute Percent Error by MCD Size and Race, All Ages, Massachusetts

	Mean Absolute Percent Error DP. Vs SF1 By Race, Ethnicity, and MCD size, All Ages				
Race Alone and Ethnicity	MCDs <5,000	MCDs 5,000 to 19,999	MCDs 20,000 to 49,999	MCDs 50,000+	All MCDs
White, NH	0.4%	0.1%	0.1%	0.1%	0.2%
Black, NH	36.4%	4.7%	1.4%	0.5%	14.8%
Asian, NH	39.0%	3.8%	1.5%	0.6%	14.4%
Two or More, NH	38.8%	5.3%	2.8%	1.8%	14.8%
Some Other Race, NH	66.4%	20.0%	6.0%	1.8%	27.3%
AI-AN, NH	67.2%	25.4%	17.6%	8.7%	38.1%
HN-PI, NH	78.5%	74.6%	79.0%	30.3%	83.3%
Hispanic/Latino, Any Race	19.9%	3.6%	1.4%	0.5%	8.4%

(All Ages)

Table 10 Mean Absolute Error by MCD Size and Race, All Ages, Massachusetts

	Mean Absolute Error DP. Vs SF1 By Race, Ethnicity, and MCD size, All Ages				
Race Alone and Ethnicity	MCDs <5,000	MCDs 5,000 to 19,999	MCDs 20,000 to 49,999	MCDs 50,000+	All MCDs
White, NH	5	8	12	25	9
Black, NH	2	4	8	17	5
Asian, NH	3	5	9	20	6
Two or More Races, NH	4	7	12	30	9
Some Other Race, NH	2	3	5	9	3
AI-AN, NH	2	3	6	12	4
HN-PI, NH	1	2	3	5	2
Hispanic/Latino, Any Race	4	8	12	37	10

At the national level, we also see mean absolute errors increasing as the MCD population increase and the mean absolute percent error decrease as the MCD population size increases. According to the U.S. Census Bureau summary metrics, there are no cities and towns above 5,000 people where the absolute percent error exceeds 2% (Table 11).

Table 9 Detailed Summary Metrics Error, Percent Error, and Outliers by MCD Size, U.S. MCDs

	Total Population for Minor Civil Division size categories - MAE, MAPE, and outliers					
Minor Civil Division Level	Count of Units (N)	Mean Absolute Error	Mean Absolute Percent Error (%)	Count of geographies where the absolute percent difference is 2% to 5%	Count of geographies where the absolute percent difference exceeds 5%	Count of incorporated places where the absolute numeric difference exceeds 200
All Minor Civil Divisions	11,862	2.74	0.65	333	133	-
Minor Civil Divisions with total population less than 1,000	4,809	2.12	1.46	328	132	-
Minor Civil Divisions with total population 1,000 to 4,999	4,484	2.65	0.14	5	1	-
Minor Civil Divisions with total population 5,000 to 9,999	1,118	2.78	0.04	-	-	-
Minor Civil Divisions with total population 10,000 to 49,999	1,255	4.43	0.02	-	-	-
Minor Civil Divisions with total population 50,000 to 99,999	144	8.00	0.01	-	-	-
Minor Civil Divisions with total population of 100,000 or more	52	11.35	0.01	-	-	-

In addition to MCD size and racial and ethnic group, we have developed further analysis by age cohort that includes 0-to-17 age population and 18-and-over population. Compared to the total population for all ages, both the mean absolute error and mean absolute percent error for the 0-to-17 age population are higher for all MCD sizes. The range of errors and percent errors have also increased substantially. The mean absolute error for all MCDs has increased nearly four times and the mean absolute percent error has increased nine times the mean absolute percent error for the total population (Table 12).

Table 10 MCD 0-to-17 Age Cohort, Error and Percent Error by MCD Size, Massachusetts

MCD Size	n	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs % Error
MCD Under 5,000	106	-17	15	4.5	-6.97%	15.09%	1.77%
MCD 5,000 to 19,999	152	-33	34	9.7	-1.62%	1.83%	0.42%
MCD 20,000 to 49,999	68	-64	40	16.6	-0.86%	0.80%	0.25%
MCD 50,000+	25	-87	123	41.8	-0.51%	0.87%	0.26%
All MCDs	351	-87	123	11.7	-6.97%	15.09%	0.81%

The mean absolute percent error is larger for nearly all race groups and all MCD sizes for the 0-to-17 age population compared to the total population (Tables 13 and 14). The mean absolute errors are relatively similar even though the 0-to-17 age population makes up about 22% of the total population. There are still similar trends within the 0-to-17 age population where the smaller racial and ethnic groups see larger percent errors and the MCDs with smaller populations also see larger percent errors.

Table 13 Mean Absolute Percent Error by Race and MCD Size, 0-to-17 Age Cohort, Massachusetts

Table 14 Mean Absolute Error by Race and MCD Size, 0-to-17 Age Cohort, Massachusetts

	Race Alone and Ethnicity	MCDs <5,000	MCDs 5,000 to 19,999	MCDs 20,000 to 49,999	MCDs 50,000+	All MCDs
	Mean Absolute Percent Error DP. Vs SF1 By Race, Ethnicity, and MCD size, Age 0-17					
Age 0-17	White, NH	1.3%	0.3%	0.2%	0.2%	0.6%
	Black, NH	64.6%	20.9%	6.8%	0.9%	30.4%
	Asian, NH	84.0%	12.8%	4.0%	2.0%	31.4%
	Two or More, NH	61.7%	11.2%	5.0%	4.0%	25.7%
	Some Other Race, NH	74.6%	57.1%	18.6%	6.0%	48.8%
	AI-AN, NH	76.5%	75.7%	66.1%	37.0%	77.8%
	HN-PI, NH	75.0%	86.3%	71.2%	95.1%	100.5%
	Hispanic/Latino, Any Race	50.4%	8.9%	3.6%	1.5%	20.8%

Race Alone and Ethnicity	MCDs <5,000	MCDs 5,000 to 19,999	MCDs 20,000 to 49,999	MCDs 50,000+	All MCDs
Mean Absolute Error DP. Vs SF1 By Race, Ethnicity, and MCD size, Age 0-17					
White, NH	3	6	9	14	6
Black, NH	2	4	6	12	4
Asian, NH	2	4	7	11	4
Two or More Races, NH	3	6	11	24	8
Some Other Race, NH	2	3	5	9	3
AI-AN, NH	2	3	4	10	3
HN-PI, NH	1	1	1	2	1
Hispanic/Latino, Any Race	4	7	12	31	8

The 18-and-over population makes up a little more than 78% of the Massachusetts population in the 2010 SF1 data. Compared to the total population, the 18-and-over age cohort has a mean absolute error that is four times as high, 3.5 compared to 11.7, and a mean absolute error that is slightly more than double, 0.09% compared to 0.20%, for all MCDs. The range of errors for all MCDs is also much larger going from -16 to 35 for the total population to -129 to 73 for the 18-and-over age cohort. Compared to the 0-to-17 age cohort, the mean absolute error are very similar event though the 18-and-over population is nearly four times the size of the 0-to-17 age population. The mean absolute percent error for the 18-and-over age cohort is much lower than the 0-to-17 age group for all MCD sizes.

Table 11 MCD 18-and-over Age Cohort, Error and Percent Error by MCD Size, Massachusetts

MCD Size	n	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs % Error
MCD Under 5,000	106	-17	16	4.1	-4.69%	2.00%	0.40%
MCD 5,000 to 19,999	152	-36	39	10.0	-0.43%	0.55%	0.13%
MCD 20,000 to 49,999	68	-44	66	16.6	-0.15%	0.24%	0.07%
MCD 50,000+	25	-129	73	41.0	-0.11%	0.12%	0.06%
All MCDs	351	-129	73	11.7	-4.69%	2.00%	0.20%

After breaking down the 18-and-over age cohort by racial and ethnic group, we see that the mean absolute percent errors are slightly larger than the total population by racial and ethnic group but significantly smaller than the 0-to-17 age cohort. Furthermore, the mean absolute errors are similar to both the total population and 0-to-17 age group even though the 18-and-over group is nearly four times as large as the 0-to-17 cohort. When comparing the different racial and ethnic groups, it seems that Two or More Races, not Hispanic, and Hispanic or Latino of any race tend to have the highest mean absolute errors across all age cohorts even though their populations are far smaller than the White alone, not Hispanic racial and ethnic group. Like the other age cohorts, the 18-and-over cohort still sees the smaller

racial and ethnic groups, as well as the smaller sized MCDs have larger mean absolute percent errors compared to the more popular and larger sized racial and ethnic groups and MCDs (Tables 16 and 17).

Table 16 Mean Absolute Percent Error by Race and MCD Size, 18-and-over Age Cohort, Massachusetts

Table 17 Mean Absolute Error by Race and MCD Size, 18-and-over Age Cohort, Massachusetts

	Race Alone and Ethnicity	MCDs <5,000	MCDs 5,000 to 19,999	MCDs 20,000 to 49,999	MCDs 50,000+	All MCDs
	Mean Absolute Percent Error DP. Vs SF1 By Race, Ethnicity, and MCD size, Age 18+					
Age 18+	White, NH	0.4%	0.1%	0.0%	0.0%	0.2%
	Black, NH	42.3%	7.7%	1.8%	0.6%	17.1%
	Asian, NH	39.5%	5.8%	1.9%	0.8%	15.3%
	Two or More, NH	42.5%	11.6%	5.3%	3.0%	20.0%
	Some Other Race, NH	57.5%	35.4%	9.8%	1.4%	34.2%
	AI-AN, NH	70.9%	34.8%	20.9%	8.0%	43.5%
	HN-PI, NH	87.0%	73.9%	82.3%	26.4%	84.4%
	Hispanic/Latino, Any Race	26.8%	6.1%	2.0%	0.8%	11.7%

Race Alone and Ethnicity	MCDs <5,000	MCDs 5,000 to 19,999	MCDs 20,000 to 49,999	MCDs 50,000+	All MCDs
Mean Absolute Error DP. Vs SF1 By Race, Ethnicity, and MCD size, Age 18+					
White, NH	4	6	9	20	7
Black, NH	2	4	7	15	5
Asian, NH	3	5	8	17	6
Two or More Races, NH	3	7	12	28	8
Some Other Race, NH	2	3	5	8	3
AI-AN, NH	2	3	5	8	3
HN-PI, NH	1	1	3	4	1
Hispanic/Latino, Any Race	3	8	13	31	9

Census Tracts

At the census tract level, the error range is relatively small for the total population. Errors range from a minimum of 8 to a maximum of 9 with a mean absolute error of 1.9 (Table 18). The range of percent errors is also reasonably tight, ranging from -16.7% to 11.1%, with a mean absolute percent error of 0.1%. The average size of all 1,463 non-zero population tracts in Massachusetts is 4,451 people, with populations ranging from 6 to 12,079.

Table 12 Tract Level Total Population Error and Percent Error, Massachusetts

Difference in Tract-Level Population by Race, DP vs. SF1, All Ages							
Race/Ethnicity	# Tracts	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs. % Error
All	1,463	-8	9	1.9	-16.7%	11.1%	0.1%

When comparing the mean absolute error and range across all geography, tract level data performs well with the second smallest range of errors behind the county level data. The mean absolute percent error for all tracts is on par with the MCD-level mean absolute percent error and far less than the block-level mean absolute percent error (Table 19). The range of percent errors is a bit wider than the MCD level data. However, this larger gap in percent errors can be expected due to the smaller size of some of the tracts in Massachusetts which can be as small as 6 people in 2010 SF1 data set.
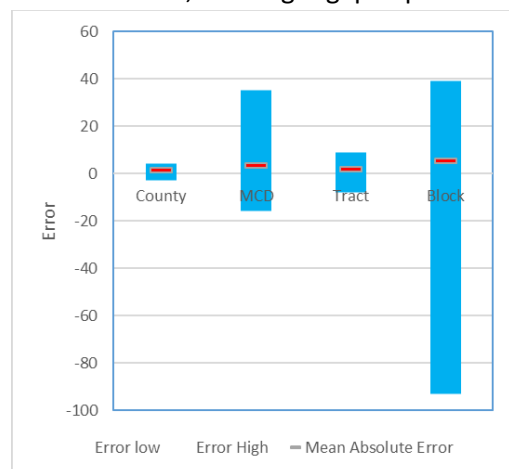

Table 19 and Figure 1 (left): Range of Errors, Mean Absolute Error, and Mean Absolute Percent Error by Geography, Total Population, Massachusetts

Geography	Percent Error Low	Percent Error High	Percent Error Range	Mean Absolute Percent Error
County	0.0%	0.0%	0.0%	0.0%
MCD	-2.7%	1.7%	4.3%	0.1%
Tract	-16.7%	11.1%	27.8%	0.1%
Block	-100.0%	3100.0%	3200.0%	32.8%

According to the Census Bureaus detailed summary metrics for the latest DAS production settings, Massachusetts and the United States both see a mean absolute error of 1.9 people for all tracts. Only 153 tracts in the entire country see an absolute percent difference in total population that exceeds 5% and there are no tracts with an error of 200 or more.

Table 13 Detailed Summary Metrics Tract Level Total Population Error and Percent Error, U.S. Tracts

Total Population for tracts - MAE and outliers					
	Count of Units (N)	Mean Absolute Error	Count of geographies where the absolute percent difference is 2% to 5%	Count of geographies where the absolute percent difference exceeds 5%	Count of where the absolute numeric difference exceeds 200
All tracts	73,057	1.93	45	153	0

When analyzing the Massachusetts tract-level data by race, the range of errors stretches to a low of -38 to a high of 49 with mean absolute errors for the eight racial and ethnic groups ranging from 1.3 to 5.8. In addition, the percent error range increases across the eight racial and ethnic groups to a minimum -100% (which means the racial and ethnic group was zeroed out, a common occurrence for small racial and ethnic populations at the tract and block level) to a maximum of 1000%. The mean absolute percent error range across all race groups, for all ages, ranges from 0.4% to 72.1%, with the maximum mean absolute percent error coming from the smaller Hawaiian Native or Pacific Islander alone, not Hispanic or Latino racial and ethnic group. A detailed table of error metrics for all eight racial and ethnic groups as well as the total population can be found in the appendix to this report (Table 26).

Once again, we can see that the tract level data by race and ethnicity stacks up well compared to other geographies. The large majority of mean absolute errors and mean absolute percent errors for the tract level data by racial and ethnic group are lower than the MCD level data. Furthermore, the tract level mean absolute percent errors are all lower than the block level mean absolute percent errors.

Figures 2 and 3. Mean Absolute Error and Mean Absolute Percent Error across Geographic Levels by Race and Ethnicity, All Ages, Massachusetts
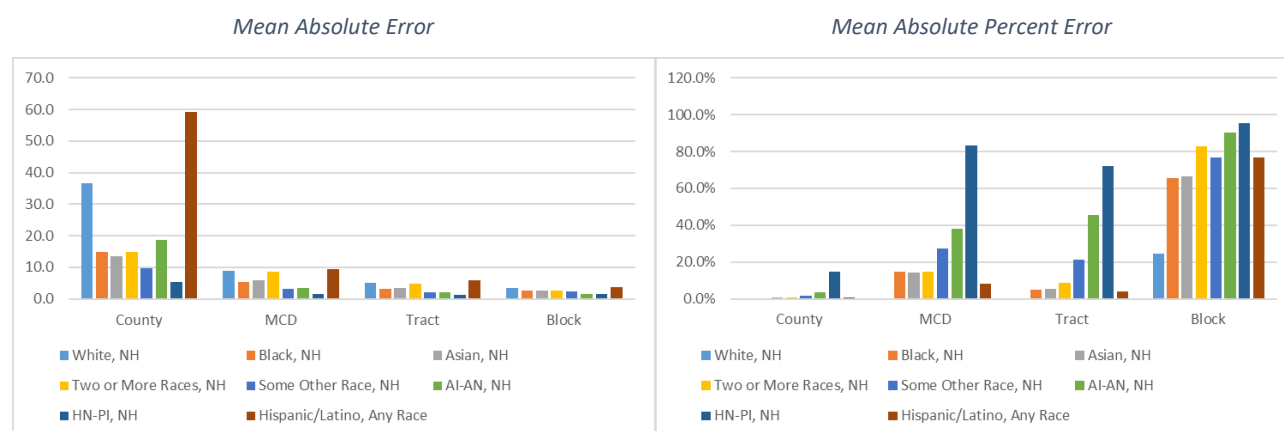
After breaking down the tract level data by age cohort, we see that for the total population for the 0-to-17 age group that the mean absolute error increases to 7.7 people, compared to 1.9 for all ages. In addition, we see that the range of errors is almost identical to the range of errors for all ages, and is now

-37 to 44 for the 0-to-17 age group. We do see the range of percent error spike to -80% to 650%. However, it should be noted that there are various tracts with small populations, after segmenting the data into the 0-to-17 age cohort, which can be adversely affected by differential privacy due to the difficulty of keeping a smaller population in a specific geography private. After filtering the tract level data and removing any tracts smaller than 50 people, the maximum percent error was reduced to 47% and the minimum percent error to -11%. The mean absolute percent error for all non-zero population tracts for the 0-to-17 age cohort has also increased compared to the mean absolute percent error for all ages from 0.1% now to 3.0% (Table 20).

Table 14 Tract Level 0-to-17 Age Cohort Error and Percent Error, Massachusetts

Difference in Tract-Level Population by Race, DP vs. SF1, 0-17							
Race/Ethnicity	# Tracts	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs. % Error
All	1,459	-37	44	7.7	-80.0%	650.0%	3.0%

When analyzing the 0-to-17 age cohort by race and ethnicity, we see that the range of errors across all racial and ethnic groups is -24 to 37, which both happen to fall in the Hispanic or Latino of any race category. The mean absolute errors range from 1.3 to 5.5 across all racial and ethnic groups. While the mean absolute errors tend to be smaller for the 0-to-17 age population, due to smaller population than all ages, the mean absolute percent errors have increased for 0-to-17 with the range now being 2.3% to 90.6% across all race groups. The range of percent errors is -100% to 1300%, with the same caveat as all ages tracts that there are very small tract populations that can show dramatic increases in percent error. The detailed error metrics table for all racial and ethnic groups analyzed for the 0-to-17 age cohort can be found in the appendix as well (Table 27).

The last age cohort analyzed for tract level data was the 18 years and over age group (Table 21). The mean absolute error for the total population of the 18-and-over age group was 7.7, which is identical to the 0-to-17 age group. The range of errors is also similar to the 0-to-17 age group as well, and is now -41 to 35 compared to -37 to 44 for 0-to-17. However, the mean absolute error for the 18 years and older group is much smaller and closer to the mean absolute error for all ages. The mean absolute error for 18 years and older is 0.3%, compared to all ages which is 0.1% and the 0-to-17 age group mean absolute error is 3%. The range of percent errors is also much tighter for the 18 and older age cohort with the minimum error being -25% and the maximum percent error being 1.9%. The range of the 18 and older age cohort is 26.9%, while the range for all ages is 27.8%.

Table 15 Tract Level 18-and-over Age Cohort Error and Percent Error, Massachusetts

Difference in Tract-Level Population by Race, DP vs. SF1, 18+							
Race/Ethnicity	# Tracts	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs. % Error
All	1,463	-41	35	7.7	-25.0%	1.9%	0.3%

For the 18 and older age cohort by racial and ethnic group, the range of errors is -37 to 34, where both the maximum and minimum are once again found in the Hispanic or Latino of all races category. The

mean absolute error across all eight racial and ethnic groups ranges from 1.3 to 5.9 which is very similar to both the 0-to-17 age cohort and the range for all ages at the tract level. The mean absolute percent error ranges from 0.4% to 79% with percent errors ranging from a minimum of -100% to a maximum of 800%, which is seen in the American Indian or Alaska Native, not-Hispanic racial and ethnic group. Once again, it is important to note that smaller racial and ethnic groups, like the American Indian and Alaska Native, not-Hispanic group, can see higher percent errors due to their smaller population numbers especially when the geographic levels are scaled down to census tracts and blocks. Like the 0-to-17 and all-ages cohorts, the detailed breakdown of the error metrics for the eight racial and ethnic groups aged 18 and over can be found in the appendix to this report (Table 28).

The Census Bureau published tract level racial data for the population 18 years and over by Hispanic or Latino origin for all tracts in the United States, of which there are 73,057 (Table 22). The mean absolute error across all race groups, for both Hispanic or Latino and Not Hispanic or Latino, ranges from 0.41 to 4.73. Massachusetts sees a similar range in mean absolute error in its 18-and-over population. Furthermore, the Census Bureau provided one outlier percent error metric for the tract level data by race and ethnicity which is the count of tracts where the absolute percent difference exceeds 10%. Four of the fourteen race and ethnicity categories that the Census Bureau analyzed see over 60% of tracts have an absolute percent difference that exceeds 10%. The total population of all ages tract level data looks quite promising after analyzing the 2010 SF1 data against the most recent June 6th, 2021 DAS PPMF files. However, after breaking the data down by race and ethnicity and age, the errors and percent errors increase which could cause some concern for stakeholders using census tract level data.

Table 16 Detailed Summary Metrics Tract Level Error, and Outliers by Race and Ethnicity, U.S. Tracts

Hispanic or Latino Origin by Race Alone for the Population 18 Years and Over for tracts - MAE and outliers			
Race Group and Hispanic Origin	Count of Units (N)	Mean Absolute Error	Count where the absolute percent difference exceeds 10%
Hispanic or Latino			
White alone	73,057	3.04	14,040
Black alone	73,057	1.64	42,840
AIAN alone	73,057	1.62	44,232
Asian alone	73,057	0.87	33,664
NHPI alone	73,057	0.41	18,020
SOR alone	73,057	3.15	23,631
Two or more races	73,057	3.49	49,994
Not Hispanic or Latino			
White alone	73,057	3.93	1,203
Black alone	73,057	3.13	15,095
AIAN alone	73,057	2.2	48,761
Asian alone	73,057	2.79	23,309
NHPI alone	73,057	1.08	37,340
SOR alone	73,057	1.52	48,071
Two or more races	73,057	4.73	40,283

Census Blocks

Census block-level data is the smallest geographic level of data that the Census Bureau releases and is therefore subject to the most pronounced effects of differential privacy as evident from the analysis to follow.

In the 2010 SF1 data file, there are 157,508 blocks in Massachusetts with the average total population per block coming in at 41.6 people. However, only 96,334 blocks, or about 61% of all, contain population, referred to as "non-zero" blocks. In the non-zero blocks, the total averages 68 people. The largest block by total population contains 4,025 people. The mean absolute error for the block total population is 5.6 people with a range of errors that goes from -93 to 39. The mean absolute percent error is 32.8% with a drastic range of percent errors spanning from -100%, meaning the population was zeroed out for that specific block, to 3100% percent (Table 23). It should be noted that a minimum percent error of -100% could be as simple as a block with two people loses its two people and is therefore zeroed out. A percent error of 3100% people could potentially be that a block of two people gains 62 people and now has 64 people in the block, therefore showing a 3100% increase. Due to the extremely small populations in some Massachusetts blocks, percent error and mean absolute percent error can see significant increases and swings, especially when the data gets further segmented into age cohorts and racial and ethnic groups where populations, and therefore denominators in the percent error calculations, are much smaller.

Table 17 Block Level Total Population Error and Percent Error, Massachusetts

Difference in Block-Level Population by Race, DP vs. SF1, All Ages							
Race/Ethnicity	# Blocks	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs. % Error
All	96,334	-93	39	5.6	-100.0%	3100.0%	32.8%

After breaking down the block level data by race and ethnicity for all ages, the error range is now -63 to 31 with both the maximum and minimum coming from the Hispanic or Latino of all races category. The mean absolute errors range from 1.5 to 3.6. The percent errors range from -100%, which is the minimum for every racial and ethnic category, to 1900%. The mean absolute percent errors range from 24.5% to 95.4%. The detailed race and ethnicity table with all error metrics can be found in the appendix to this report (Table 29).

Like all other geographic levels, the block level data was available for analysis by two age cohorts, 0 to 17 and 18 and older. When looking at the 0-to-17 age cohort, we see that there are 83,549 blocks with population, which represents about 53% of total blocks in the 2010 census SF1 data file. The error range for the 0-to-17 age group is smaller than the total population which is to be expected since, as previously stated the 0-to-17 age group is only 21.7% of the total population. Errors range from -52 to 24 with a mean absolute error of 3.4 for all 0-to-17 age people. The percent error range for the non-voting age population is also smaller than the total population and is -100% to 2100% with a mean absolute error of 51.2% (Table 24).

Table 18 Block Level 0-to-17 Age Cohort Error and Percent Error, Massachusetts

Difference in Block-Level Population by Race, DP vs. SF1, 0-17							
Race/Ethnicity	# Blocks	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs. % Error
All	83,549	-52	24	3.4	-100.0%	2000.0%	51.2%

The 0-to-17 age cohort can also be broken down into racial and ethnic groups. For the 0-to-17 age group, the range of errors across all of the racial and ethnic groups is -35 to 19, where both the maximum and minimum errors are found in the Hispanic or Latino of any race category. The range of mean absolute errors is relatively tight and goes from 1.4 to 2.5. The range of percent errors is -100% to 1500% and once again, both are found in the Hispanic or Latino of any race category. The mean absolute percent error ranges from 46.9% to 97.5%, where the maximum is similar to the maximum for all ages but the minimum is nearly double that of all ages. The detailed error metrics table for all racial and ethnic categories analyzed can be found in the appendix (Table 30).

The 18-and-over age cohort seems to perform better than the total population for all ages across all error metrics (Table 25). The 18-and-over age group has a smaller range of errors, -64 to 38, as well as a smaller mean absolute error at 3.9 compared to the all ages mean absolute error of 5.6. The percent error range for the 18-and-over age cohort is -100% to 1900% and the mean absolute percent error is 26.9% compared to 32.8% for the all-ages group. This is the only age cohort across all geographies that we have analyzed that has a mean absolute percent error for the total population that is less than the mean absolute percent error for all ages. One potential reason for the smaller mean absolute percent error is the redistricting and voting rights act accuracy targets that the Census Bureau put in place to ensure that the census files were completely usable for these purposes. Redistricting legislative districts is done at the block level for the voting age population which is the 18-and-over age cohort. It is possible that the Census Bureau used more of the privacy loss budget on the 18-and-over age cohort for the block level data because of its specific use in redistricting.

Table 19 Block Level 18-and-over Age Cohort Error and Percent Error, Massachusetts

Difference in Block-Level Population by Race, DP vs. SF1, 18+							
Race/Ethnicity	# Blocks	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs. % Error
All	96,327	-64	38	3.9	-100.0%	1900.0%	26.9%

Finally, the last age and race and ethnicity analysis that was performed was on the 18-and-over age cohort by the eight racial and ethnic groups that have been previously used (appendix Table 31). The range of errors for the 18-and-over age group is smaller than both the 0-to-17 age group and the all ages range. The range of errors for 18-and-over goes from -31 to 22, which are both found in the Hispanic and Latino of any race category. The mean absolute errors range are tight, ranging from 1.3 to 2.6. The percent error range for the 18-and-over age cohort is -100% to 1400% with a mean absolute percent error range of 22.3% to 95.9%. The percent error maximum for every racial and ethnic group for the 18-and-over age cohort is either the same or lower than it is for the corresponding racial and ethnic group

for population of all ages. The mean absolute percent errors are also very similar for the 18-and-over age cohort when compared to the population of all ages.

For more information on differential privacy and the applied Census Bureau Disclosure Avoidance System (DAS) please visit the Census Bureau's webpage at https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance.html

Appendix to UMDI Census 2020 Differential Privacy Brief, October 26, 2021

Table 20 Tract Level Total Population Error and Percent Error by Race and Ethnicity, Massachusetts

Difference in Tract-Level Population by Race, DP vs. SF1, All Ages							
Race/Ethnicity	# Tracts	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs. % Error
All	1,463	-8	9	1.9	-16.7%	11.1%	0.1%
White Alone, NH	1,463	-38	29	5.0	-100.0%	33.3%	0.4%
Black or African American Alone, NH	1,461	-26	27	3.1	-100.0%	200.0%	5.0%
Asian Alone, NH	1,460	-22	39	3.4	-100.0%	300.0%	5.5%
Two or More Races Alone, NH	1,459	-23	29	4.8	-66.7%	450.0%	8.7%
Some Other Race Alone, NH	1,406	-17	12	2.0	-100.0%	1000.0%	21.2%
American Indian or Alaska Native Alone, NH	1,372	-12	12	2.1	-100.0%	500.0%	45.3%
Hawaiian Native or Pacific Islander Alone, NH	636	-7	6	1.3	-100.0%	500.0%	72.1%
Hispanic or Latino, Any Race	1,461	-29	49	5.8	-100.0%	250.0%	4.1%

Table 21 Tract Level 0-to-17 Age Cohort Error and Percent Error by Race and Ethnicity, Massachusetts

Difference in Tract-Level Population by Race, DP vs. SF1, 0-17							
Race/Ethnicity	# Tracts	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs. % Error
All	1,459	-37	44	7.7	-80.0%	650.0%	3.0%
White Alone, NH	1,456	-17	15	3.7	-100.0%	100.0%	2.3%
Black or African American Alone, NH	1,425	-15	18	2.9	-100.0%	500.0%	22.5%
Asian Alone, NH	1,428	-12	18	2.8	-100.0%	600.0%	21.0%
Two or More Races Alone, NH	1,447	-21	27	5.0	-100.0%	1300.0%	22.8%
Some Other Race Alone, NH	1,234	-11	12	2.3	-100.0%	1200.0%	50.0%
American Indian or Alaska Native Alone, NH	850	-6	11	1.9	-100.0%	900.0%	87.4%
Hawaiian Native or Pacific Islander Alone, NH	171	-4	5	1.3	-100.0%	300.0%	90.6%
Hispanic or Latino, Any Race	1,452	-24	37	5.5	-100.0%	900.0%	15.3%

Table 22 Tract Level 18-and-over Age Cohort Error and Percent Error by Race and Ethnicity, Massachusetts

Difference in Tract-Level Population by Race, DP vs. SF1, 18+							
Race/Ethnicity	# Tracts	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs. % Error
All	1,463	-41	35	7.7	-25.0%	1.9%	0.3%
White Alone, NH	1,463	-23	28	4.1	-100.0%	14.3%	0.4%
Black or African American Alone, NH	1,461	-20	19	3.0	-100.0%	200.0%	7.1%
Asian Alone, NH	1,459	-17	27	3.4	-100.0%	120.0%	7.4%
Two or More Races Alone, NH	1,459	-32	20	5.3	-69.2%	500.0%	16.8%
Some Other Race Alone, NH	1,387	-11	11	2.4	-100.0%	600.0%	32.7%
American Indian or Alaska Native Alone, NH	1,345	-13	11	2.2	-100.0%	800.0%	58.8%
Hawaiian Native or Pacific Islander Alone, NH	614	-6	6	1.3	-100.0%	400.0%	79.0%
Hispanic or Latino, Any Race	1,461	-37	34	5.9	-100.0%	200.0%	6.1%

Table 23 Block Level Total Population Error and Percent Error by Race and Ethnicity, Massachusetts

Difference in Block-Level Population by Race, DP vs. SF1, All Ages							
Race/Ethnicity	# Blocks	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs. % Error
All	96,334	-93	39	5.6	-100.0%	3100.0%	32.8%
White Alone, NH	94,837	-26	24	3.4	-100.0%	1600.0%	24.5%
Black or African American Alone, NH	31,625	-19	19	2.6	-100.0%	1200.0%	65.7%
Asian Alone, NH	33,656	-23	22	2.7	-100.0%	1200.0%	66.7%
Two or More Races Alone, NH	31,146	-44	17	2.5	-100.0%	1300.0%	82.7%
Some Other Race Alone, NH	12,663	-18	16	2.2	-100.0%	1100.0%	77.0%
American Indian or Alaska Native Alone, NH	5,962	-11	7	1.5	-100.0%	700.0%	90.3%
Hawaiian Native or Pacific Islander Alone, NH	932	-8	6	1.5	-100.0%	200.0%	95.4%
Hispanic or Latino, Any Race	44,930	-63	31	3.6	-100.0%	1900.0%	76.6%

Table 24 Block Level 0-to-17 Age Cohort Error and Percent Error by Race and Ethnicity, Massachusetts

Difference in Block-Level Population by Race, DP vs. SF1, 0-17							
Race/Ethnicity	# Blocks	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs. % Error
All	83,549	-52	24	3.4	-100.0%	2000.0%	51.2%
White Alone, NH	78,854	-14	15	2.3	-100.0%	1000.0%	46.9%
Black or African American Alone, NH	17,558	-13	15	2.0	-100.0%	1000.0%	71.7%
Asian Alone, NH	19,269	-11	12	1.9	-100.0%	900.0%	75.5%
Two or More Races Alone, NH	20,187	-20	13	1.9	-100.0%	1100.0%	84.4%
Some Other Race Alone, NH	6,104	-10	10	1.7	-100.0%	800.0%	84.5%
American Indian or Alaska Native Alone, NH	1,527	-7	6	1.4	-100.0%	300.0%	91.6%
Hawaiian Native or Pacific Islander Alone, NH	181	-4	4	1.5	-100.0%	100.0%	97.5%
Hispanic or Latino, Any Race	30,095	-35	19	2.5	-100.0%	1500.0%	76.3%

Table 25 Block Level 18-and-over Age Cohort Error and Percent Error by Race and Ethnicity, Massachusetts

Difference in Block-Level Population by Race, DP vs. SF1, 18+							
Race/Ethnicity	# Blocks	Error Min.	Error Max.	Mean Abs. Error	% Error Min.	% Error Max.	Mean Abs. % Error
All	96,327	-64	38	3.9	-100.0%	1900.0%	26.9%
White Alone, NH	94,781	-23	17	2.5	-100.0%	1400.0%	22.3%
Black or African American Alone, NH	30,646	-15	16	2.1	-100.0%	1000.0%	66.2%
Asian Alone, NH	32,415	-16	16	2.1	-100.0%	1000.0%	66.2%
Two or More Races Alone, NH	24,719	-24	17	1.9	-100.0%	1200.0%	84.2%
Some Other Race Alone, NH	11,567	-11	11	1.8	-100.0%	800.0%	78.1%
American Indian or Alaska Native Alone, NH	5,466	-10	6	1.3	-100.0%	600.0%	92.3%
Hawaiian Native or Pacific Islander Alone, NH	879	-6	6	1.3	-100.0%	200.0%	95.9%
Hispanic or Latino, Any Race	42,399	-31	22	2.6	-100.0%	1400.0%	74.8%