

# Census Demographic and Housing Characteristics File (DHC) Demonstration Product Webinar

Jason Devine  
Population Division

Alexandra Krause  
Population Division

**March 22, 2022**

# Background

- Census Bureau has a long history of collecting, disseminating, and protecting data
  - Legal obligation to protect data under Title 13
  - With more detailed data being published at lower levels of geography, Increased availability of external data, and advances in computer science the risk of disclosure has increased
  - To mitigate, implement disclosure avoidance methods based on differential privacy
  - With the new approach noise is added to tabulations and less detail is released
- Challenge: produce accurate and detailed data while protecting individual responses

# Background (cont.)

- Demonstration products allow the public to see progress and the level of noise that is being added through comparisons between the published 2010 Census data and 2010 Census data protected using the new disclosure avoidance methodology
  - 5 demonstration products and a production settings product were provided leading up to the release of the redistricting data (2 of these included limited DHC content, the other 3 only included redistricting content)
  - Two DHC specific demonstration products (including the 2022-03-16 demonstration product) and a production settings product are planned for the DHC

## Previous Demonstration Products

### ^ About DHC Development

The census data product formerly known as "Summary File 1" is now the Demographic and Housing Characteristics File (DHC). Please follow our [newsletter](#) for updates and to learn how to contribute to its development. Review the [2020 Census Data Product Planning Crosswalk](#) to better understand the proposed changes to this data product.

#### ∨ DAS Development Update 2022-03-16

[Top of Section](#)

## Detailed Demographic and Housing Characteristics File (Detailed DHC)

### ^ About Detailed DHC Development

Improvements in the design, processing and coding of the 2020 Census allow the release of data for almost five times as many detailed race and ethnic groups than were possible in 2010. Development of disclosure protections to accommodate this improvement are underway. See the "Detailed DHC" tabs in the [2020 Census Data Product Planning Crosswalk](#).

[Top of Section](#)

## 2020 Census Redistricting Data (P.L. 94-171)

[EXPAND ALL](#) | [COLLAPSE ALL](#)

#### ∨ About 2020 Census Redistricting Data (P.L. 94-171) Summary File Development

#### ∨ Production Settings: 2021-06-08

#### ∨ DAS Development Update 2021-04-28

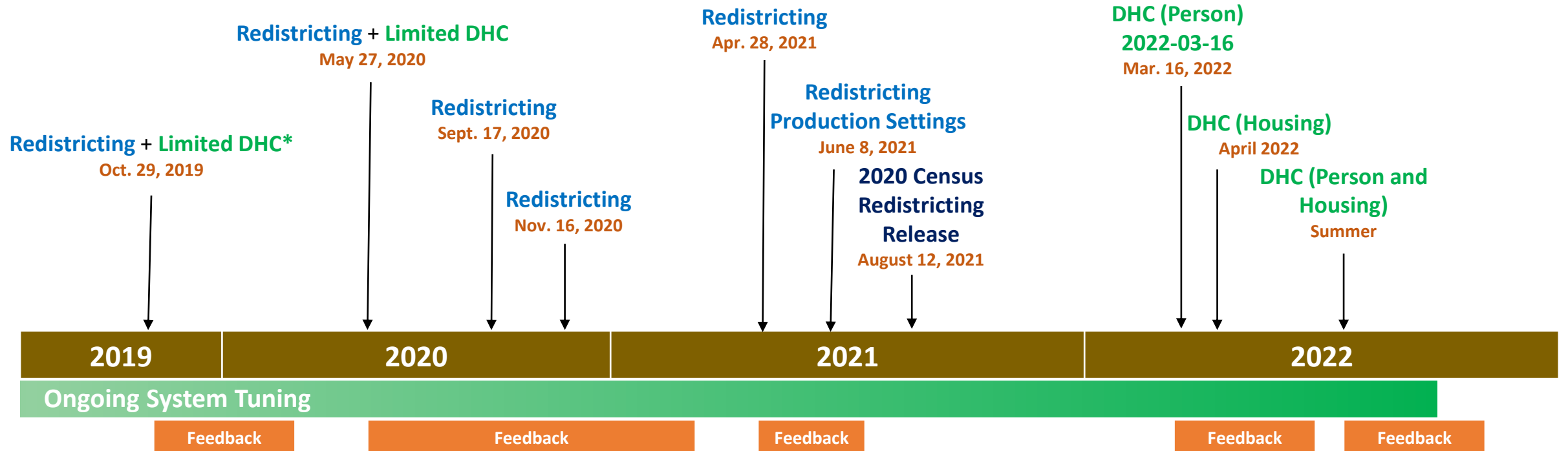
#### ∨ DAS Development Update 2020-11-16

#### ∨ DAS Development Update 2020-09-17

#### ∨ DAS Development Update 2020-05-27

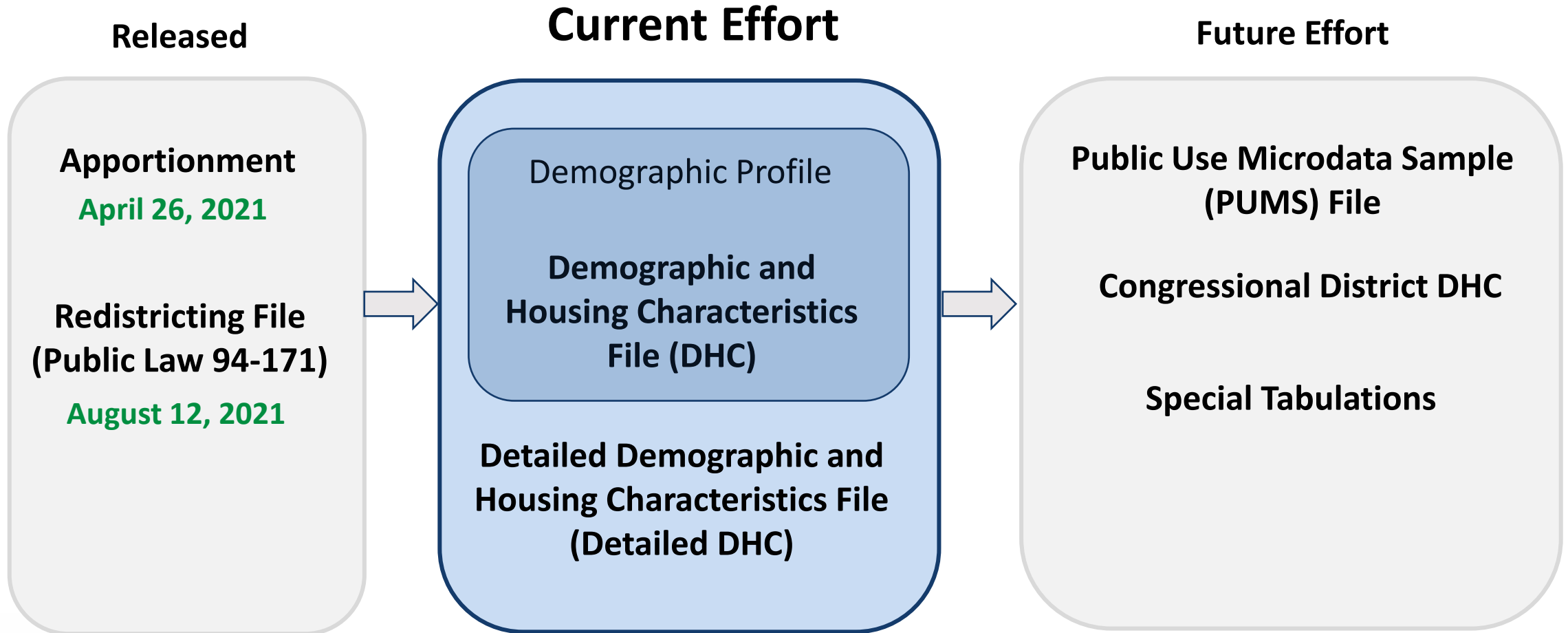
#### ∨ 2010 Demonstration Data Products Baseline 2019-10-29

# Demonstration Product Timeline



# Understanding the DHC Demonstration Product

## 2020 Census Data Products



# Understanding the DHC Demonstration Product

## Determining the Contents

- The 2010 Census Summary File 1 (SF1) was the starting point for the design of the DHC with the DHC including many of the tables included in SF1
- The initial proposal for the design and contents of the DHC was developed and shared in 2019 based on subject matter expert knowledge and feedback from a Federal Register Notice
- The contents of the DHC were reviewed and additional tables are being considered based on the feedback received documenting uses of the data
- The cost for privacy and the accuracy of the new tables must be considered during the tuning of the disclosure avoidance system (DAS) before they are added to the DHC
- Any additions or changes to the contents of the DHC will be reflected in the second DHC demonstration product

# Understanding the DHC Demonstration Product

## Topics Included in the 2020 Census

- Relationship to Householder
- Sex
- Age
- Hispanic or Latino origin
- Race
- Housing Tenure (owned, rented, occupied without payment)
- Vacancy Status

# Understanding the DHC Demonstration Product

## Topics Included in the DHC

### Counts of People by:

- Sex by single year-of-age
- Hispanic or Latino origin by race by sex by age
- GQ population by sex by age
- Household Type by Relationship to householder

### Counts of Households/Families by:

- Multigenerational households
- Hispanic or Latino origin of householder by race of householder
- Family type by presence of children
- Presence of people of different ages
- Tenure by household size
- Tenure by household type by age of householder
- Vacancy Status



# Understanding the DHC Demonstration Product

## Protecting from Disclosure

- Data from the Census Edited File (CEF) are coded to create the variables needed to generate the tables in the DHC and then put into **Person** and **Housing Unit (Unit)** microdata files
- The variables included in the **Person** and **Unit** microdata files are designed to only provide the information necessary for the DHC tables
- Variables in the DHC **Person** and **Unit** microdata files are independently protected by our new disclosure avoidance procedures
- Redistricting data (released August 2021) is used as an input allowing the DHC tables to match the Redistricting tables exactly
- Tables are produced for dissemination by tabulating the protected DHC **Person** and **Unit** microdata files

# Understanding the DHC Demonstration Product

## Refining Our Disclosure Avoidance Methodology

- Experimental runs from our Disclosure Avoidance System (DAS) using 2010 Census data allow us to assess the resulting privacy and accuracy
- A re-identification study is being used to assess the risk of re-identification for a specific attack
- Leading experimental run along with the results for the accuracy analysis and re-identification study are discussed with DSEP\* to obtain approval to release as the demonstration product
- The demonstration product is made available to the public for their analysis with their comments helping guide work on improvements to the DAS

\* Data Stewardship Executive Policy Committee

# Understanding the DHC

## The Demonstration Product

- Two rounds of DHC demonstration products are planned with this webinar focusing on the first round
- The 2010 Demonstration Data Product for the DHC (2022-03-16) for tables sourced from the **Person** File in the form of Summary Files was released for public analysis on March 16
- Accuracy metrics for the tables sourced from both the **Person** and **Unit** file demonstration products were also released on March 16
- A DHC demonstration product for tables sourced from the **Unit** file is being prepared for release
- The public will have 30 days to review and comment on the **Person** file and 30 days to review and comment on the **Unit** file demonstration products
- This is the most extensive set of demonstration data to be released using the new disclosure avoidance methods (113 Person File tables, 121 Unit File tables, 50+ different geographies)

# Understanding the DHC Demonstration Product

## Census Table Names

Table names indicate the lowest level of geography at which the table is proposed to be available.

**P** = Population table - Census Block Level

**PCT** = Population table - Census Tract Level

**PCO** = Population table - County Level

**PST** = Population table - State Level

**H** = Housing table – Census Block Level

**HCT** = Housing table – Census Tract Level

# Understanding the DHC Demonstration Product

## Table Iterations by Race and Hispanic Origin

Many tables are repeated (iterated) by race and Hispanic origin. Some iterations are for categories A through I:

A: White alone

B: Black or African American alone

C: American Indian and Alaska Native alone

D: Asian alone

E: Native Hawaiian and Other Pacific Islander alone

F: Some Other Race alone

G: Two or More Races

H: Hispanic or Latino

I: White alone, not Hispanic or Latino

Some tables are repeated J through AH, which includes race alone and race alone or in combination with one or more races by Hispanic or not Hispanic

# Understanding the DHC Demonstration Product

## Person and Unit Files

### Person File

Geographic Variables (76)  
 Group Quarters Type  
 Relationship  
 Sex  
 Age  
 Hispanic Origin  
 Race  
 Lives Alone

Census data are coded to create Person and Unit file variables and disclosure avoidance is applied independently to the Person and Unit files. Everything in the DHC is tabulated from the protected Person and Unit variables

### Unit File

Geographic Variables (76)  
 Group Quarters Type  
 Tenure  
 Vacancy Status  
 Population Count  
 Household/Family Type  
 Household/Family Type (Includes Cohabiting)  
 Couple Type  
 Presence and Type of Unmarried Partner Household  
 Multigenerational Household  
 Age of Householder  
 Hispanic Householder  
 Race of Householder  
 Presence and Age of Own Children Under 18  
 Presence of People Under 18 Years in Household  
 Presence of People 60 Years and Over in Household  
 Presence of People 65 Years and Over in Household  
 Presence of People 75 Years and Over in Household  
 Presence and Age of Children Under 18  
 Sex of Householder

No links between members of the same household on the **Person** file

No characteristics of individual household members on the **Unit** file except for those encoded in **Unit** file variables

Tables that require a link between household members are included in the Detailed DHC

# Understanding the DHC

## The Person File

<b>Person File</b>
Geographic Variables (76)
Group Quarters Type
Relationship
Sex
Age
Hispanic Origin
Race
Lives Alone
Number of Race Groups

Some of the tables tabulated from the protected DHC Person file include:

**P2** Hispanic or Latino Origin By Race

**P3** Race for the Population 18 Years and Over

**P5** GQ Population by Major GQ Type

**P12A-AH** Sex by Age for Selected Age Categories

**P14** Sex by Age for the Population Under 20 Years (SYA)

**PCO11** Age of Grandchildren Under 18 Years  
Living with a Grandparent Householder

**P19** Household Type (Household vs GQ) by Relationship

**PCO19A-I** Household Type (Household vs GQ) by Relationship

**PST1** Group Quarters Population by Sex by Age by Group Quarters Type  
(detailed)

These tables will match the P.L. Production Setting demonstration product – 2021-06-08



# Understanding the DHC

## The Unit File

<b>Unit File</b>
Geographic Variables (76)
Group Quarters Type
Tenure
Vacancy Status
Population Count
Household/Family Type
Household/Family Type (Includes Cohabiting)
Couple Type
Presence and Type of Unmarried Partner Household
Multigenerational Household
Age of Householder
Hispanic Householder
Race of Householder
Presence and Age of Own Children Under 18
Presence of People Under 18 Years in Household
Presence of People 60 Years and Over in Household
Presence of People 65 Years and Over in Household
Presence of People 75 Years and Over in Household
Presence and Age of Children Under 18
Sex of Householder

Some of the tables tabulated from the protected DHC Unit file include:

**H2** Occupancy Status



Will match the P.L.  
Production Settings

**H4** Tenure (repeated by race)

**H5** Vacancy Status (detailed)

**P16** Household Type (repeated by race)

**P18** Households by Presence of People 65 Years and Over, Household Size, and Household Type

**P21** Households by Presence of People Under 18 Years

**PCO3** Household Type by Age of Householder

**PCO4** Households by Presence of People 60 Years and Over by Household Type

**PCO6** Households by Presence of People 75 Years and Over, Household Size, and Household Type

**PCO10** Family Type by Presence and Age of Own Children



# Understanding the DHC Demonstration Product

## Our Internal Analysis

### What We Looked For

- Results that would be noticeably incorrect
- Measures of average difference that exceed standards (numeric or percent)
- Differences that change patterns or that would change the story or headline
- Inconsistencies from applying disclosure avoidance to the **Person** and **Unit** files independently

Block-level data are fit-for-use when aggregated into geographically contiguous larger entities. They are not intended to be fit-for-use as a unit of analysis

# Understanding the DHC Demonstration Product

## Internal Analysis, What Looks Good and What Is Getting Additional Attention

**Despite substantial improvements from earlier results, below are some examples of areas for improvement:**

**Records may have values that were not present within a geography during enumeration**

- In some cases, the GQ Population was assigned a Detailed GQ type not present within the state
- Housing may have a detailed vacancy type that does not exist in the area such as “For migrant workers”

**Because protected values cannot be negative, small groups often have small numeric but large percent differences combined with a positive bias**

- Larger differences for some ages such as 18 to 24-year-olds (GQ population) and centenarians
- Lower accuracy for counts of people by some relationship to the householder categories
- Lower accuracy for same-sex married and same-sex unmarried couple households
- Bias towards higher counts of households with younger children present

# Understanding the DHC Demonstration Product

## Mean Absolute Error for Sex by Age Groups – Counties

	Demonstration Data (2019-10-29)	Demonstration Data (2020-05-27)	Full DHC Demonstration Data (2022-03-16) $\rho = 3.3$
Total			
0 to 17 years	46.43	42.95	<b>9.84</b>
18 to 64 years	99.17	231.70	<b>12.83</b>
65 years and over	103.04	230.85	<b>12.66</b>
Male			
0 to 17 years	65.52	92.39	<b>7.59</b>
18 to 64 years	90.78	195.33	<b>10.29</b>
65 years and over	74.38	137.97	<b>7.31</b>
Female			
0 to 17 years	64.36	93.18	<b>7.47</b>
18 to 64 years	93.85	181.53	<b>8.03</b>
65 years and over	73.49	147.95	<b>7.24</b>

# Understanding the DHC Demonstration Product

## Mean Absolute Error for 5 Year Age Groups – Counties

	Demonstration Data (2019-10-29)	Demonstration Data (2020-05-27)	Full DHC Demonstration Data (2022-03-16) $\rho = 3.3$
Under 5 years	96.50	83.67	<b>7.75</b>
5 to 9 years	111.47	81.73	<b>7.42</b>
10 to 14 years	110.19	81.50	<b>7.36</b>
15 to 19 years	100.38	92.31	<b>12.03</b>
20 to 24 years	100.12	106.48	<b>18.20</b>
25 to 29 years	107.86	95.39	<b>16.06</b>
30 to 34 years	114.02	89.38	<b>12.53</b>
35 to 39 years	97.79	87.08	<b>12.17</b>
40 to 44 years	97.65	84.14	<b>12.07</b>
45 to 49 years	116.99	87.00	<b>11.95</b>
50 to 54 years	110.95	84.56	<b>11.17</b>
55 to 59 years	97.19	82.07	<b>10.93</b>

# Understanding the DHC Demonstration Product

## Mean Absolute Error for 5 Year Age Groups – Counties (cont.)

	Demonstration Data (2019-10-29)	Demonstration Data (2020-05-27)	Full DHC Demonstration Data (2022-03-16) $\rho = 3.3$
60 to 64 years	99.92	81.74	<b>6.76</b>
65 to 69 years	99.45	76.22	<b>7.02</b>
70 to 74 years	99.40	76.12	<b>9.69</b>
75 to 79 years	85.14	72.49	<b>9.43</b>
80 to 84 years	80.09	69.43	<b>8.87</b>
85 to 89 years	78.29	65.74	<b>23.63</b>
90 to 94 years	65.03	53.44	<b>16.83</b>
95 to 99 years	39.15	35.26	<b>14.71</b>
100 to 104 years	10.77	12.86	<b>6.46</b>
105 to 109 years	3.71	2.80	<b>2.00</b>
110 to 115 years	4.27	2.43	<b>2.24</b>

# Understanding the DHC Demonstration Product

## Mean Absolute Error and Mean Absolute Percent Error by Household Type – Counties

2010 Demonstration Data Product for the DHC (2022-03-16)

	MAE	MAPE (%)	Count of Counties with a Difference Exceeds 5%
Opposite-sex married couple household	6.01	0.21	5
Same-sex married couple household	3.56	20.75	1,897
Opposite-sex unmarried partner household	5.09	2.24	271
Same-sex unmarried partner household	3.45	31.31	1,932

# Understanding the DHC Demonstration Product

## Mean Absolute Error and Mean Absolute Percent Error for Relationship Categories – Counties

2010 Demonstration Data Product for the DHC (2022-03-16)

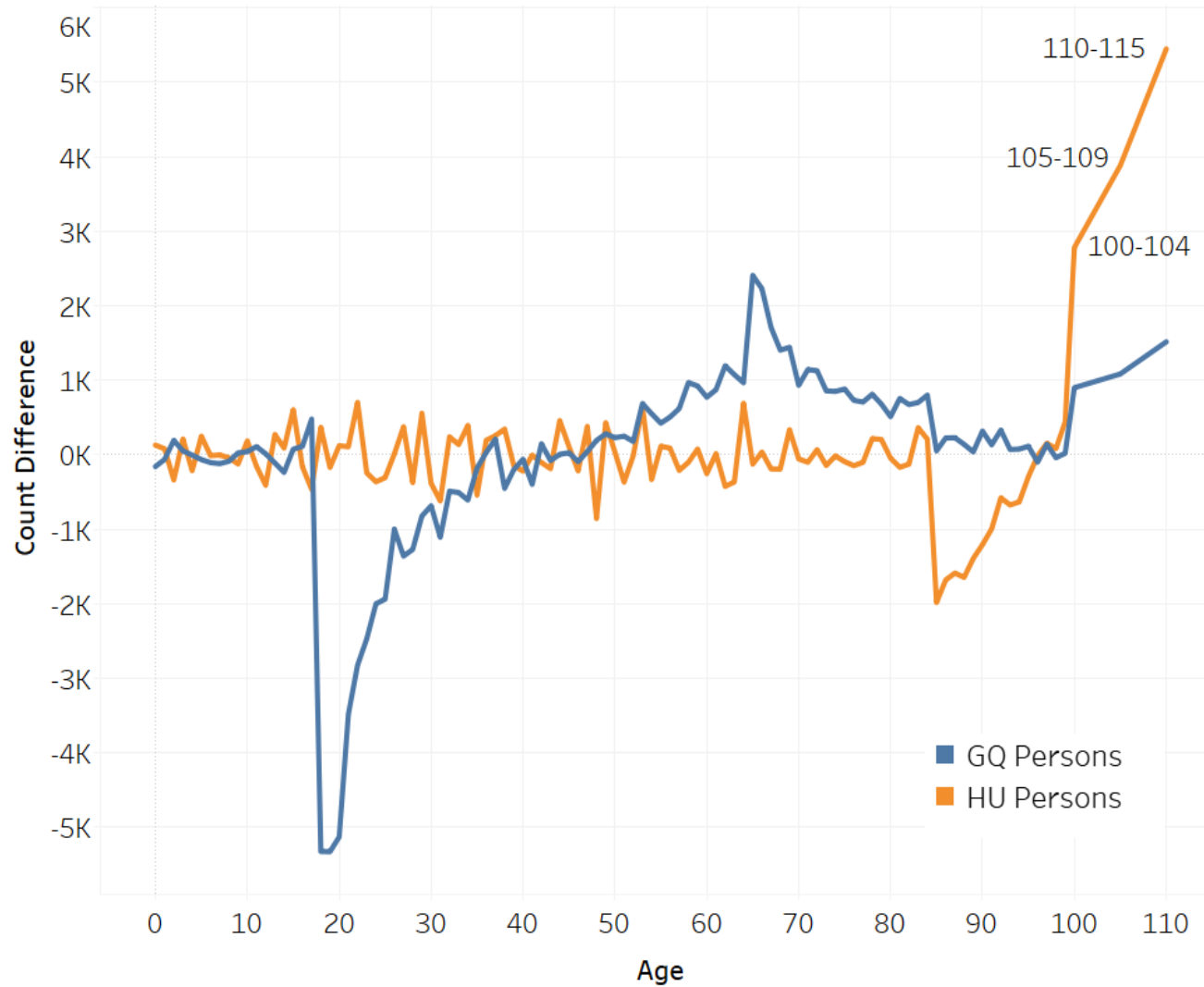


CBDRB-FY22-DSEP-003

	MAE	MAPE (%)
Householder	135.81	1.78
Opposite-sex husband/wife/spouse	74.24	1.95
Opposite-sex unmarried partner	28.10	9.24
Same-sex husband/wife/spouse	19.24	70.55
Same-sex unmarried partner	25.39	132.98
Biological son or daughter	88.76	1.73
Adopted son or daughter	29.32	21.97
Stepson or stepdaughter	30.55	10.67
Brother or sister	42.31	37.61
Father or mother	27.44	32.03
Grandchild	32.08	9.93
Parent-in-law	22.62	82.32
Son-in-law or daughter-in-law	25.24	50.01
Other relative	50.24	38.42
Roommate or housemate	42.80	48.02
Foster child	13.84	122.09
Other nonrelative	41.63	24.51

# Understanding the DHC Demonstration Product

## Difference by Age, Housing Unit and Group Quarters Population – Nation



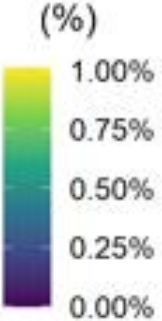
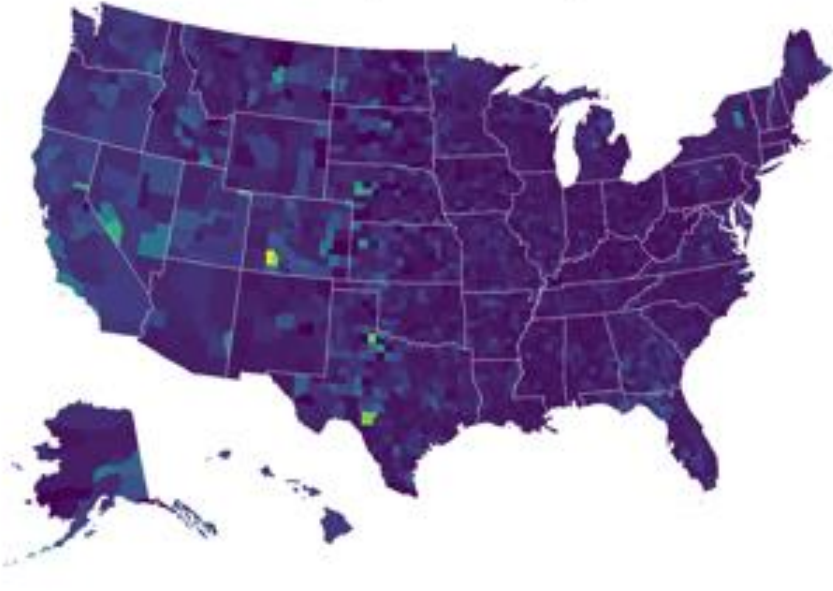
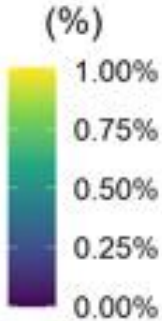
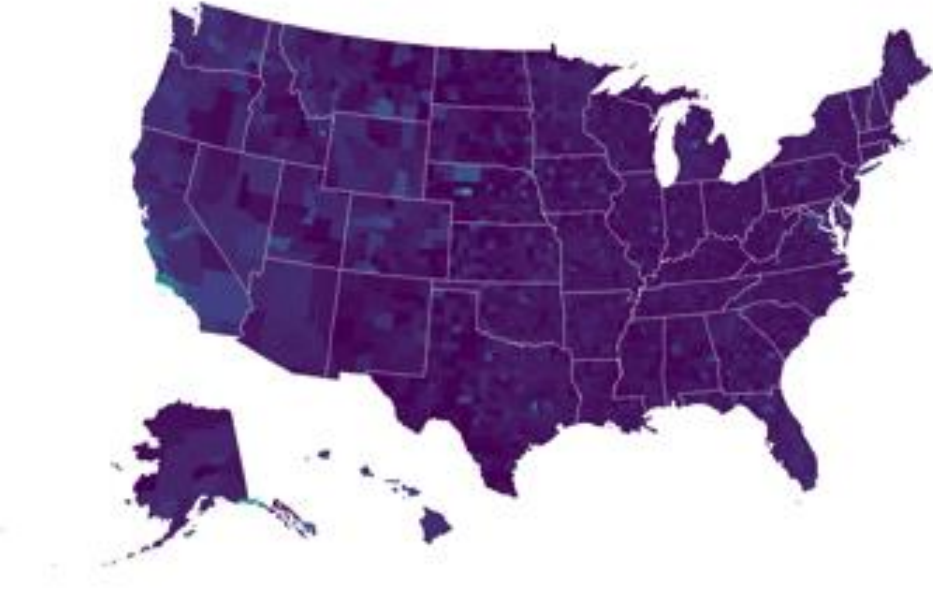


# Understanding the DHC Demonstration Product

## Percent Foster Child - Counties

**Percent Foster Child**  
2010 Census Summary File

**Percent Foster Child**  
2010 Demonstration Data Product for the  
DHC (2022-03-16)



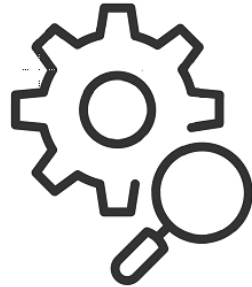
# Understanding the DHC Demonstration Product

## How to see the Impact of Disclosure Avoidance Using the Detailed Summary Metrics

The Detailed Summary Metrics are provided in a Microsoft Excel worksheet with worksheets for different sets of metrics, all you have to do is open it and start looking at the results!



Developed with  
input from data  
users



A tool to evaluate  
incremental  
improvements



Measures:  
accuracy  
bias  
outliers  
impossible or  
improbable results



Includes metrics for data  
user-provided use cases

# Understanding the DHC Demonstration Product

## Using the Summary Files to See the Impact of Disclosure Avoidance

Compare the published table to the demonstration table to see impact from applying disclosure avoidance\*

[Published Table]

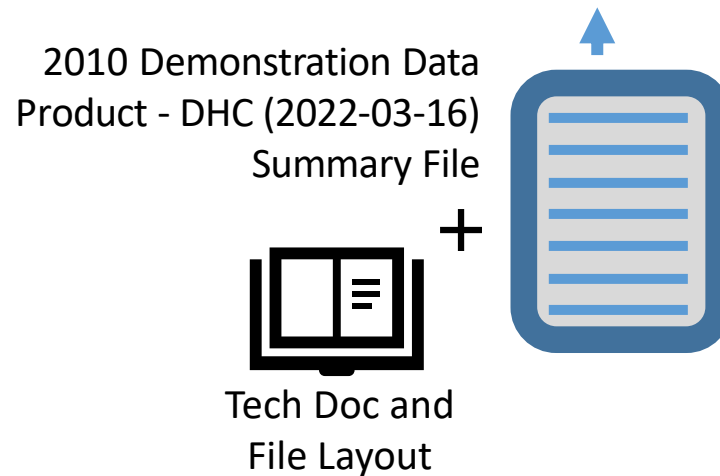
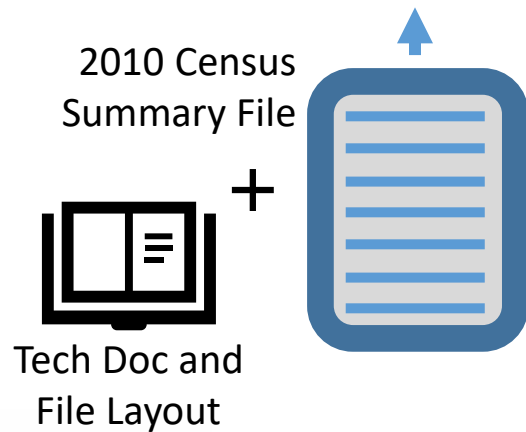
Total:	182,420
Male:	89,340
Under 5 years	5,616
5 to 9 years	5,945
10 to 14 years	6,012
15 to 17 years	3,704
18 and 19 years	2,225
20 years	1,011
21 years	8610
...	

[Demonstration Table]

Total:	182,430
Male:	89,350
Under 5 years	5,614
5 to 9 years	5,950
10 to 14 years	6,020
15 to 17 years	3,745
18 and 19 years	2,226
20 years	1,013
21 years	8620
...	

Impact from applying disclosure avoidance\*

Total:	-10
Male:	-10
Under 5 years	+2
5 to 9 years	-5
10 to 14 years	-8
15 to 17 years	-41
18 and 19 years	-1
20 years	-2
21 years	-10
...	



\* Differences will also include differences due to the swapping that was applied to the 2010 published tables

# Understanding the DHC Demonstration Product

## Online Walk-Through of Key Items

### **What we will show you:**

- 2020 Census Disclosure Avoidance Website
- 2010 Demonstration Detailed Accuracy Metrics
- 2010 Census Summary Files and Demonstration Product Summary Files
- 2010 Demonstration Summary File Technical Documentation
- 2020 Census Data Product Planning Crosswalk