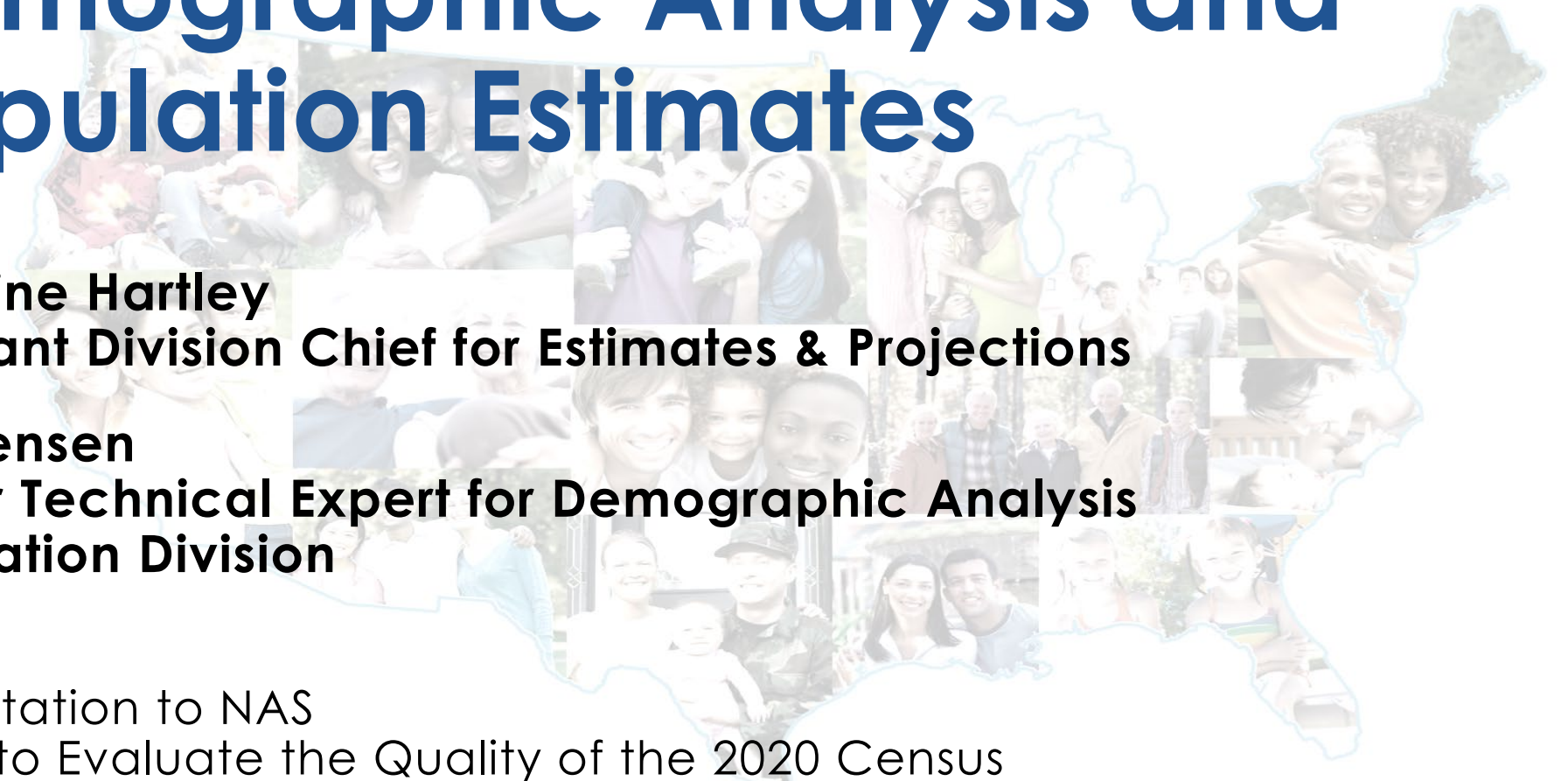


Demographic Analysis and Population Estimates



Christine Hartley
Assistant Division Chief for Estimates & Projections

Eric Jensen
Senior Technical Expert for Demographic Analysis
Population Division

Presentation to NAS
Panel to Evaluate the Quality of the 2020 Census
July 19-21, 2021

Shape
your future
START HERE >

United States[®]
Census
2020

What is Demographic Analysis?

- **Demographic Analysis (DA) is a method that the Census Bureau uses to evaluate the quality of the decennial census**
- **We produced national estimates of the population on April 1, 2020 by age, sex, DA race categories, and Hispanic origin**
- **Estimates are developed using current and historical vital records, data on international migration, and Medicare records**
- **The results are used to develop estimates of net coverage error at the national level by demographic detail**

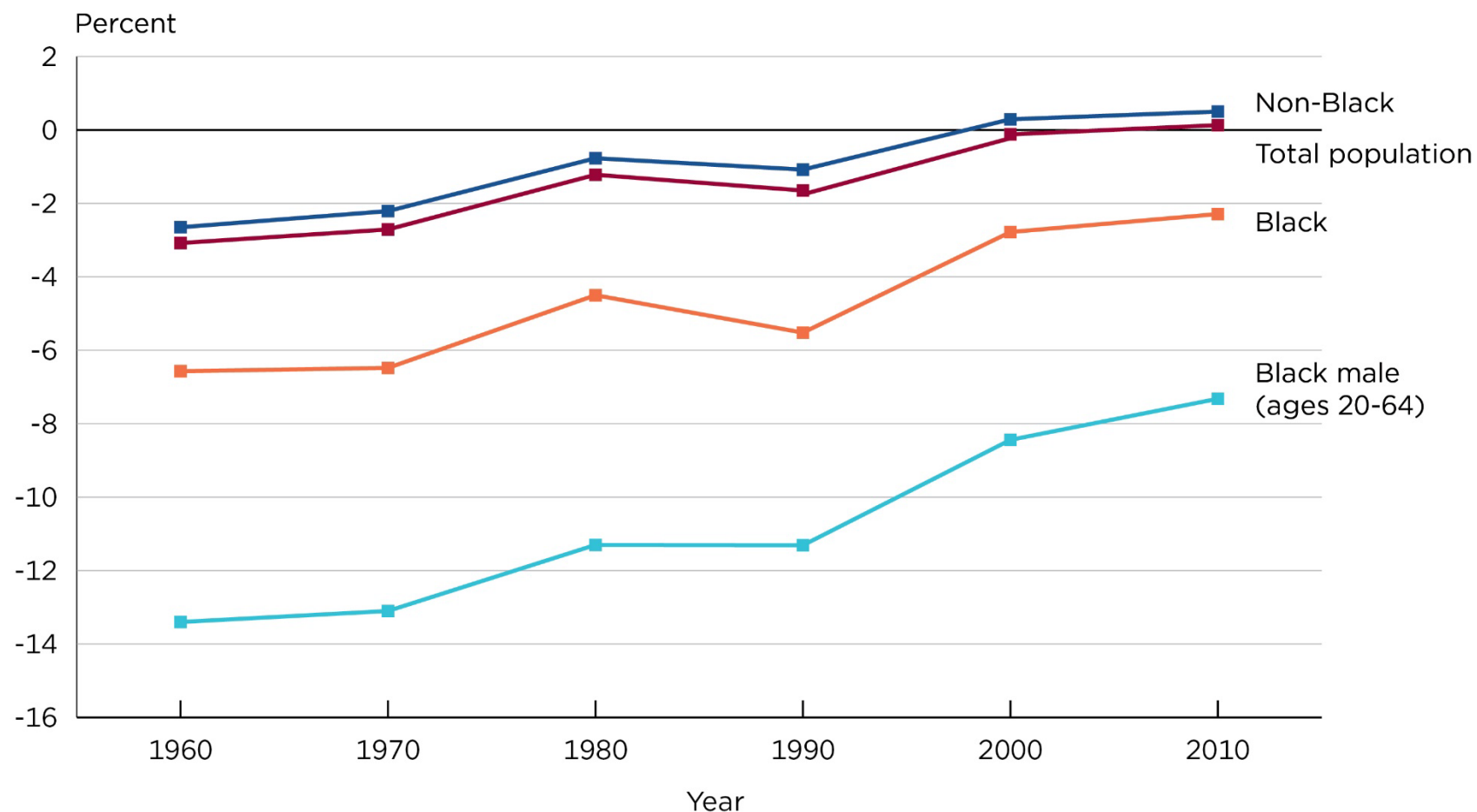
What Demographic Analysis Tells Us About the Quality of the 2020 Census Results

- DA can tell us about the net coverage error for specific age, sex, DA race, and Hispanic origin groups at the national level

$$\frac{100 * (\textit{Census count} - \textit{DA estimate})}{\textit{DA estimate}}$$

- The strength of DA is what it tells us about differential coverage for demographic groups either within the same census or over time
 - Coverage differentials by race
 - Undercount of young children

Demographic Analysis Estimates of Net Coverage Error From 1960 to 2010



Note: Net coverage error is calculated as the percent difference between the census counts and the Demographic Analysis estimates.

Source: U.S. Census Bureau, Population Division, Demographic Analysis Program, Special Tabulation.

Demographic Analysis Method

$$\text{Population} = \text{Births} - \text{Deaths} + \text{Immigration} - \text{Emigration}$$

- Birth records are the foundation of the DA estimates
- The total estimate for each cohort is calculated by answering two questions:
 - Who was in the birth records from 1945 to 2020 but was not living in the United States on April 1, 2020?
 - Who was living in the United States on April 1, 2020 but was not in the birth records?

Demographic Analysis Method (cont.)

- **We use Medicare enrollment records to estimate the population born before 1945, or ages 75 and older on April 1, 2020**
 - Adjustments are made to account for underenrollment
- **To calculate the total population, we sum the estimates for each birth cohort from 1945 to 2020 and add them to the estimates of the population aged 75 and older**

2020 Demographic Analysis

- **Three sets of Demographic Analysis (DA) population estimates**
 - Black Alone/non-Black Alone
 - Black Alone or in Combination/non-Black Alone or in Combination
 - Hispanic (ages 0-29 only)
- **National estimates by age, sex, DA race groups, and Hispanic origin**
 - DA estimates have been limited to the Black and non-Black categories
 - Hispanic origin information was not available on the birth certificates for all states until 1990
- **Low, middle, and high series for each set to measure uncertainty**
- **Components of population change for all three sets and series**
 - Births, deaths, international migration, Medicare-based estimates

Census Counts, Demographic Analysis (DA) Estimates of the Resident Population, and Net Coverage Error: April 1, 2020, and April 1, 2010

(Population in thousands. Net coverage error in percent)

Measure	2020 Census	2020 DA series			2010 Census	2010 DA series				
		Low	Middle	High		Low	Low middle	Middle ¹	High middle	High
Population.	331,449	330,730	332,601	335,514	308,746	305,684	307,415	308,346	310,038	312,713
Net coverage error ² . .	X	0.22	-0.35	-1.21	X	1.00	0.43	0.13	-0.42	-1.27

X Not applicable.

¹ The 2010 middle estimate comes from the May 2012 revision to the DA estimates. The other 2010 DA estimates were part of the original December 2010 release.

² Calculated as (Census count - DA estimate)/DA estimate * 100.

Source: U.S. Census Bureau, 2020 and 2010 Censuses, and 2020 and 2010 Demographic Analysis.

Census Counts, Demographic Analysis (DA) Estimates of the Resident Population, and Net Coverage Error: April 1, 2020, and April 1, 2010

(Population in thousands. Net coverage error in percent)

Measure	2020 Census	2020 DA series			2010 Census	2010 DA series				
		Low	Middle	High		Low	Low middle	Middle ¹	High middle	High
Population.....	331,449	330,730	332,601	335,514	308,746	305,684	307,415	308,346	310,038	312,713
Net coverage error ² ..	X	0.22	-0.35	-1.21	X	1.00	0.43	0.13	-0.42	-1.27

X Not applicable.

¹ The 2010 middle estimate comes from the May 2012 revision to the DA estimates. The other 2010 DA estimates were part of the original December 2010 release.

² Calculated as (Census count - DA estimate)/DA estimate * 100.

Source: U.S. Census Bureau, 2020 and 2010 Censuses, and 2020 and 2010 Demographic Analysis.

Census Counts, Demographic Analysis (DA) Estimates of the Resident Population, and Net Coverage Error: April 1, 2020, and April 1, 2010

(Population in thousands. Net coverage error in percent)

Measure	2020 Census	2020 DA series			2010 Census	2010 DA series				
		Low	Middle	High		Low	Low middle	Middle ¹	High middle	High
Population.	331,449	330,730	332,601	335,514	308,746	305,684	307,415	308,346	310,038	312,713
Net coverage error ² . .	X	0.22	-0.35	-1.21	X	1.00	0.43	0.13	-0.42	-1.27

X Not applicable.

¹ The 2010 middle estimate comes from the May 2012 revision to the DA estimates. The other 2010 DA estimates were part of the original December 2010 release.

² Calculated as (Census count - DA estimate)/DA estimate * 100.

Source: U.S. Census Bureau, 2020 and 2010 Censuses, and 2020 and 2010 Demographic Analysis.

Census Counts, Demographic Analysis (DA) Estimates of the Resident Population, and Net Coverage Error: April 1, 2020, and April 1, 2010

(Population in thousands. Net coverage error in percent)

Measure	2020 Census	2020 DA series			2010 Census	2010 DA series				
		Low	Middle	High		Low	Low middle	Middle ¹	High middle	High
Population.....	331,449	330,730	332,601	335,514	308,746	305,684	307,415	308,346	310,038	312,713
Net coverage error ² ..	X	0.22	-0.35	-1.21	X	1.00	0.43	0.13	-0.42	-1.27

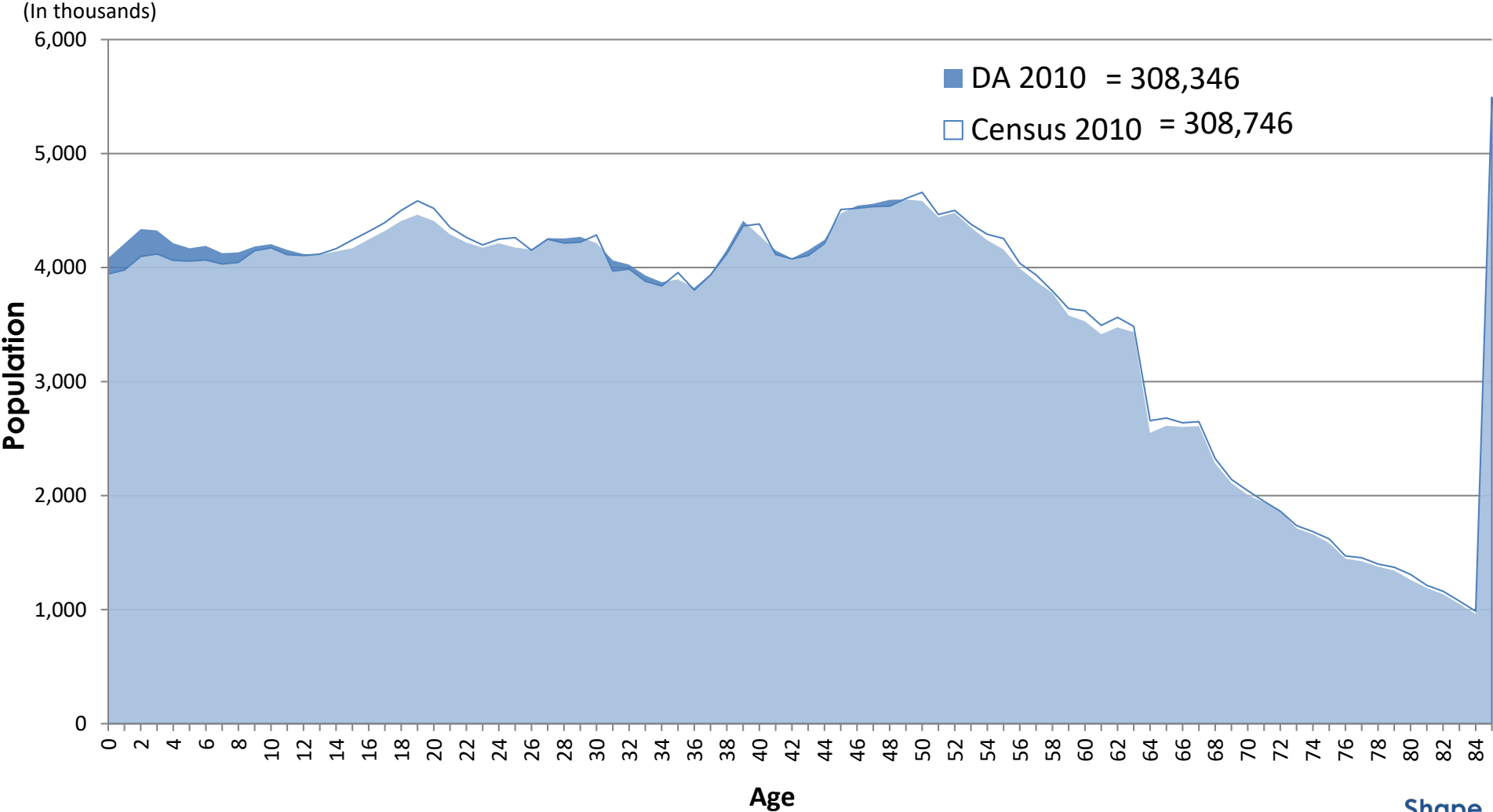
X Not applicable.

¹ The 2010 middle estimate comes from the May 2012 revision to the DA estimates. The other 2010 DA estimates were part of the original December 2010 release.

² Calculated as (Census count - DA estimate)/DA estimate * 100.

Source: U.S. Census Bureau, 2020 and 2010 Censuses, and 2020 and 2010 Demographic Analysis.

Demographic Analysis Estimates and 2010 Census Counts by Age in 2010



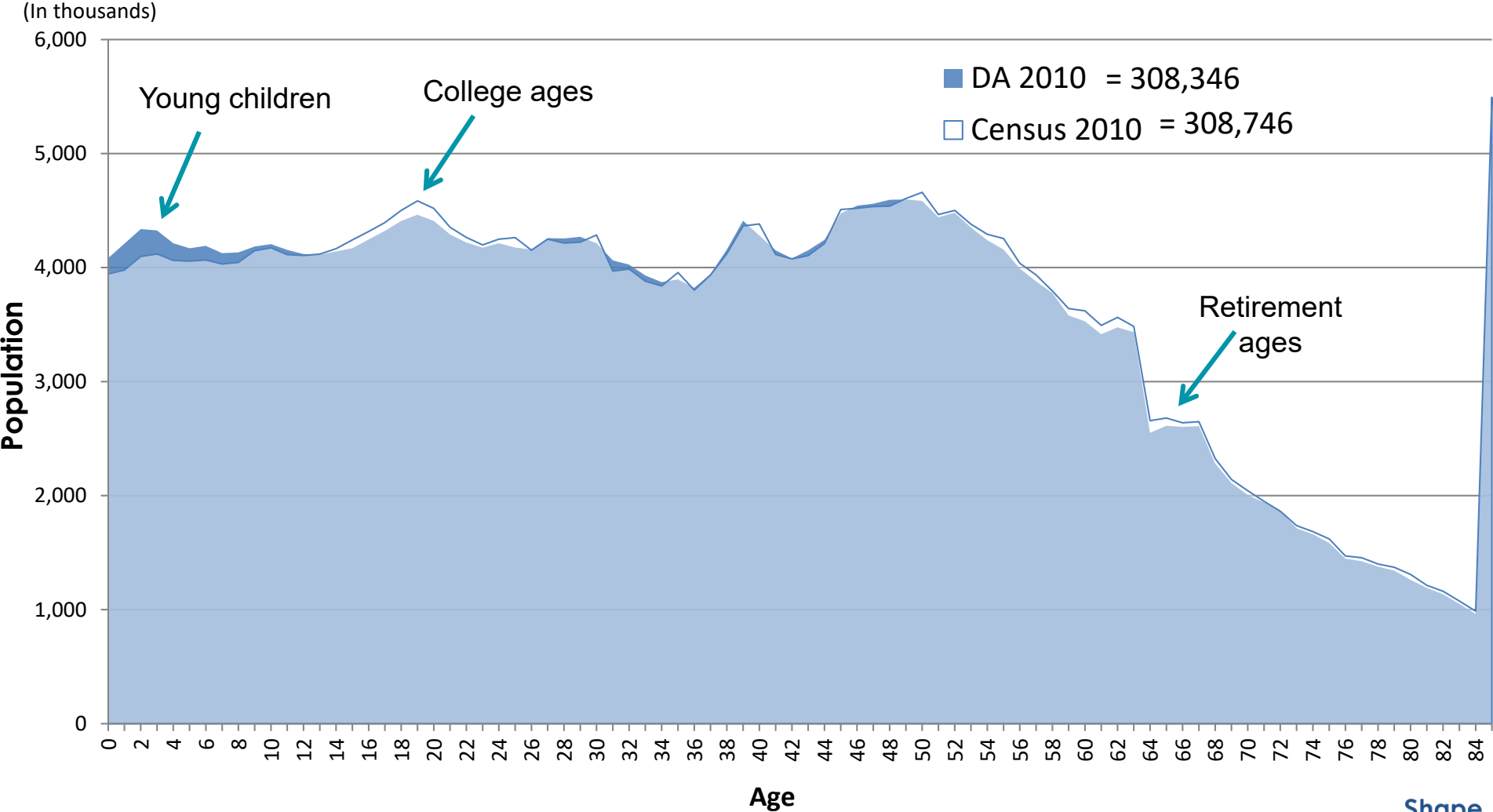
Source: U.S. Census Bureau, 2010 Census, and 2010 Demographic Analysis (May 2012 DA Release).

2020CENSUS.GOV

Shape
your future
START HERE >

United States[®]
Census
2020

Demographic Analysis Estimates and 2010 Census Counts by Age in 2010



Source: U.S. Census Bureau, 2010 Census, and 2010 Demographic Analysis (May 2012 DA Release).

2020CENSUS.GOV

Shape
your future
START HERE >

United States[®]
Census
2020

Next Steps for DA

- **Report on DA estimates of net coverage error**
- **Experimental sets of estimates**
 - Extended Hispanic origin (ages 0-39)
 - Churning race (ages 0-85+)
 - Full race and Hispanic origin detail (ages 0-17)
 - State and county estimates (ages 0-4)

Annual Estimates Detail

Estimate Type	U.S.	States	Metro/ Micro Statistical Areas	Counties	Cities and Towns	Puerto Rico	Puerto Rico Municipios
Total Resident Population	X	X	X	X	X	X	X
Components of Change (Births, Deaths, Net Migration)	X	X	X	X		X	
Population Characteristics (Age, Sex, Race, Hispanic Origin*)	X	X		X		X	X
Monthly Population (Five Universes**)	X						
Group Quarters Population	X (by char)	X (totals)		X (totals)			
Housing Units	X	X		X			

*For Puerto Rico Commonwealth and its municipios, population estimates are produced by age and sex only.

**The five estimates universes are: resident, resident plus Armed Forces overseas, civilian, civilian noninstitutionalized, household.

Methods

- County-level geographies and above: cohort-component method measuring population change since the last census using the most current administrative records on births, deaths, and migration
- Subcounty geographies: distributive housing unit method to estimate the household population, which is added to an estimate of the population in group quarters
- Housing units: component-based method beginning with the most recent decennial census and measuring change in the housing stock using survey and administrative data

Using the Vintage 2020 Estimates to Evaluate the Decennial Census

- **The annual estimates series are typically based on the latest decennial census**
 - Vintage 2020 estimates are based on the 2010 Census
- **Differences between the Vintage 2020 estimates and the 2020 Census may result from several sources:**
 - Error in the previous census (2010 Census)
 - Estimation error in the components of demographic change (V2020)
 - Error in the new census (2020 Census)
- **Careful, deliberate comparisons are still possible**

2020 Estimates Evaluation (E2)

- **Estimates Evaluation (E2) is an error-of-closure analysis where the results of the census are compared to the population estimates**
- **A team is working on planning the 2020 E2**
 - Examining what we have done past decades
 - Assessing our current needs
 - Identifying and coding metrics
 - Working closely with our partners in the FSCPE

Measures of Difference

- **Percent Error: Mean Absolute Percent Error (MAPE)**

$$((\sum(|\text{Estimate} - \text{Census}|)/\text{Census}))/N * 100$$

- **Bias: Mean Algebraic Percent Error (MALPE)**

$$((\sum((\text{Estimate} - \text{Census})/\text{Census}))/N) * 100$$

- **Numeric Error: Root Mean Squared Error (RMSE)**

$$\text{SQRT}(\sum((\text{Estimate} - \text{Census})^2)/N)$$

- **Outliers: Percent Difference Threshold**

Number of areas with differences plus/minus 5% and 10%

- **Distribution: Total Absolute Error of Shares**

$$\sum |(\text{Estimate}/\sum \text{Estimate}) - (\text{Census}/\sum \text{Census})|$$

Blending the 2020 Population Estimates Base

- The census typically forms the base for the population estimates
- Research is needed to evaluate the 2020 Census data for this purpose, but the files will not be available with sufficient time
- “Blended base” method: control April 1, 2020 from Vintage 2020 (based on the 2010 Census) to other sources to generate a plausible base
- **Potential data sources**
 - Invariant state total population from the 2020 Census
 - National age/sex detail from 2020 Demographic Analysis
 - Official county totals from the Microdata Detail File (PL-MDF)

Next Steps for the Vintage 2021 Blended Base

- **Finalize methodology**
- **Stakeholder outreach**
 - Webinar for the Federal-State Cooperative for Population Estimates
 - Other webinars
- **Estimates production**

Beyond Vintage 2021

- **Evaluations of coverage measures and the results from E2 will help us to determine how the incorporation of the 2020 Census will affect the estimates base**
 - Are there specific geographies/characteristics that are concerning?
- **Depending on our findings, the blended base could be extended to Vintage 2022 (and beyond)**
 - This could permit additional time for research and the development and implementation of mitigation strategies
 - Findings from E2 may result in the incorporation of additional controls from the 2020 Census into the blended base
- **Another option is to begin to use 2020 Census CEF data as the base (with differentially private noise infused into the estimates)**

Q&A

END