Status Update on the Population Estimates

Christine Hartley, Ph.D.

Assistant Division Chief for Estimates & Projections

Population Division

U.S. Census Bureau

Presentation to the Committee on National Statistics
2020 Census Data Products: Workshop on the Demographic and Housing Characteristics Files
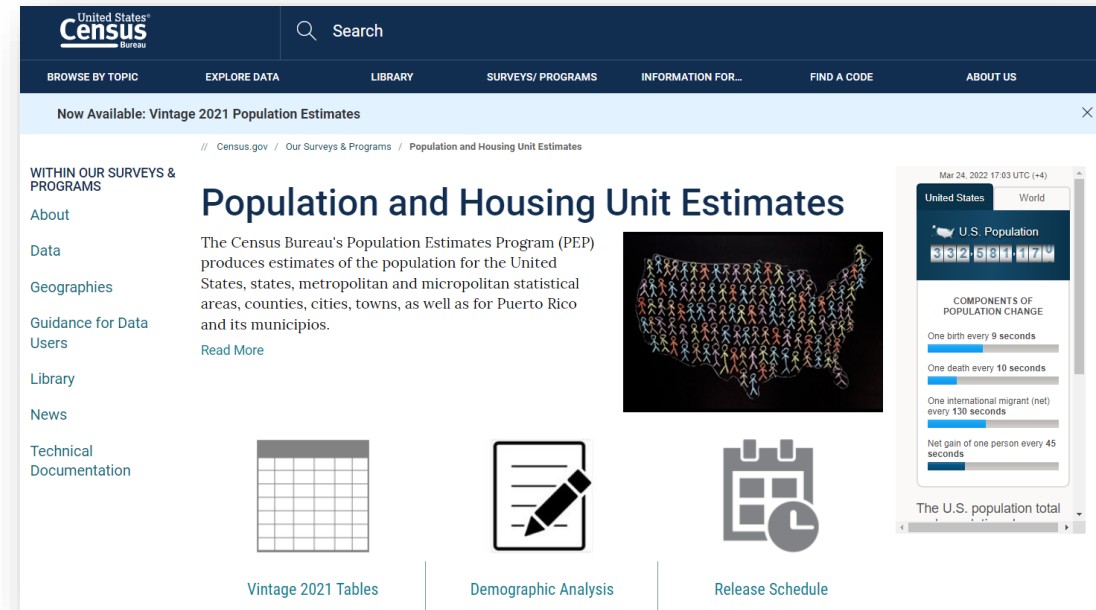June 21, 2022

United States® Census Bureau

1

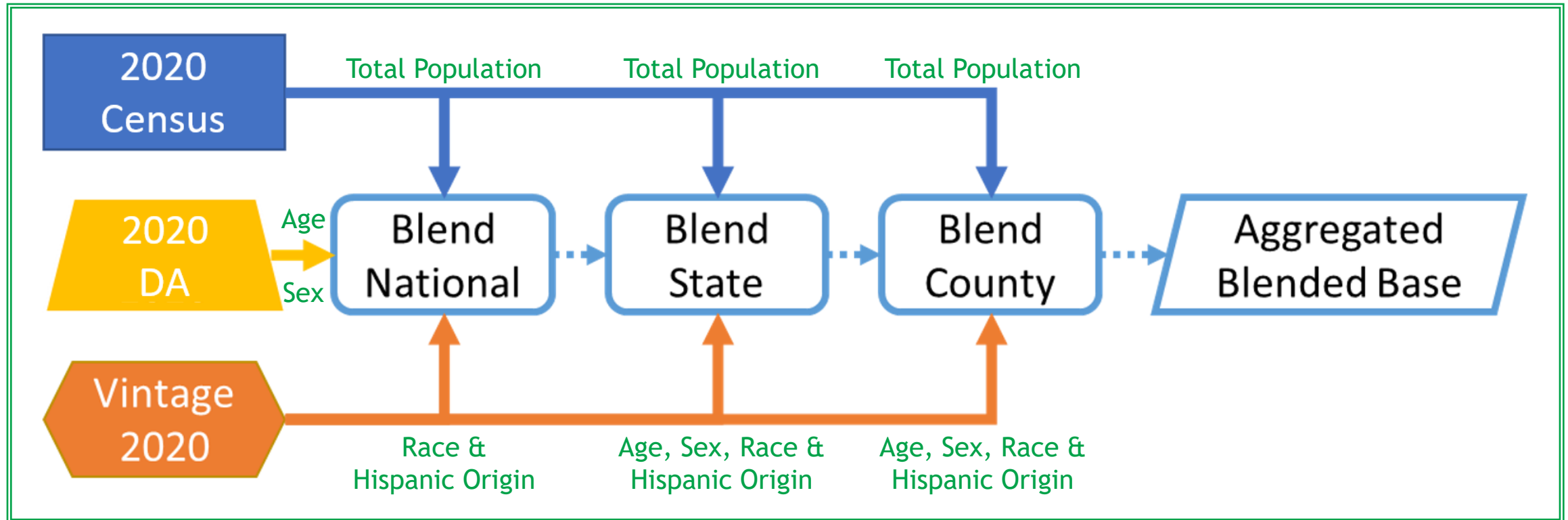# The Population Estimates Program

- The Population Estimates Program (PEP) disseminates official measures of population and housing units between decennial censuses

- Mandated by federal law

- Use cases include:
  - Population controls and denominators
  - Academic and business research
  - Program planning in the public and private sectors

- Time series are released annually featuring data for the date of the last census through the vintage year, which represents the latest year of estimates available

- Current estimates series is Vintage 2021 (April 1, 2020 to July 1, 2021)

Decennial Census Data in the Estimates

- Typically, the decennial census is used as the base population from which the annual times series of estimates are built
  - Nation, states, and counties: cohort-component method
  - Cities and towns: distributive housing unit method
- Census data also used for:
  - Kidlink File (assigning race and Hispanic origin to births)
  - Demographic Characteristics File (assigning race and Hispanic origin to domestic migrants)

United States® Census Bureau

Blended Base Approach

- Blended inputs (April 1, 2020 reference date)
  - Vintage 2020 postcensal, full-detail county-level population estimate (based on the 2010 Census)
  - 2020 Demographic Analysis (DA) national estimates by sex and age (0-84, 85+)
  - 2020 Decennial Census total population counts (PL 94-171)

Plans for Future Vintages

- V2022 base population methodology improvements
  - Use GQ and household population from the 2020 Census at the county level
  - Expand use of DA age detail to 100+ to maximize its utility
- PEP staff will continue annual research efforts to improve other elements of our estimates methodology
- Ongoing research will determine other base population updates for V2023 and beyond
- Key considerations for a 2020 Census estimates base:
  - Data quality (coverage issues)
  - Disclosure avoidance

The Base Evaluation and Research Team (BERT)

- BERT is researching the feasibility of taking coverage measures from DA and the Post-Enumeration Survey (PES) into account in the development of the population estimates
- BERT is made up of Census Bureau subject-matter experts in the areas of:
  - Population estimates
  - Age and sex statistics
  - Coverage measurement
  - Race and ethnicity
  - Demography
  - Disclosure avoidance

Census Data in the Estimates

- Typically, the decennial census is used as the base population from which the annual times series of estimates are built
  - Nation, states, and counties: cohort-component method
  - **Cities and towns: distributive housing unit method** ⟵

    > V2021 estimates already based on 2020 Census data + differentially private noise

- Census data also used for:
  - Kidlink File (assigning race and Hispanic origin to births)
  - Demographic Characteristics File (assigning race and Hispanic origin to domestic migrants)

Options for Creating Differentially Private Population Estimates

Microdata Detail File

- No privacy loss budget (PLB) required
- Imputed race variable must be created
- Methodology changes needed to apply annual geographic updates, assign race and Hispanic origin to births, and assign race and Hispanic origin to domestic migrants

Census Edited File

- Requires PLB and development of a differential privacy mechanism to infuse noise into a very detailed tabulation
- PLB would need to cover each year of the decade so the estimates base could be redeveloped featuring geographic updates and results from decennial programs like CQR and PCGQR
- No methodology changes needed

Issues to Address

Microdata Detail File

- What updates to the methodology will enable us to apply annual geographic updates, assign race and Hispanic origin to births, and assign race and Hispanic origin to domestic migrants?

- What will the impact be on data quality by using these alternative methods?

Census Edited File

- Is it possible to develop a differential privacy mechanism to infuse acceptable levels of noise into the CEF at the level of demographic/geographic detail required for estimates processing?

- Is the amount of PLB that would be needed to infuse noise into such a detailed tabulation realistic?

Thank you.

Questions?