2020 Census: Post-Enumeration Survey

JASON Winter Study: 2020 Census Quality Process Assessment
Date: January 7, 2021

Patrick J. Cantwell
Chief, Decennial Statistical Studies Division

Timothy Kennel
Assistant Division Chief, Statistical Methods

Cleared for Release
DRB Number: CBDRB-FY21-DSSD007-0007
Outline

1. Introduction to Census Quality
2. 2020 Post-Enumeration Survey Design
3. Schedule and Reports
Introduction to Census Quality
Measuring Census Quality

- Census process indicators
- Assessments and evaluations of census operations
- Demographic Analysis (DA) and other methods
- Post-Enumeration Survey (PES)
Coverage Measurement in the Past

• Demographic analysis for decades
• Coverage measurement surveys
• 1980 Post-Enumeration Program
• 1990 Post-Enumeration Survey
• 2000 Accuracy and Coverage Evaluation
• 2010 Census Coverage Measurement survey
2010 Census and Demographic Analysis (in Thousands)

2010 Census
308,746

DA Low: 305,684
DA Middle Low: 307,415
DA Middle: 308,475
DA High Middle: 310,038
DA High: 312,713
Net Coverage Error

Net census coverage error (defined as)

\[ \text{Net census coverage error} = \text{Census count} - \text{True population size} \]

**Negative** coverage error means fewer people are included in the census than are in the true population (undercount)

**Positive** coverage error means more people are included in the census than are in the true population (overcount)

Note: In the past, that is, up through 2010, documentation of results from the Census Bureau’s post-enumeration surveys has generally defined the net undercount using the opposite of the difference above. When providing results from 2020, and in these slides covering past results, we will use the definition above.
# Net Undercount of US Household Population

<table>
<thead>
<tr>
<th>Year</th>
<th>Census Count</th>
<th>Net Coverage Error</th>
<th>Percent Coverage Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010</td>
<td>300,700,000</td>
<td>36,000.</td>
<td>0.01 %</td>
</tr>
<tr>
<td>2000</td>
<td>273,600,000</td>
<td>1,332,000*</td>
<td>0.49 %*</td>
</tr>
<tr>
<td>1990</td>
<td>248,700,000</td>
<td>– 3,994,000*</td>
<td>– 1.61 %*</td>
</tr>
</tbody>
</table>

* Statistically different from zero

Census counts for 2000 and 2010 exclude Group Quarters population.
Census count for 1990 included people in housing units and in noninstitutional and nonmilitary Group Quarters.
U.S. Total Percent Net Coverage Error from 1990 - 2010

Error bars indicate 90% confidence interval.
## Components of Census Coverage for the 2010 United States Household Population (in Thousands)

<table>
<thead>
<tr>
<th>Component of Census Coverage</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>Percent</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Census Count</td>
<td>300,700</td>
<td>0</td>
<td>100.0</td>
<td></td>
</tr>
<tr>
<td>Correct enumerations</td>
<td>284,700</td>
<td>199</td>
<td>94.7</td>
<td>0.07</td>
</tr>
<tr>
<td>Erroneous enumerations</td>
<td>10,040</td>
<td>199</td>
<td>3.3</td>
<td>0.07</td>
</tr>
<tr>
<td>Due to duplication</td>
<td>8,521</td>
<td>194</td>
<td>2.8</td>
<td>0.06</td>
</tr>
<tr>
<td>For other reasons</td>
<td>1,520</td>
<td>45</td>
<td>0.5</td>
<td>0.01</td>
</tr>
<tr>
<td>Whole-Person Census Imputations</td>
<td>5,993</td>
<td>0</td>
<td>2.0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Estimate of Population</strong></td>
<td>300,700</td>
<td>429</td>
<td>100.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Correct enumerations</td>
<td>284,700</td>
<td>199</td>
<td>94.7</td>
<td>0.1</td>
</tr>
<tr>
<td>Omissions</td>
<td>16,000</td>
<td>440</td>
<td>5.3</td>
<td>0.1</td>
</tr>
<tr>
<td><strong>Net Coverage Error</strong></td>
<td>36</td>
<td>429</td>
<td>0.01</td>
<td>0.14</td>
</tr>
</tbody>
</table>
2020 Post-Enumeration Survey Design
What is the Post-Enumeration Survey?

Goal
- Measure coverage of people and housing units in the decennial census
- Measure net coverage error and components of coverage
- No intent to adjust the census counts

How are the data used?
- To assess the success of the current census
- To provide information to improve the coverage of future censuses

How is coverage estimated?
- Conduct a survey in sample blocks
- Match people in the PES independent survey with enumerations in the census to determine who was missed or counted in error

History
- Used in the United States census since 1950
2020 Post-Enumeration Survey Universe

Household Population and Housing Units

Excludes

• Group Quarters and Group Quarters Population
• Remote Alaska
Sample Sizes

First sample (Basic Collection Units)
- Selected 9,800 Basic Collection Units in 50 states and District of Columbia
- Selected 400 Basic Collection Units in Puerto Rico

PES listing
- 510,000 housing units in 50 states and District of Columbia
- 18,000 housing units in Puerto Rico

After subsampling (Expected)
- 180,000 housing units in 50 states and District of Columbia
- 8,000 housing units in Puerto Rico
Two enumerations in sample areas

Capture (Enumeration sample, E sample)
- A sample of census records (enumerations) in the same block clusters, called E Sample
- Measure correct enumerations in census
- Determine if enumeration is correct or erroneous; duplicate? correct location?

Recapture (PES sample, population sample, P sample)
- A sample of blocks, independent of the census, in the post-enumeration survey (PES), called P Sample; interview people in units
- Match people in P Sample to Census
### Post-Enumeration Survey (PES)

#### How PES Estimates are Created

<table>
<thead>
<tr>
<th>Enumerated in PES?</th>
<th>YES</th>
<th>NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enumerated in Census?</td>
<td>YES</td>
<td>Found in Both</td>
</tr>
<tr>
<td>NO</td>
<td>Missing in Census Found in PES</td>
<td>Missing in Both</td>
</tr>
</tbody>
</table>
Current version of Dual System Estimator (DSE)

\[ \overline{TOT}_D = \sum_{j \in D} \left[ \pi_{dd,j}^{\pi_{ce,j}} / \pi_{m,j} \right] \]

Where
\( \pi_{dd,j} \) is the predicted probability census case \( j \) is data-defined
\( \pi_{cd,j} \) is the predicted probability that the census case is a correct enumeration
\( \pi_{m,j} \) is the predicted probability that the census case was matched
\( j \) is a census enumeration, and
\( D \) is the estimation domain (total, tenure status, age/sex groupings, etc.).
Schedule and Reports
## Revised field milestone schedule

<table>
<thead>
<tr>
<th>Operation</th>
<th>Original Dates</th>
<th>New Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Start</td>
<td>Finish</td>
</tr>
<tr>
<td>Independent Listing</td>
<td>1/16/2020</td>
<td>3/13/2020</td>
</tr>
<tr>
<td>Initial HU Followup</td>
<td>5/6/2020</td>
<td>6/19/2020</td>
</tr>
</tbody>
</table>

Operations in blue are complete.
Three major groups:

- Demographic or housing unit characteristics
  - Age Group
  - Age Group by Sex
  - Sex
  - Race and Hispanic Origin (alone or in-combination and mutually exclusive groups)
  - Relationship to Householder
  - Tenure (owner or renter)

- Geographic areas (states)
  - Region
  - State

- Census operations
  - Coverage Improvement Universe
  - GSS Partnership County
  - In-File Address Canvassing Outcome
  - Nonresponse Followup Respondent Type
  - Nonresponse Followup Workload
  - Response Mode
  - Response Rate Decile
  - Type of Enumeration Area
Schedule to Release Coverage Reports and Tables

November 2021

- National demographic estimates of net coverage of people
- National demographic estimates of components of coverage for people
- Tables can be downloaded on Census Bureau website

February 2022

- State coverage estimates for people
- More detailed national coverage estimates for people
- All housing unit coverage estimates
- Coverage estimates for Puerto Rico
Schedule to Release Methodology Reports

November 2021
• Source and Accuracy Statement – National Person

February 2022
• Estimation Design
• Estimation Methods: Characteristic Imputation
• Estimation Methods: Missing Data
• Estimation Methods: Net Coverage Estimation
• Estimation Methods: Component Estimation
• Source and Accuracy – All of PES
References


Questions