

**A Geographic Analysis of Characteristic Allocation in the 2010 Census**

Christine Flanagan Borman, Population Division, U.S. Census Bureau

**Abstract**

As the U.S. Census Bureau prepares to conduct the 2020 Census, it is useful to review outcomes from the 2010 Census. The *2010 Decennial Census: Item Nonresponse and Imputation Assessment Report* provides a detailed review of data quality, specifically regarding item non-response and characteristic imputation, for the person-level and household-level items from the 2010 Census at the national level. Another way to look at data quality is to break down the allocation rates at different geographic levels. In this paper, the 2010 Census allocation rates for the following person and housing-unit items – relationship to householder, sex, age/date of birth, Hispanic origin, race, tenure (if housing unit is occupied), and detailed vacancy status (if housing unit is vacant) – are analyzed at the region, state, and county levels. This paper will present levels and geographic patterns of allocations that will help inform expectations for the 2020 Census.

**Introduction**

The *2010 Decennial Census: Item Nonresponse and Imputation Assessment Report* provides a detailed review of data quality, specifically regarding item non-response and characteristic imputation, for the person-level and household-level items from the 2010 Census at the national level. The overall item non-response and imputation rates from the report are presented in the table below:

| <b>Overall Item Nonresponse and Imputation Rates</b> |                            |            |                          |                        |             |                                      |
|--|----------------------------|------------|--------------------------|------------------------|-------------|--------------------------------------|
|  | <b>Person-Level Items</b>  |            |                          |                        |             | <b>Household<br/>-Level<br/>Item</b> |
|  | <b>Relationship</b>        | <b>Sex</b> | <b>Age/Date of Birth</b> | <b>Hispanic Origin</b> | <b>Race</b> | <b>Tenure</b>                        |
| Item Nonresponse                                     | 1.5                        | 1.5        | 3.5                      | 3.9                    | 3.3         | 4.5                                  |
| Total  | 100.0                      | 100.0      | 100.0                    | 100.0                  | 100.0       | 100.0                                |
| “As Reported”  | 97.9                       | 98.4       | 95.0                     | 95.5                   | 95.9        | 96.5                                 |
| Imputed  | 2.1                        | 1.6        | 5.1                      | 4.5                    | 4.1         | 3.5                                  |
| Assigned   | 0.5                        | 1.3        | 1.5                      | 1.7                    | 1.2         | n/a                                  |
| Allocated  | 1.7                        | 0.3        | 3.6                      | 2.8                    | 2.9         | 3.5                                  |
| Substituted  | 1.9 percent of all persons |            |                          |                        |             |                                      |

While the *2010 Decennial Census: Item Nonresponse and Imputation Assessment Report* analyzes item-non response and characteristic imputation by return type (i.e. self-response returns, enumerator returns, etc.), another way to look at data quality is to explore the data at

different geographic levels. In this paper, the 2010 Census allocation rates are analyzed at the region, state, and county levels. Key research questions related to allocation include:

1. What are the allocation rates and geographic patterns for the following data items by region, state, and county: relationship to householder, sex, age/date of birth, Hispanic origin, race, tenure (if housing unit is occupied), and detailed vacancy status (if housing unit is vacant)?
2. What are the rates and geographic patterns of person substitution by region, state, and county?
3. What are the rates and geographic patterns of persons with more than one person-level item – i.e. relationship to householder, sex, age/date of birth, Hispanic origin, and race – allocated by region, state, and county? This is a distribution of the level of item nonresponse by person (i.e., number of persons with zero to five person-level item responses).

## **Background**

In the *2010 Decennial Census: Item Nonresponse and Imputation Assessment Report*, item nonresponse rates are described as an indicator of respondent cooperation, while characteristic imputation rates reflect respondent cooperation, but also incorporate inconsistent and unusable responses. Characteristic imputation ensures that each person and housing unit on the final census data file has valid values in all of the person and housing items. When a response is not provided or is invalid for some or all demographic data items – relationship to householder, sex, age/date of birth, Hispanic origin, race, tenure (if housing unit is occupied), or detailed vacancy status (if housing unit is vacant) – then these values must be obtained from another source or imputed.

There are three methods of characteristic imputation: assignment, allocation, and substitution. Assignment occurs when missing or inconsistent responses can be determined based on other information provided for that same person or household. Allocation occurs when missing or inconsistent responses cannot be determined based on other information provided for that same person and must be allocated from another person within the housing unit or from a person in a nearby housing unit. Substitution is a special type of allocation that occurs when all the characteristics for every person in the housing unit are missing.

For this paper, I will focus my analysis on rates of characteristic allocation and substitution. The primary difference between characteristic-level allocation and substitution is how the item-level characteristics are allocated. For person items needing allocation in individual person records

that are not filled by substitution and housing items needing allocation in individual housing unit records, the variables in the records are filled by allocating the fields one at a time from either within the household (for Hispanic origin and race only) or from a hot deck matrix (all characteristics). For characteristic-level hot deck allocation, this means that when allocating a value, one finds in the hot deck matrix specific characteristics that the housing unit or person being allocated needs and then retrieves the corresponding value (or means for deriving that value) from the matrix for those characteristics. For substitution, all of the persons in the housing unit have their person items (relationship to householder, sex, age/date of birth, Hispanic origin, and race) either allocated from administrative records or by referencing a hot deck matrix for the first six people in the household. For substitution via the hot deck, a hot deck matrix essentially copies a household with the same number of people from another housing unit to the housing unit needing substitution. For substitution households with more than six people, the remaining six-plus persons have their characteristics allocated individually through the main edit and allocation. Persons and housing units with sufficiently good reported data populate the hot decks.

## Results by Region

Table 1 contains the item allocation rates for relationship, sex, age/date of birth, Hispanic origin, race, tenure, and detailed vacancy status. Table 2 shows the overall substitution rate, and Table 3 displays the rate at which more than one person-level item was allocated per person.

The Midwest Region has the lowest allocation rates in almost every single category, including lowest individual allocation rates for relationship, sex, age/date of birth, Hispanic origin, race, and tenure, as well as the lowest substitution rates and rates at which more than one person-level item was allocated. The only category for which the Midwest Region does not have the lowest allocation rates is detailed vacancy status.

The West Region has the highest allocation rates overall when you sum all of the person and housing unit allocation rates. Particularly for race, the West Region has the highest allocation rate by 1.7 percentage points at 4.7 percent, with the Northeast Region having the next highest allocation rate at 3.0 percent. The South Region has the next highest allocation rates, with the highest individual allocation rates in Hispanic origin, tenure, and detailed vacancy status.

*Table 1: Characteristic Allocation Rates by Region*

Universe: Population not substituted in households\*, Population not substituted\*\*, Occupied housing units not substituted\*\*\*, Vacant housing units\*\*\*\*

| Geography        | Relationship* | Sex** | Age/Date of Birth** | Hispanic Origin** | Race** | Tenure*** | Vacant**** |
|------------------|---------------|-------|---------------------|-------------------|--------|-----------|------------|
| Northeast Region | 1.6           | 0.3   | 3.7                 | 3.1               | 3.0    | 3.5       | 3.2        |
| Midwest Region   | 1.4           | 0.2   | 3.0                 | 2.5               | 2.1    | 3.1       | 3.3        |

|              |     |     |     |     |     |     |     |
|--------------|-----|-----|-----|-----|-----|-----|-----|
| South Region | 1.7 | 0.3 | 3.8 | 3.3 | 2.8 | 3.7 | 4.0 |
| West Region  | 1.9 | 0.3 | 3.9 | 3.1 | 4.7 | 3.6 | 3.2 |

Source: U.S. Census Bureau, 2010 Census, Summary File 1

*Table 2: Substitution Rates by Region*

Universe: Total Population

| <b>Geography</b> | <b>Substitution</b> |
|------------------|---------------------|
| Northeast Region | 1.4                 |
| Midwest Region   | 1.3                 |
| South Region     | 2.3                 |
| West Region      | 2.0                 |

Source: U.S. Census Bureau, 2010 Census, Summary File 1

*Table 3: Rates at which more than one person-level item was allocated by Region*

Universe: Population not substituted

| <b>Geography</b> | <b>More than one item allocated</b> |
|------------------|-------------------------------------|
| Northeast Region | 8.6                                 |
| Midwest Region   | 6.6                                 |
| South Region     | 8.6                                 |
| West Region      | 9.9                                 |

Source: U.S. Census Bureau, 2010 Census, Summary File 1