

2020 Census: Count Imputation

The National Academies of Sciences, Engineering, and Medicine:
Panel to Evaluate the Quality of the 2020 Census
Date: July 21, 2021

Timothy Kennel
Assistant Division Chief, Statistical Methods

Patrick J. Cantwell
Chief, Decennial Statistical Studies Division

Outline

Housing Unit Addresses

- Count Imputation for Housing Unit Addresses
- Results from the Census
- The Procedure for the 2020 Census

Count Imputation for Group Quarters

Count Imputation for Housing Unit Addresses

Note: In the next several sections, the information refers to unresolved cases and imputation for housing unit addresses only. Analogous concepts for group quarters are discussed in the final section.

Count Imputation in the Census is a Last Resort

We use imputation as a last resort. We first urge households to respond on their own or when a census taker visits.

- We encouraged households to respond online, by phone, or by mail.
- Census taker follows up on non-responders in person.
- If a household didn't respond after one or more census taker visits, we checked to see if other high-quality data sources could provide the same information for that address.
- Otherwise, we continued to visit the household or tried to get information about the address from a neighbor or landlord.
- As a last resort, we impute missing information about the address.

Resolved and Unresolved Addresses

At the beginning of data collection, the objective is to assign a status to every address in the Census address list. Through self-response and nonresponse follow-up, we account for each address in one of two ways:

- **Resolved**: We know if the address is occupied, vacant, or delete (nonexistent), and, if occupied, how many people live there, or;
- **Unresolved**: We have insufficient or conflicting information about the address. This unresolved address goes to count imputation to be resolved.

Eligible Address Status for each Type of Imputation

Status Imputation

Non-Existent	Vacant	Size								
		1	2	3	4	5	6	7	8	9+

Occupancy Imputation

	Vacant	Size								
		1	2	3	4	5	6	7	8	9+

Household-Size Imputation

	Size								
	1	2	3	4	5	6	7	8	9+

When Is Count Imputation Implemented?

- Conducted after all data collection is complete, that is, after self-response and Nonresponse Follow-Up
- Conducted after administrative record (AR) enumeration has taken place; *addresses with “good” (high-quality) ARs have already been resolved*
- It’s an important step in creating the Census Unedited File (CUF), which is used to provide apportionment counts

Results from the Census

Shape
your future
START HERE >

United States®
Census
2020

Several Important Numbers: Census of 1970 Through 2010

Census	Self-Response Rate ¹	Percent of Addresses Unresolved	People Added as a % of Resident Population	Percent Net Undercount ²
1970	78.3%	n/a	0.44%	n/a
1980	75.0%	n/a	0.34%	−0.8%* to −1.4%*
1990	65.0%	n/a	0.02%	−1.61%*
2000	64.3%	0.55%	0.42%	0.49%*
2010	63.5%	0.38%	0.39%	0.01%
2020	Not confirmed	0.93%	Not yet released	
¹ Response rates as of the start of the Nonresponse Follow-Up operation.			[*] Statistically different from 0	
² In 1980, the Census Bureau produced 12 estimates of net undercount, ranging from −0.8% to −1.4%.				

2020 Census: Proposed Quality Metrics

- Percent of addresses unresolved (Done)
- Percent of total resident population added through count imputation
- These two percents by state

2020 Count Imputation Procedure

Shape
your future
START HERE >

United States[®]
Census
2020

2020 Count Imputation Procedure: Cell Creation and Sort

- All addresses—resolved and unresolved—are assigned to one of several mutually exclusive groups called **imputation cells**
 - *Imputation cells are defined by characteristics related to operational paradata and information from various AR sources*
- Within each state and imputation cell, addresses are **sorted** by geography

2020 Count Imputation Procedure: Nearest Neighbor

- For unresolved addresses: within same cell, impute the status or population count (or both) from its **nearest neighbor**, that is, the nearest resolved record preceding it on the file
 - *Takes advantage of spatial correlation among nearby addresses*
 - *A donor can be used more than once—no limit*
 - *2010 procedure was not nearest neighbor*

2020 Count Imputation Procedure: Example of Nearest Neighbor Imputation

- Unresolved cases (in red, below in table) at 107 and 111 Elm St.
- Based on various characteristics, they are placed in Imputation Cells 1 & 3
- In that cell, resolved nearest neighbor (in blue) is selected as donor

Address	Resolved?	Imputation Cells			
		Cell 1	Cell 2	Cell 3	Cell 4
101 Elm	Yes			X	
103 Elm	Yes	X			
105 Elm	Yes	X			
105a Elm	Yes			X	
107 Elm	No			X	
109 Elm	Yes				X
111 Elm	No	X			

Count Imputation for Group Quarters

Shape
your future
START HERE >

United States®
Census
2020

Count Imputation for Group Quarters (GQs): Considerations

GQ operations: Advance Contact before Census Day; Enumeration in summer 2020

- Some addresses classified as likely occupied, with population count of 0, including
 - Refusals
 - Reported as occupied
 - Open on Census Day, closed during visit for GQ enumeration
- Occurred across all states, various types of GQs
- Imputation? Differences between housing units and GQs: complexity of GQs

Activities to Reduce the Amount of Imputation in GQs

Operations and review by staff and subject matter experts

- Calling operation, cases with larger expected counts
- Review of universities, using data from Department of Education
- Review of counts for GQs and GQ facilities
- Matching GQs within GQ facilities to reduce duplicated GQs
- Matching people across GQs to remove duplicated people
- Distributing counts and person records within GQ facilities

Simplified Summary of GQ Count Imputation Procedure

Where available, use information available about the specific GQ

- Within type of GQ, after minor editing, apply **ratio imputation**, using
 - Expected count or maximum count from Advance Contact operation
 - Expected count or maximum count from current surveys, such as ACS
- Limited use in count imputation for GQs
 - Information from outside administrative records
 - Percentile of distribution from reported counts of resolved GQs

Questions

Shape
your future
START HERE >

United States®
Census
2020

Additional Material

Shape
your future
START HERE >

United States®
Census
2020

Count Imputation: Objective

For various reasons, including the complexity and massive size of census operations, the inability to find knowledgeable respondents, and the receipt of inconsistent information, there are records that do not have status and size information by the end of data processing. These records are referred to as missing data records.

To fill in the missing information, the Census Bureau conducts count imputation, which assigns an address status to those without one, and a household size to occupied units without one.

Characteristic Imputation is a different procedure that refers to filling in demographic characteristics (e.g. Race, Hispanic Origin, Age, Sex, Tenure). Characteristic imputation occurs after Count Imputation.

Resolved Addresses, Three Types

After data collection (enumeration) is complete, a resolved address must be one of 3 types:

- **Occupied:** an occupied, valid housing unit
- **Vacant:** a valid housing unit, with no occupants
- **Delete (nonexistent):** not a valid, livable housing unit

Address Status

Non-existent	Vacant	Size								
		1	2	3	4	5	6	7	8	9+

Three (3) Types of Count Imputation

Status imputation - Not determined if a living quarter exists at the address

- E.g., Enumerator cannot verify a living quarter exists at the address
- E.g., Conflicting information about the existence of the address
- Impute address status (nonexistent, vacant, or occupied with count)

Occupancy imputation - Determined a living quarter at the address, but not if it's occupied or vacant

- E.g., Enumerator finds the address; cannot find a respondent or proxy, or determine if someone is living there
- Impute address status (vacant or occupied with count)

Household-size imputation - Know an occupied housing unit is at the address, but not the number of people living there

- E.g., Enumerator concludes house is occupied; cannot confirm number of residents
- Impute the number of people in the unit (1 to 9)

2020 Count Imputation Procedure: Sort Procedure

- Within each state and imputation cell, sort addresses by geography: Area Census Office (ACO), county, tract, block, walking sequence
- For unresolved addresses, impute the address status from its nearest neighbor, that is, the nearest resolved record preceding it on the file
- No donor in same tract and cell? Go to bottom of list in same tract and cell
- Still no donor? Search in ACO and cell, starting with preceding tract; repeat
- Still no donor? Search in state and cell; repeat
- Still no donor? In same state, search in preceding cell. (Note: Imputation cells are ordered, so that nearby cells share similar distributions of address status, based on data from 2010 Census)

2020 Count Imputation Procedure: Characteristics Used to Create Cells

For records not from self-response, there are 86 cells, based on the following **characteristics**:

- Nearest-neighbor Household Type: occupied, vacant, delete
- Master Address File (MAF) Unit Status: valid living quarter or not
- MAF Delivery Status: likely delete, likely vacant, or other
- Spring Delivery Sequence File Status: residential or not
- NRFU Proxy Type: household member, unknown proxy, etc.
- Undeliverable as Addressed (UAA) code: “no such number,” all other, none
- AR Count: pop count of unresolved record according to ARs, 0, 1, ..., 9+

For records from self-response, 9 cells, based on number of ARs

END