

The Leadership Conference *on* Civil *and* Human Rights



CENSUS

FREQUENTLY ASKED QUESTIONS: *IMPUTATION*

Why is the census important?

The decennial census is the most inclusive civic activity in our country because it involves every person in every household and group facility. The U.S. Constitution requires an accurate count of the nation's population every 10 years as the foundation of our system of governance.

Census results are used to allocate seats in Congress; draw district lines for the U.S. House of Representatives, state legislatures, and local jurisdictions; target approximately \$1.5 trillion in federal assistance annually to states and localities; guide community decision-making affecting schools, housing, health care services, transportation, and more; and promote wise business investments. Each of these functions depends on a fair and accurate census.

What happens when the U.S. Census Bureau does not have all the answers to the census questionnaire for households?¹

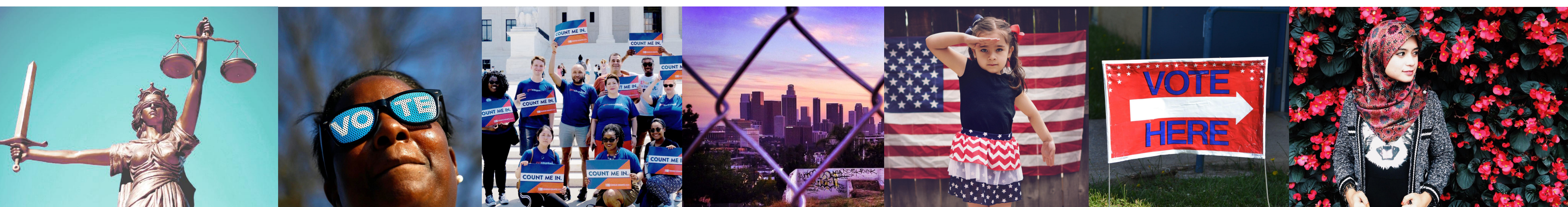
To achieve a fair and accurate census, the Census Bureau needs to know the following information about every housing unit in its address file of April 1, 2020:²

- Whether the address is a real and livable housing unit
- If it is a real and livable housing unit, whether it is vacant or occupied
- If it is occupied, how many people live there
- Demographic characteristics about people who live there, such as age, race, and sex

¹ The imputation for group facilities (known as group quarters) is discussed later in the document.

² Cantwell, Pat, "How We Complete the Census When Households or Group Quarters Don't Respond." *U.S. Census Bureau*. April 2021.

<https://www.census.gov/newsroom/blogs/random-samplings/2021/04/imputation-when-households-or-group-quarters-dont-respond.html>.





To secure data for a household address, the Census Bureau employs several techniques, which it attempts in sequential order:³

- Encouraging households to self-respond online, by phone, or by mail
- Visiting an unresponsive address to gather the information
- Using high-quality administrative records, which contain information persons in the household may have already provided to the government, such as responses to previous surveys, data from Social Security Administration files, or tax records from the IRS⁴
- Seeking information from a proxy respondent like a neighbor or landlord

After all of the above steps are exhausted, the Census Bureau turns to imputation to fill in missing data.

What is imputation?

Imputation is a statistical technique used by the Census Bureau to fill in unknown or missing data, using information that the Bureau knows about a household and its members or similarly situated households.

The Census Bureau can impute the entire count for a household (which is called “**count imputation**”), or it can impute data for unanswered questions (which is called “**characteristics imputation**”). In the 2020 Census, the Census Bureau also applied imputation to group quarters (residences where a number of unrelated people live together and an organization is providing housing or services).

Imputation is not a new technique. It is used by many other statistical agencies in the United States and around the world. Count imputation has been used to complete our decennial census since 1960. In each of the last five censuses, count imputation made up less than half of one percent of the total population count.⁵ The count imputation rate for the 2020 Census was 0.9 percent of the total population count compared to 0.4 percent in the 2010 Census.⁶ The Census Bureau will provide the characteristic imputation rates for key demographic and housing items in 2022.⁷

³ Cantwell, Pat and Kennel, Timothy, “2020 Census: Count Imputation.” Presentation at the National Academy of Sciences, Committee on National Statistics meeting. July 21, 2021.

⁴ Cantwell, Pat, “How We Complete the Census When Households or Group Quarters Don’t Respond.” *U.S. Census Bureau*. April 2021.

<https://www.census.gov/newsroom/blogs/random-samplings/2021/04/imputation-when-households-or-group-quarters-dont-respond.html>.

⁵ *Id.*

⁶ Cantwell, Pat and Kennel, Timothy, “2020 Census: Count Imputation” Presentation at the National Academy of Sciences, Committee on National Statistics meeting July 21, 2021.

⁷ Borman, Pat and Ramirez, Roberto “How We Complete the Census When Demographic and Housing Characteristics Are Missing.” *U.S. Census Bureau*. August 2021.

<https://www.census.gov/newsroom/blogs/random-samplings/2021/08/census-when-demographic-and-housing-characteristics-are-missing.html>.



How does imputation work for housing units?

Count imputation for housing units resolves whether units were occupied and how many people lived there (addresses for which occupancy and number of residents are unknown are called “unresolved” and all others are called “resolved”).⁸ Missing information is inserted about an address’s population size by copying information from its nearest neighbor (this procedure is sometimes referred to as the “hot deck” method).⁹ Persons imputed into the count are included in the census data used to apportion seats in the House of Representatives.

Characteristics imputation, which occurs after count imputation, is used to ensure that the final census file for each person, housing unit, and group quarters resident includes valid data for all questions on the census questionnaire. Unlike count imputation, characteristics imputation does not affect the total population count. But the assignment of race or other missing characteristics, such as age, can impact the population totals for subgroups.

For housing units where the Census Bureau must impute people because only occupancy status is known, all of the demographic characteristics from the donor household are imputed to the unresolved unit. But in many other cases, households respond to the census but don’t complete all of the questions (known as item non-response), or they supply an answer that doesn’t make sense (for example, the age and birthdate are inconsistent). In these cases, characteristics are imputed. There are several ways to impute characteristics:¹⁰

⁸ Prior to housing size (count) imputation, the Census Bureau conducts status imputation (when census takers do not know whether an address is a real, livable residence) and occupancy imputation (when occupancy status is unknown). Cohn, D’Vera, “Imputation: Adding People to the Census.” Pew Research Center. May 2011. <https://www.pewresearch.org/social-trends/2011/05/04/imputation-adding-people-to-the-census/>.

⁹ Data records are processed geographically, so all the addresses on a block are processed sequentially: Housing units with an unresolved status will be processed close in the processing queue to neighboring housing units that have been resolved. All addresses (resolved and unresolved) are placed in mutually exclusive groupings with some similarities. For unresolved addresses, the Bureau imputes the occupancy status, or population count (i.e., household size), or both from the “nearest neighbor” in the same grouping, so that the data imputed are based on information from a nearby household sharing some characteristics.

¹⁰ Cantwell, Pat, “How We Complete the Census When Households or Group Quarters Don’t Respond.” *U.S. Census Bureau*. April 2021. <https://www.census.gov/newsroom/blogs/random-samplings/2021/04/imputation-when-households-or-group-quarters-dont-respond.html>.



Assignment is used when answers to one or more questions are missing or inconsistent, but the missing information can be determined from (a) other responses provided for the same person; or (b) a previous census, another Census Bureau survey, or administrative records. For example, if a respondent includes their date of birth but leaves their age blank, age will be imputed based on date of birth; and, if a person did not provide their date of birth, the Census Bureau might use data the respondent provided in the previous census or in another Census Bureau survey.

Allocation is used when answers are missing or inconsistent and information can be determined from (a) responses provided for other persons in the household or (b) similar nearby households. For example, if two members of a household identify as Black and a third member of the household leaves the race question blank, the Census Bureau would impute the third person's race as Black. This is also referred to as within-household imputation. If, however, all members of a household omit their race, then the Census Bureau can impute data from the nearest neighboring household. For example, if 90 percent of the neighbors are Asian, the Census Bureau is likely to impute the missing race of a household as Asian.¹¹

How and why was count imputation conducted for group quarters in the 2020 Census?

Group quarters are typically residences where a number of unrelated people live together and an organization is providing housing or services, such as college dormitories, nursing homes, or prisons.¹² According to the Census Bureau's 2019 American Community Survey (ACS), there were a little over 8 million people living in group quarters, or 2.4 percent of the population.¹³

There are several ways information is gathered from group quarters. In some cases, the facility administrator will send information about residents to the Census Bureau. In other situations, census takers go to the facility to collect the information. A hybrid method involves census takers dropping off (and later picking up) census questionnaires, which the facility administrator distributes to and collects from residents.

¹¹ O'Hare, William P., "STANDARD DEVIATIONS: Which Census Bureau Are We Talking About?" The Census Project. *June 25, 2021*. <https://thecensusproject.org/blog/>.

¹² Cantwell, Pat, "How We Complete the Census When Households or Group Quarters Don't Respond." *U.S. Census Bureau*. April 2021. <https://www.census.gov/newsroom/blogs/random-samplings/2021/04/imputation-when-households-or-group-quarters-dont-respond.html>.

¹³ "Characteristics of the Group Quarters Population by Group Quarters Type." *U.S. Census Bureau*. 2019. <https://data.census.gov/cedsci/table?q=S2602&tid=ACSST1Y2019.S2602>.



Unfortunately, COVID-19 pandemic disruptions meant that many people who normally would have been living in group quarters on or around April 1, such as college students, were no longer there at the time of enumeration, which was scheduled for the spring but was extended by several months. In addition, pandemic-related restrictions for facilities like nursing homes and prisons created obstacles for enumeration. These conditions made group quarters data collection particularly difficult. If not corrected, the missing information could lead to an undercount.¹⁴

As a result, the Census Bureau is using count imputation for group quarters in the 2020 Census for the first time. Instead of drawing from the nearest similar neighbor, group quarters imputation uses data from a variety of sources, including information already available on the facility, such as the expected count or the maximum capacity the group quarter reported during the Census Bureau's Advance Contact Operation,¹⁵ and data from other surveys such as the ACS.^[16] Alternatively, the Census Bureau will sometimes use data from similar group quarters.

Why not just leave the count blank?

It may seem counterintuitive that imputation makes the census more accurate, but in fact, it does. The National Research Council (the research arm of the National Academy of Sciences) concluded that the 1.2 million count imputations in the 2000 Census were problematic but concluded that, without this method, the census “would have undoubtedly underestimated the true number of household resident[s]...”¹⁷ The U.S. Supreme Court has noted that the census would be “a far less accurate assessment of the population” without imputation.¹⁸

Leaving the count for an address blank because there was no response would be equivalent to assigning a count of zero. For the vast majority of housing units that have data imputed, there is evidence that the housing unit is occupied, so a population count of zero would usually be wrong. Thus, an imputed number is likely to be more accurate than zero would be. While imputation is less accurate than data directly gathered from a respondent, it is more accurate than assuming no one lives in the housing unit in question.

¹⁴ Stempowski, Deborah and King, Ryan, “2020 Census Group Quarters.” *U.S. Census Bureau*. Mar. 2021.
<https://www.census.gov/newsroom/blogs/random-samplings/2021/03/2020-census-group-quarters.html>

¹⁵ The Group Quarters Advance Contact Operation was an operation in which the Census Bureau reached out to each group facility ahead of Census Day (April 1) to verify and collect information including facility name, address, contact name, contact phone number, expected population, and preferred form of enumeration for each GQ. See “2020 Census Fact Sheet: Group Quarters Operation.” *Georgetown Center on Poverty and Inequality*. November 2019.
<http://www.georgetownpoverty.org/wp-content/uploads/2019/12/GCPI-ESOI-Correctional-Facilities-GQ-Fact-Sheet-20191210.pdf>

¹⁶ Cantwell, Pat, “How We Complete the Census When Households or Group Quarters Don’t Respond.” *U.S. Census Bureau*. April 2021.
<https://www.census.gov/newsroom/blogs/random-samplings/2021/04/imputation-when-households-or-group-quarters-dont-respond.html>

¹⁷ Citro, Constance F. et al., “The 2000 Census: Counting Under Adversity.” *National Research Council*. 2004.
<https://www.nap.edu/read/10907/chapter/1#i>

¹⁸ *Utah et. al v. Evans*, 536 U.S. 452 (2002).



Learn More!

- [How We Complete the Census When Households or Group Quarters Don't Respond](#)
- [How We Complete the Census When Demographic and Housing Characteristics Are Missing](#)
- [Imputation: Adding People to the Census](#)
- [2020 Census Group Quarters](#)
- [Evaluating the Accuracy of the Decennial Census](#)

Please direct questions about the information in this fact sheet to Meeta Anand, census consultant to The Leadership Conference Education Fund, at me@meeta.nyc.