2020 Census: Item Nonresponse, and Characteristic Imputation

National Academies of Sciences, Engineering, and Medicine
July 21, 2021

Christine Borman
2020 Decennial Census Count Review – Population Division

Shape
your future
START HERE >

United States®
Census
2020

2020 Census
Presentation Roadmap

This presentation provides an overview of:

1. Characteristic Item Nonresponse Rates

2. Edit and Characteristic Imputation Overview

3. 2020 Imputation Methodology by Characteristic

Shape
your future
START HERE >

United States®
Census
2020

2020 Census
Item Nonresponse Rates

The Item Nonresponse Rate (INR) is essentially the proportion of missing responses before pre-editing or imputation procedures for a given item.

- For example, the respondent provides the number of people living at the housing unit, but does not answer any of the demographic questions on the Census questionnaire.

- For INR, missing values are included in the rates, but inconsistent responses (i.e., incompatible with other responses) are considered non-missing responses.

The Bureau plans to release the INR Assessment Report in 2022.

Shape
your future
START HERE >

United States®
Census
2020

2020 Census
Purpose of Edit and Characteristic Imputation

Edits are used to ensure certain consistencies among characteristics.

Characteristic imputation is used to ensure that each person, housing unit and group quarter unit on the final census file has valid values for all items.

Edit and Characteristic Imputation

Has been used in some form since the 1960's

Utilizes administrative records data to improve quality of imputations

Occurs after Count Imputation and does not change the population count

Shape your future START HERE >

United States®
Census
2020
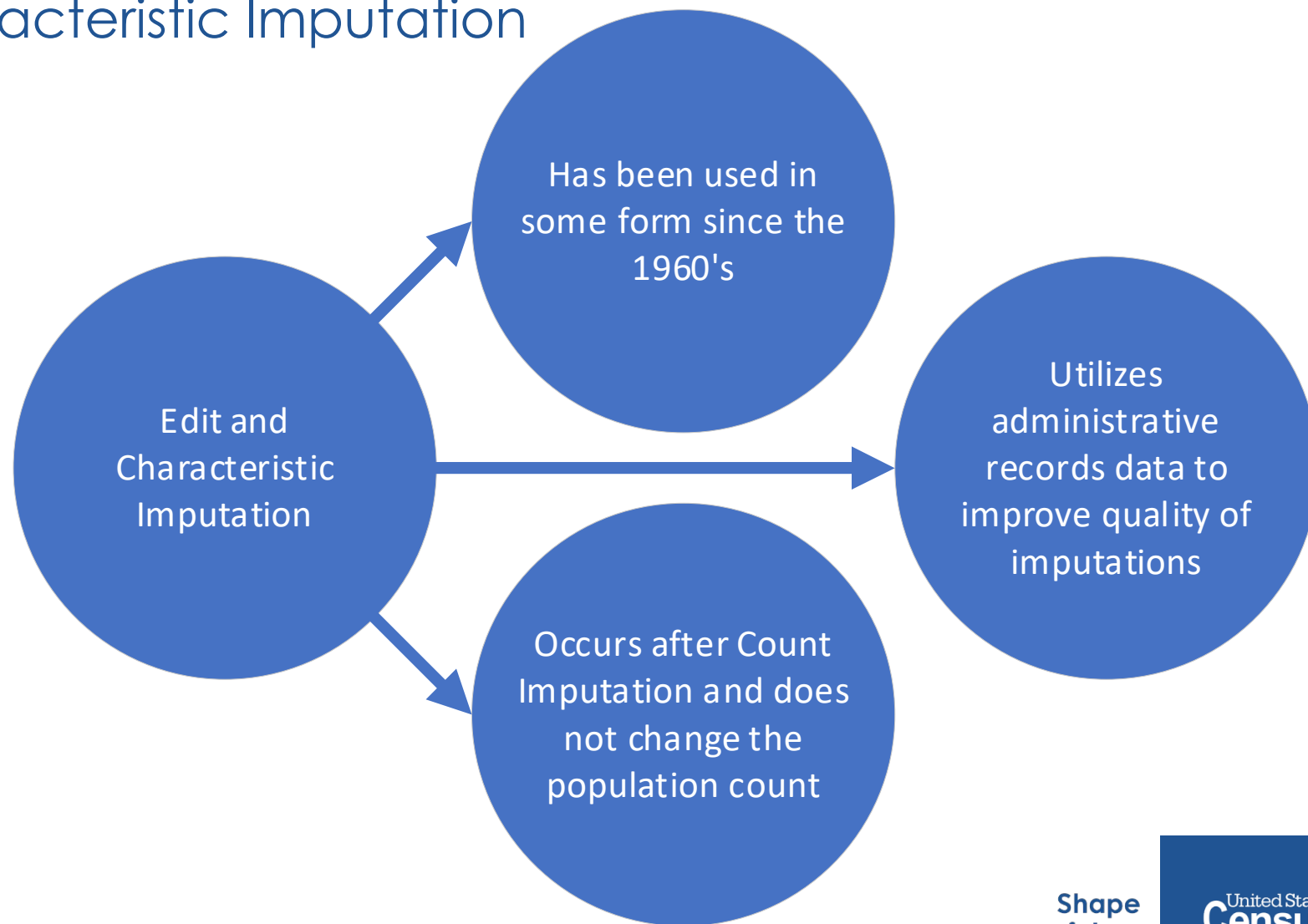
2020 Edit and Characteristic Imputation Overview

Edits

1. During the editing phase, responses are run through a series of checks to:

a) Detect and correct out-of-range or inconsistent values

b) Remove invalid or duplicate responses

c) Convert date of birth values to age values and race/Hispanic origin responses to a numeric code

The best case scenario is that all responses are valid and consistent – we call those "reported" and no imputation is needed. When that doesn't happen, we use characteristic imputation.

Characteristic Imputation

2. Assignment – responses are missing or inconsistent and information can be determined from:

a) Other responses provided for that same person

b) Previous Census or Administrative Records (AR)

3. Allocation – responses are missing or inconsistent and information can be determined from:

a) Responses provided for other persons in the household

b) Similar nearby households using "nearest neighbor hot decks"

Shape
your future
START HERE >

United States®
Census
2020

2020 Census
What if sex is missing, invalid, or inconsistent?

What is Person 1's sex? *Mark* ☒ *ONE box.*
☐ Male   ☐ Female

Some respondents choose not to provide sex information or provide invalid responses to the sex question (both male and female are selected). When this happens, we:

1. Use first name to try to fill in missing sex

2. Link the person to high-quality administrative records, such as their 2010 Census response or information from the Social Security Administration (i.e. the Census Numident)

3. Assign sex to maintain household consistency

4. Allocate from the nearest neighbor hot deck

For names like Emma or Isabella that are highly reported as female, we can assign sex based on first name.

John Doe was matched to his 2010 Census record, where he reported that he was male.

If sex is missing for the householder's opposite-sex spouse, we assign the sex that fits with that response.

Shape your future START HERE >

United States® Census 2020

2020 Census
What if age/date of birth (DOB) are missing, invalid, or inconsistent?

What is this person's age and what is this person's date of birth? *For babies less than 1 year old, do not write the age in months. Write 0 as the age.*

Print numbers in boxes.

Age on April 1, 2020 ☐☐☐ years | Month ☐☐ | Day ☐☐ | Year of birth ☐☐☐☐

Some respondents choose not to provide age and DOB information or provide invalid or inconsistent age and DOB. Invalid would be an age of 299 years old; we know that no one has lived that long (yet)! Inconsistent would be when the age calculated based on the provided DOB does not match the age provided. When this happens, we:

1. Calculate age from DOB and visa versa by creating a random DOB if only age is provided

2. If age and DOB are inconsistent, choose the value that is consistent with the person's relationship to the householder

3. Assign age/DOB from the 2010 Census or the Census Numident

4. Allocate from the nearest neighbor hot deck

The respondent only provided the person's DOB, but not their age. That's okay! We can calculate it.

If age and DOB are inconsistent for a person in group quarters, we choose the value that is closest to the median age for the group quarters unit.

Going to the hot deck is always our last resort. When we do, we usually allocate an age difference using relationship.

Shape your future START HERE >

United States® Census 2020

2020 Census
What if Hispanic Origin and Race are missing or invalid?

Some respondents choose not to provide Hispanic origin or race information, or provide responses that have to be edited to a single Hispanic origin response and up to eight race responses. Note – we do not impute detailed Hispanic origin and race responses if only a major group is reported. When Hispanic origin and race responses are missing, we:

1. Assign Hispanic origin and race from the 2010 Census, American Community Survey, Census Numident, or other federal sources

2. Assign race from Hispanic origin and visa versa

3. Allocate from another household member's response

4. Allocate from the nearest neighbor hot deck

It is important to note that we have made a lot of improvements to our race and Hispanic origin coding over the decade. We've also added write-in lines for the White and Black race categories.

The respondent wrote into the Some other race field that they are Hispanic. We can use that info to fill in their missing Hispanic origin value.

The householder marked the "Korean" checkbox but didn't provide a race for their child. We can allocate the child's race to Korean.

8. Is Person 1 of Hispanic, Latino, or Spanish origin?
Mark ☒ one or more boxes AND print origins.

☐ **No,** not of Hispanic, Latino, or Spanish origin

☐ **Yes, Mexican,** Mexican Am., Chicano

☐ **Yes, Puerto Rican**

☐ **Yes, Cuban**

☐ **Yes, another Hispanic, Latino, or Spanish origin** — Print, for example, Salvadoran, Dominican, Colombian, Guatemalan, Spaniard, Ecuadorian, etc. ⬆

9. What is Person 1's race?
Mark ☒ one or more boxes AND print origins.

☐ White – Print, for example, German, Irish, English, Italian, Lebanese, Egyptian, etc. ⬆

☐ Black or African Am. – Print, for example, African American, Jamaican, Haitian, Nigerian, Ethiopian, Somali, etc. ⬆

☐ American Indian or Alaska Native – Print name of enrolled or principal tribe(s), for example, Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Nome Eskimo Community, etc. ⬆

☐ Chinese ☐ Vietnamese ☐ Native Hawaiian
☐ Filipino ☐ Korean ☐ Samoan
☐ Asian Indian ☐ Japanese ☐ Chamorro
☐ Other Asian – Print, for example, Pakistani, Cambodian, Hmong, etc. ⬆ ☐ Other Pacific Islander – Print, for example, Tongan, Fijian, Marshallese, etc. ⬆

☐ Some other race – Print race or origin. ⬆

Shape your future START HERE >

United States® Census 2020

2020 Census
What if Relationship is missing, invalid, or inconsistent?

Some respondents choose not to provide relationship information or provide invalid or inconsistent responses. Invalid would be multiple relationship categories selected. Inconsistent would be when the relationship status does not match the age or sex of other members in the housing unit. When this happens, we:

1. Assign or allocate relationship to maintain consistency with other household members

2. Assign relationship from the Census Bureau's Kidlink file, which contains Social Security Administration information and links parent-child relationships

3. Allocate from nearest neighbor hot deck

How is this person related to Person 1? *Mark* ☒ *ONE box.*

- ☐ Opposite-sex husband/wife/spouse
- ☐ Opposite-sex unmarried partner
- ☐ Same-sex husband/wife/spouse
- ☐ Same-sex unmarried partner
- ☐ Biological son or daughter
- ☐ Adopted son or daughter
- ☐ Stepson or stepdaughter
- ☐ Brother or sister
- ☐ Father or mother
- ☐ Grandchild
- ☐ Parent-in-law
- ☐ Son-in-law or daughter-in-law
- ☐ Other relative
- ☐ Roommate or housemate
- ☐ Foster child
- ☐ Other nonrelative

We've changed the relationship categories a little between 2010 and 2020. We now distinguish spouse and unmarried partner categories by same-sex and opposite-sex.

If the householder is younger than who they report as their child, they probably meant to respond that they are the child and the other person is their parent. We'll switch the relationship category to match.

Shape your future START HERE >

United States® Census 2020

2020 Census
What if Tenure status (owned versus rented) is missing?

When tenure status is missing or invalid (multiple values selected), we:

1. Assign tenure based on a set of rules using federal administrative records sources, such as information from Housing and Urban Development (HUD), and commercial tax and deed information

2. Allocate from nearest neighbor hot deck with administrative records values used as a covariate

3. Allocate from nearest neighbor hot deck

Is this house, apartment, or mobile home — *Mark* **X** *ONE box.*

☐ Owned by you or someone in this household with a mortgage or loan? *Include home equity loans.*

☐ Owned by you or someone in this household free and clear (without a mortgage or loan)?

☐ Rented?

☐ Occupied without payment of rent?

Housing units that receive rental assistance through public housing or other federal programs are assigned to be renter-occupied units.

Shape your future START HERE >

United States® Census 2020

2020 Census
What if Detailed Vacancy Status is missing?

Detailed vacancy status is typically collected by enumerators during the Nonresponse Followup operation when a housing unit is identified as vacant, but housing units can also be identified as vacant using high-quality administrative records and through a method called Count Imputation when no status is provided for a housing unit.

When detailed vacancy status is not completed by the enumerator or a housing unit is determined to be vacant by another method, there is only one way for the Census Bureau to collect this detailed status:

1. Allocate from nearest neighbor hot deck

Vacancy question: What is the primary reason why no one will be living or staying at <ADDRESS> on April 1, 2020?

The unit will be – *For rent; Rented, not occupied; For sale only; Sold, not occupied; For seasonal, recreational, or occasional use; For migrant workers; Other*

Missing detailed vacancy status can only be filled through allocation; there are currently no administrative records or assignment rules that we use to determine vacancy status otherwise.

Shape your future START HERE >

United States® Census 2020

2020 Census
What if all information is missing for all persons in a housing unit?

Sometimes, we do not have any information for the people living in an occupied housing unit. This occurs when respondents or enumerators only provide a population count (also referred to as "pop count only" cases) and when the housing unit status is count imputed. To resolve this, we:

1. Allocate characteristics from administrative records persons associated with the housing unit when administrative records are available

2. Allocate entire neighboring households from nearest neighbor hot deck

When we allocate all demographics from every person in a housing unit, we call this "substitution."

Shape
your future
START HERE >

United States®
Census
2020

2020 Census
2010 Imputation Rates and Expectations for 2020

For 2020, item nonresponse rates were higher than in the 2010 Census, which resulted in higher imputation rates. We have carefully reviewed the Census Edited File to ensure that our edits worked as expected and the data are reasonable.

Overall Item Nonresponse and Imputation Rates

	Person-Level Items					Household -Level Item
	Relationship	Sex	Age/Date of Birth	Hispanic Origin	Race	Tenure
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					

Source: 2010 Decennial Census: Item Nonresponse and Imputation Assessment Report

Shape
your future
START HERE >

United States®
Census
2020

Questions?