

Exhibit B

How We Unduplicated Responses in the 2020 Census

CB [census.gov/newsroom/blogs/random-samplings/2021/04/how_we_unduplicated.html](https://www.census.gov/newsroom/blogs/random-samplings/2021/04/how_we_unduplicated.html)

Andrew Keller, Mathematical Statistician, Decennial Statistical Studies Division, and Ryan King, Mathematical Statistician, Decennial Statistical Studies Division

April 22, 2021

Written by:

Andrew Keller, Mathematical Statistician, Decennial Statistical Studies Division, and Ryan King, Mathematical Statistician, Decennial Statistical Studies Division

At the U.S. Census Bureau, we often say our goal is to count everyone once, only once, and in the right place. Sometimes in an effort to count everyone in a census, we end up counting some people more than once. The Census Bureau refers to a person counted more than once as a “duplicate.”

Today, we’ll talk through situations where that can happen and how we resolve duplicates in the 2020 Census.

Reasons for Duplication

There are several reasons for duplicates in a census:

- We receive more than one response for an address.
- People are counted in more than one place because of potentially complex living situations.
- There might be an issue with the address — a housing unit is on our address list more than once or census materials are misdelivered.

We use a special algorithm to resolve the first situation and a series of steps to resolve the second and third.

More Than One Response for an Address

We might receive more than one response for an address if, for example, a roommate or spouse responds to the census without realizing another member of the household has already responded.

For the 2020 Census, we allowed households to respond online or by phone with or without their Census ID — a unique 12-digit number that links the household’s response to our address list. (The paper invitations and questionnaires had the Census ID pre-printed.)

Allowing responses without an ID made it even easier for households to respond, but it also made it easier for more than one person to respond for the household.

We've developed sophisticated procedures that take these situations into account and build upon our decades of census and survey-taking experience. Each census, we use what we call the "Primary Selection Algorithm" (the details of which are protected for quality assurance reasons) to determine whom to count when we receive more than one response for a single address.

Complex Living Situations

Sometimes people are initially counted in more than one place because of the complexity of their living situation.

A few examples include:

- College students counted at both their college residence and at their parents' home.
- Children counted by both divorced parents who share custody.
- People with more than one residence, such as a seasonal or vacation home.
- Workers who live near their workplace during the week and commute to a different residence for the weekend.
- People who moved and were counted at both their old and new addresses.

These situations are trickier to untangle, and we must rely on people providing us information about their living situations to do so. From there, we determine where to count them using what we call the "[residence criteria](#)," which are based on a longstanding principle set by Congress to count people at their usual residence, which is where they live and sleep most of the time.

To help sort out people's living situations, the 2020 Census included a question that asked, "Does this person usually live or stay somewhere else?" If "yes," people could select among multiple options to indicate the reason.

2. Does this person usually live or stay somewhere else?

Mark ☒ all that apply.

☐

No

☐

Yes, for college

☐

Yes, for a military assignment

☐

Yes, for a job or business

☐

Yes, in a nursing home

☐

Yes, with a parent or other relative

☐

Yes, at a seasonal or second residence

☐

Yes, in a jail or prison

☐

Yes, for another reason

We then had a special operation called Coverage Improvement to call a subset of households that responded “yes.”

From those phone interviews, we try to determine:

- **Where the person lives and sleeps most of the time.** This determines where they should be counted in the census.
- **The other address where the person lives or stays.** This helps us determine whether their information was duplicated and resolve that duplication.

Sometimes when we initially counted people living in group quarters (places such as college dorms, prisons and nursing homes), the facility would provide an address for where the person stayed when they were not at the group quarters. (More information about how we count group quarters is available in the recent [2020 Census Group Quarters](#) blog.)

During data processing, we used the alternate address in conjunction with the residence criteria to resolve instances when an individual was initially counted in both places.

The process described above was not enough to resolve all duplicated people for this subset of the population. For example, some households didn’t cooperate with the follow-up phone interviews, and some group quarters didn’t provide alternate addresses for their residents.

To further resolve duplication:

- We used statistical matching. This process uses software to match names and characteristics like sex and birthdate to look for and link individuals who were possibly duplicated. We attempted to match the people from the Coverage Improvement operation and the group quarters population to all other people in the census.
- If the person was found at two addresses, we applied the residence criteria to remove the duplicate enumeration.

Issues With the Address

Sometimes, duplicates occur because of an issue with the address, such as:

- **We may have two addresses on our list that refer to the same housing unit.** This occasionally happens if an address update we receive from the U.S. Postal Service or state and local governments is different enough in spelling or formatting that it’s not clear the address is the same as one already on our list.

- **A mail carrier or a census worker may have delivered the household's census invitation to the wrong address.** Imagine that the census invitations for apartments A and B get switched in delivery. If only A responds (using B's address), a census taker follows up with A because their original census mailing was delivered to B's address, and B, whose form has A's address, didn't respond. This actually results in collecting A's information twice because of the mix up in delivery — once at B's address (which was identified using the pre-printed Census ID) and once at A's address through the census taker interview.

We relied on statistical matching that considered geographic distance to identify and resolve these situations.

Our research suggests that if we find duplicates within a limited area, such as the same block, duplicates are more likely an issue of the address being duplicated than an issue related to the living situation. We used enhanced address matching methods for addresses for which the people were linked to identify and remove duplicated addresses in these select areas.

Taking Additional Steps to Resolve Duplicates

As we tally census results, we compare these results to other benchmark data, as discussed in the recent [2020 Census Data Review](#) blog. From our review, we could tell that more duplicates remained in the 2020 Census, even after taking all the steps we describe above.

Consequently, after data collection was complete, we took the following steps to identify and remove additional duplicated individuals from the census:

1. **We did a broad search for duplicates within each state and the District of Columbia.**

We checked for links between people in:

- Housing units and group quarters.
- Across all housing units.
- Across group quarters of the same type.

2. We determined where to count the duplicated people.

- If we found individuals who were initially counted at both a group quarters and a housing unit, we followed the residence criteria and kept them at one address and removed them from the other.
- If we found individuals who were initially counted at two addresses that really referred to the same housing unit, we picked one address based on its attributes and removed the other address.
- If we found individuals who were initially counted at two different housing units, we first checked if there was any indication from census responses that one of the units was vacant or nonexistent. If so, we removed the duplicated people from the vacant or nonexistent unit and kept them at the other address.
- If neither housing unit appeared vacant or nonexistent, we checked whether the people were only associated with one of the addresses in administrative records. If so, we kept them at that address.

3. We imputed the status for addresses that remained after removing duplicated people.

For some addresses, removing duplicates meant removing everyone. In those cases, we used a statistical technique called imputation to determine whether the remaining address was occupied, vacant or nonexistent, and if occupied, how many people lived there. We took into account information already available, including from administrative records and from the post office, about the delivery of census materials to help fill in the missing information.

An earlier blog, [How We Complete the Census When Households or Group Quarters Don't Respond](#), discusses imputation in more detail.

In summary, we used long-established procedures to unduplicate multiple 2020 Census responses for the same address. We also worked to resolve duplication when individuals were enumerated at two different addresses by following up with households, using statistical matching techniques, and examining potentially duplicated addresses.

These extensive steps enable us to get closer to our goal of counting everyone once, only once, and in the right place.

This article was filed under:

[Data Quality \(Missing Data, Edit, and Imputation\)](#)