

2020 Decennial Census – Apportionment Data Processing

Original timeframe: 7/31-12/31/20 (5 months); Post COVID extension: 10/31/20 – 4/30/21 (6 months)

The Census Bureau's process to generate apportionment data is more than simply tallying real time responses and aggregating them into state-wide totals. The time spent on data processing is essential to ensuring an accurate and complete count.

- Data processing activities can only begin once incoming data collection concludes. Running these operations concurrently would result in duplicative work and ultimately delay the issuance of final apportionment data. The Bureau needs to work from a static dataset in order to assess its validity, correct discrepancies where necessary, and use administrative records and statistical methods to complete the count if field data collection could not gather the data.
- Data processing activities must occur sequentially because of the data's interdependencies. For example, collected responses must be validated to determine which households are still outstanding and can be enumerated with administrative records or other statistical methods.
- The three month delay in the largest field data collection operations, which impacted more than 35 percent of all responding households, will require additional data processing to ensure people are accurately counted in the correct location as of the April 1st reference date, which accounts for the difference in the durations of the two timeframes.
- All Decennial products are aggregates of household level data. As a result, it is imperative that the bureau takes steps to ensure both the accuracy of the count and the geographic location; even if the first product, the delivery of apportionment counts, is only presented at the state level.
- At each step of the process, we conduct and document rigorous quality assurance to avoid undercounts of specific groups, including evaluating the data at multiple levels of geography against benchmark data to ensure demographic reasonableness of data.
- Actions that would condense or remove parts of this process run the risk of:
 - Incorrect geographic placement of housing units or missing units that were added through peak field operations.
 - Duplicative or conflicting data for certain households.
 - Unreliable characteristic data for redistricting files.
 - Additional legal challenges of apportionment counts, redistricting results, or other data products as a result of diminished quality of decennial data.

The specific data processing activities conducted after data collection are:

- De-duplicating data from households who may have inadvertently responded multiple times in order to avoid “double counting.” This includes reviewing instances where a household provided different information across multiple responses, and determining what information should be counted as part of the official record.
- Merging and standardizing the format of data received from different sources, including mail-in, phone, or internet self-responses; nonresponse followup information provided to an enumerator; and “special operations” such as group quarters. This includes clerical coding of write-in responses.
- Ensuring that responses from households that responded without the unique identifier or who responded from a previously unidentified address are assigned to the correct location.
- Performing statistical techniques to account for missing housing unit status and occupancy information.
- Processing the counts for military and civilian personnel and their dependents living overseas as part of the Federally Affiliated Count Overseas operation.
- Incorporating new quality assurance processes to adjust for temporary geographic shifts due to the pandemic (such as students coming home early from college) and assure consistent residence criteria are correctly applied.
- Tabulating all of the above data and calculating population data to apportion the Congressional seats and delivering these data to the Secretary of Commerce for transmittal to the President.