

## 2020 Census

### The Bureau Concluded Field Work but Uncertainty about Data Quality, Accuracy, and Protection Remains

*The Census Bureau has completed data collection operations for the 2020 Decennial Census. As the Bureau begins processing responses to deliver data for apportionment and redistricting, it will need to ensure the quality, accuracy, and protection of the data collected.*

*In recent years, GAO has identified challenges to the Bureau's ability to conduct a cost-effective count of the nation, including new innovations, acquisition and development of IT systems, and other challenges. In 2017, these challenges led GAO to place the 2020 Census on its High-Risk List.*

*Since 2007, GAO has made 113 recommendations specific to the 2020 Census. As of December 2020, 20 of the recommendations had not been fully implemented.*

GAO was asked to provide regular updates on the 2020 Census. This report examines the cost and progress of key 2020 Census operations critical to a cost-effective enumeration, and early warnings that may require Census Bureau or congressional attention.

The Bureau provided technical comments that were incorporated as appropriate.

This report is the fifth in a series of updates on the Census Bureau's (Bureau) 2020 Census activities and operations. This update includes information from GAO's ongoing work on the conclusion of 2020 Census field operations, selected potential indicators of quality, and changes made to response processing operations as the Bureau produces its data products.

The Bureau changed the dates for completing the nonresponse follow-up (NRFU) operation and delivering data for apportionment several times between August 2020 and October 2020 in response to litigation. After receiving a ruling from the U.S. Supreme Court on October 13 that allowed it to stop data collection, the Bureau announced it would conclude NRFU on October 15 and deliver data for apportionment on or shortly after December 31.

At the conclusion of its data collection operations on October 15, the Bureau reported it had achieved a national enumeration rate of 99.98 percent of housing units. The Bureau reported 67.0 percent of enumerations came from self-response via internet, paper, or phone, and an additional 32.9 percent of households were enumerated during NRFU.

#### Data Collection Operations in the Field Have Ended, but Data Quality, Accuracy, and Protection Remain Uncertain

##### Alternative Data Collection Methods

When the Bureau cannot obtain census information directly from household members, either through self-response or a completed NRFU interview, it relies on alternative methods. The Bureau's reliance on these methods may provide insight into the quality of data collected:

- **Proxy responses.** The Bureau used proxy responses—information from a neighbor or other knowledgeable person, such as a landlord or building manager, about a household—to collect data on 24.1 percent (approximately 7.4 million, based on preliminary results) of occupied households in the NRFU workload, compared to 23.8 percent (approximately 6.8 million) in 2010. Proxy responses are generally lower quality than responses directly from a household.
- **Partial responses.** The Bureau may receive a partial response for a household through self-response or a NRFU interview. For some cases, enumerators in the field are directed to obtain, at a minimum, the status of whether the household is occupied, vacant, or not a household, and the number of people in the housing unit. The number of responses with this minimal amount of data can be an indicator of the quality of data collected. The Bureau has not yet calculated the number of partial responses it received, but plans to report on it in future operational assessments.
- **Administrative records.** The 2020 Census incorporated increased use of administrative records into its design, a major cost saving innovation. Use of these records leverages information people have already provided to the federal or state government, such as the Internal Revenue Service or prior census data. The Bureau used administrative records to resolve approximately 14 percent of households (about 8.4 million) in the NRFU workload, which was less than planned. However, the Bureau decided to use

administrative records that lack corroboration by a second source after NRFU began, introducing a data quality risk.

- **Imputation.** This statistical method draws on data from other household members, nearby households, and data on that household from past censuses and administrative records. The Bureau uses imputation to create records for housing units that appear occupied, but for which no other information is available. It has been used in some form since the 1940 Census.

#### Difficulty Completing NRFU in Some Local Areas

The census is a local endeavor and the Bureau experienced challenges completing NRFU in some local areas including difficulty hiring enumerators and accessing rural and tribal areas. Other challenges included high rates of COVID-19 and natural disasters.

To address these challenges, the Bureau instituted financial awards for enumerators who maximized hours and completed a set number of cases per hour. The Bureau also enumerated some areas by phone and used travel teams of enumerators, offering financial awards for those willing to travel to certain areas.

When the Bureau left the field on October 15, 10 of the Bureau's 248 area census offices fell short of completing 99 percent of their NRFU workload, one of the Bureau's stated indicators of completion.

#### Less Time to Ensure Accuracy during Response Processing Operations

To deliver data for apportionment on December 31, the Bureau will have only 77 days to complete response processing, an operation that was designed to take 153 days. To complete response processing in fewer days, the Bureau made changes to its process, including locking down its Master Address File prior to the end of data collection and shortening the amount of time for reviews by subject matter experts in the Bureau's statistics divisions.

The Bureau is also prioritizing tasks needed to produce apportionment counts rather than simultaneously preparing redistricting data, which involves more data elements. In doing this, the Bureau will create two separate analyses for the separate output files, which differs from its plan to produce a single analysis to support both output files.

In compressing its response processing, the Bureau also faces increased risk that system defects or other information technology issues may go undetected, affecting the quality and accuracy of the count. Additionally, the Bureau will have less time to address issues that arise.

#### Work Remains to Protect Data Privacy

The Bureau reported progress in implementing a disclosure avoidance technique to protect the confidentiality of its respondents' data in its publicly-released statistical products. However, the Bureau still has work remaining before it implements disclosure avoidance on its data products, and final decisions regarding the implementation have yet to be made.