

THE IMPORTANCE OF DATA PROCESSING & QUALITY ASSURANCE IN THE CENSUS:

Why the census isn't over when the counting stops

The census doesn't end when counting operations stop. Once the Census Bureau finishes collecting data — from 150 million housing units and thousands of group facilities, from people experiencing homelessness or living in transitory locations (e.g. RV parks) — it begins a painstaking, complex, and highly specialized series of activities to process and improve the accuracy of the raw data.

Under the pre-pandemic 2020 Census timeline, that work would have taken about five months. The Census Bureau (with the administration's support) requested six months for this phase in its COVID-adjusted operational plan,ⁱ in recognition of the greater challenges to collecting accurate data in the field caused by pandemic-related delays. The adjusted timeline set an April 30, 2021 deadline for reporting apportionment data to the president and July 31, 2021 deadline for sending redistricting data to the states. Then, without explanation in late July, Secretary Ross asked bureau officials to develop an accelerated timeline that essentially forced the Census Bureau to rush remaining counting operations and compress data processing and quality assurance activities into less than three months. Activities designed to take 153 days, before a pandemic derailed the original 2020 Census plan, would have to be done in 92 days in order to meet the current Dec. 31, 2020 apportionment deadline.ⁱⁱ *And that smaller window, based on ending data collection on Sept. 30, has now shrunk further ... and the clock is ticking.*

The shortened timeline requires the bureau to “streamline” many components of post data collection activities. **According to the Commerce Department Inspector General, “The streamlined data processing under the accelerated census plan poses a myriad of risks to accuracy and completeness.”**

Why data processing is more than simply adding up the numbers

Once data collection ends, a lot of work remains to make sure every person is counted once, only once, and in the right place (as the census motto goes). Each step must be done thoroughly before moving on to the next, in case errors in processing require reruns that would be difficult to fix later. IT system software is programmed to run each step in a certain order. Here are some of the activities and challenges:

- ➔ **Scan paper questionnaires and check for incomplete responses (blank or missing data); merge responses from different sources (on-line, phone, paper forms, Nonresponse Follow-up operation); process responses not tied to a census ID number.**
- ➔ **Remove duplicate enumerations for the same household or for people counted at two different addresses, to prevent an overcount** There were an estimated 8.5 million duplicate counts in the 2010 Census.ⁱⁱⁱ *Disruptions and delays to every 2020 Census operation make it likely that there are even more duplicates this year.* Historically, duplications are disproportionately non-Hispanic White, older, and homeowners.
 - College students who went home just as the census was starting last March might have been included on their parents' form mistakenly. They also would be counted in the Group Quarters operation if they lived in college housing, or at their off-campus address.
 - People who own two homes might have fled COVID hotspots last spring and be counted at both residences. The bureau must count them at the home where they live most of the year.
 - People who self-responded and then moved after Census Day might have been counted by an enumerator at their new residence during the delayed door-knocking operation.
- ➔ **Resolve inconsistent responses; identify and resolve fraudulent self-responses.**
- ➔ **“Enumerate” an estimated 6.2+ million households using federal administrative records** (after no self-response and one unsuccessful in-person visit). The “administrative record linkage” for enumeration purposes is new for the 2020 Census.

- **Fill in missing information, either for characteristics (age, race, gender, etc.) or for entire households lacking any data except occupancy status.** This vital work is done using several methods, including administrative records from other government agencies, data collected in other Census Bureau surveys, and statistical imputation. There were 6 million “whole person imputations” in the 2010 Census. (If the bureau must rush to meet a 12/31 apportionment deadline, it would conduct count imputation first, to produce the apportionment data, and resolve inconsistencies and impute characteristics data to produce the redistricting data files afterwards.)
- **Process data from group facilities (“Group Quarters,” e.g. college dorms, nursing homes, prisons, military barracks).** Due to pandemic-related disruptions, student data from higher education institutions, in particular, are likely to be incomplete, making processing more difficult.
- **Share preliminary Group Quarters counts with state demographers,** who review for missing or incorrectly located facilities and anomalies such as improbable facility counts. The shortened timeline has derailed this part of the Count Review Operation, which helped the Census Bureau identify and fix mistakes before final tabulations in previous censuses.
- **Run preliminary tabulations. Trained demographers review (multiple times)** to see if the numbers are much higher or lower than independent estimates of the population.
- **Add data on servicemembers and federal civilian employees stationed abroad (and dependents living with them) to the state population totals used for congressional apportionment.**
- **Produce total population counts for states. Compute the congressional apportionment using the Method of Equal Proportions.** Historically, the Census Bureau has done this as a “courtesy” for the president, whose role in apportionment has been essentially ministerial since enactment of the “automatic” apportionment law in 1929.

What the experts say about short-changing data processing and quality assurance activities

Commerce Department Inspector General

- “[T]he accelerated schedule increases the risks to the accuracy of the 2020 Census. This was the consensus view of the senior Bureau officials we interviewed. **The accelerated replan increases risks to both phases of the 2020 Census—i.e., data collection and data processing.**” (*“The Acceleration of the Census Schedule Increases the Risk to a Complete and Accurate 2020 Census,”* Final Management Alert No. OIG-20-050-M, September 18, 2020, pg. 8, *emphasis added*)
- “[I]f you can’t complete the data collection – the input to the census until the end of October, you can’t deliver those apportionment counts by the legislative requirement of December 31st. You – you can’t do it. There’s not enough time in ... November and December to put out a quality product with all of the backend processing that has to happen.” (Senior Census Bureau official quoted in Final Management Alert, pg. 7.)
- “Bureau leaders continued to believe that the statutory extension was preferable, and would give the Bureau the best chance to create a high-quality, usable census. A statutory extension would permit the Bureau to adhere, as closely as practicable, to the 2020 Census plan it developed over a decade instead of the replan it developed over a weekend.” (Final Management Alert, pg. 8)

Government Accountability Office

- “The Bureau acknowledges both delays and the compressed timeframes as a risk to data quality.” (GAO testimony before the House Committee on Oversight and Reform, 9/10/20, pg. 9)
- “[T]he Bureau is behind schedule in completing integration testing of the 12 systems needed to conduct the response processing operation. Specifically, the Bureau did not complete system integration testing by mid-June 2020 as originally planned, and now expects to complete this testing by the beginning of October 2020—approximately 3.5 months later than planned. The Bureau is also concurrently conducting operational testing, which is intended to ensure that all components are ready to conduct the operation, and will need to complete this testing before the response processing operation begins.” (*“Recent Decision to Compress Census Timeframes Poses Additional Risks to an Accurate Count,”* GAO-20-671R, pg. 16)

John Thompson, Former Census Director (2013-17) and Associate Director for the 2000 Decennial Census
Sworn expert declaration in *LUPE v. Trump*, Civil Action No. 8:19-CV-02710-PX, Oct. 3, 2020

- “The compressed schedule for post data collection processing significantly increases the risk of inaccuracies into the 2020 Census data.” (pg. 12)
- “[T]he Census Bureau is cutting 21 days of expert review and computer error remediation from the schedule. ...The 21 days were determined to be essential to detecting and repairing computer errors. [T]he incidence of an undetected computer error in the programs that will be used to produce the Apportionment counts is virtually assured. The fitness of use of the resulting data for Apportionment is in my opinion questionable.” (pg. 13)

Howard Hogan, Associate Director for Demographic Programs, U.S. Census Bureau

- “[T]he compressed schedule for post data collection processing carries a grave risk of a greatly increased differential undercount.” (Sworn expert declaration in *LUPE v. Trump*, Civil Action No. 8:19-CV-02710-PX, Sept. 1, 2020)
- “The Census Bureau has been forced into a situation of modification, revision and replanning. ...The Census Bureau has not been allowed time to carefully assess and test the programs in this modified and “streamlined” process. The result is a significant chance of a catastrophic failure for Census 2020.” Sworn expert declaration in *LUPE v. Trump*, Civil Action No. 8:19-CV-02710-PX, Oct. 3, 2020, pg. 14)

Please direct questions about the information in this memo to Terri Ann Lowenthal, freelance census consultant, at terriannlowenthal@gmail.com.

ⁱIssued April 13, 2020.

ⁱⁱ 13 U.S.C. §141(b)

ⁱⁱⁱ DSSD 2010 Census Coverage Measurement Memorandum Series #2010-G-01, May 22, 2012