Congressional Research Service
Informing the legislative debate since 1914

The 2020 Decennial Census: Assessments to Date of Census Quality

Updated March 24, 2022

Congressional Research Service
https://crsreports.congress.gov
R47046

CRS REPORT
Prepared for Members and
Committees of Congress

Congressional Research Service
Informing the legislative debate since 1914

SUMMARY

R47046

March 24, 2022

Jennifer D. Williams
Specialist in American National Government

The 2020 Decennial Census: Assessments to Date of Census Quality

The 2020 decennial census was innovative in several ways. New technology was developed to make updating the Census Bureau's Master Address File, used for contacting the public, simpler and less expensive, and to streamline field operations for collecting data from those who did not answer the census initially. For the first time, people could answer online as well as on paper census forms or by telephone. The Census Bureau spent more than $500 million, partly for extensive advertising, to reach and inform all U.S. residents about the census and encourage them to complete it.

Despite these efforts, the census encountered critical problems. The Coronavirus Disease 2019 (COVID-19) pandemic necessitated shutting down, then restarting, census operations. They included the bureau's special operations to enumerate hard-to-count people living in group quarters, in shelters, at campgrounds and other transitory locations, and outdoors, plus people visiting soup kitchens and food vans. Apart from special operations, the bureau's largest field operation, nonresponse follow-up (NRFU), called for repeatedly visiting or phoning households that were disinclined to provide their census responses. In-person NRFU visits, too, had to be postponed.

The Census Bureau reported to the Government Accountability Office a national enumeration rate of 99.98% of housing units on October 15, 2020. The occupants of 99.0 million housing units (67.0% of the total enumerated) responded by internet, on paper, or by phone, before NRFU began; the occupants of an additional 48.6 million housing units (32.9%, from preliminary data) were counted during NRFU.

These numbers told only part of the story. The highest-quality census responses, the bureau noted, came from households that answered before NRFU or with enumerators' assistance during NRFU. Preliminary data indicated that 116.1 million households (about 79%) were in this category, while the remaining 31.04 million households (approximately 21%) did not complete their questionnaires in the initial response phase or during NRFU. For these households, the bureau resorted to other means of enumeration—proxy responses, for about 7.4 million occupied households (24.1%) in the NRFU workload; partial responses, from a yet-to-be-determined number of households; administrative records, for about 8.4 million households (14.0%) covered by NRFU; and imputation, for an unspecified number of households. A proxy response was one supplied by a neighbor, building manager, landlord, or someone else presumably knowledgeable about the household's composition. Imputation generally involved inferring one household's characteristics from those of other households nearby. Group quarters, like college and university housing, proved difficult to count. The bureau had to complete about 1.2 million NRFU cases by phone instead of in person, a process that did not always go smoothly.

The bureau has conducted a Post-Enumeration Survey (PES), its principal means for assessing census quality, and has begun releasing the results. Among other important data, the PES yields estimates of differential net undercounts—the extent to which the 2020 census, like its predecessors, may have undercounted groups such as Blacks and Hispanics relative to non-Hispanic Whites. Other census quality assessments have been or are being made by the American Statistical Association; JASON, an independent science advisory group; and the National Academies of Sciences, Engineering, and Medicine's Committee on National Statistics.

Contents

Contacts

Introduction

More than a decade of planning and testing went into preparations for the 2020 census. Among the innovations for this enumeration were a first-ever online census response option, use of governmental administrative records and satellite imagery to eliminate some of the fieldwork involved in updating census addresses and maps, and greater reliance on technology to streamline census field operations.[1] The Census Bureau conducted extensive advertising, part of a more than "$500 million public education and outreach campaign," to publicize the 2020 census and encourage all U.S. residents to complete it.[2] The advertisements were in English as well as 46 other languages that, according to the bureau, reached "over 99% of households more than 300 times."[3] The bureau accepted census responses online, on paper census forms, and by telephone. Households that needed language assistance to respond online or by phone could receive it in 13 languages. In addition, help was available through guides in 59 languages "in print and video on 2020census.gov." The bureau mailed households up to seven "invitations" and reminders to answer the census, including up to two questionnaires.[4] Enumerators either were to leave questionnaires with households lacking home mail delivery or visit them to collect census data in person. The bureau planned special operations to enumerate hard-to-count people living in group quarters, in shelters, at campgrounds and other transitory locations, and outdoors, plus people visiting soup kitchens and food vans. The bureau also was to try repeatedly to visit or phone households that were disinclined to provide their census responses.[5]

Some considered these efforts to be inadequate and the census to be seriously flawed.[6] The COVID-19 pandemic, which was beyond what the bureau could have foreseen or encompassed in its plans, necessitated shutting down, then restarting, census field operations, delaying data collection[7] and causing some confusion about where certain respondents, notably college and university students, were to be counted.[8] The Trump Administration's attempts to put a

[1] For background information on the 2020 census, see CRS Report R44788, *The Decennial Census: Issues for 2020*, by Jennifer D. Williams; CRS Report R46237, *The 2020 Census: Frequently Asked Questions*, by Jennifer D. Williams; CRS In Focus IF11015, *The 2020 Decennial Census: Overview and Issues*, by Jennifer D. Williams; and CRS In Focus IF11486, *2020 Census Fieldwork Delayed by COVID-19*, by Jennifer D. Williams.

[2] U.S. Census Bureau, "2020 Census Advertising Campaign Expands to Reach All Audiences," press release CB20-CN.13, February 18, 2020, at https://www.census.gov/newsroom/press-releases/2020/2020-ad-campaign-expands.html.

[3] U.S. Census Bureau, "2020 Census Data Quality," at https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/data-quality.html#ensuring.

[4] Ibid.

[5] Ibid.

[6] See, as examples, Jonathan Tilove, "Robert Santos Warned That the 2020 Census Would Be Sabotaged. Now He'll Oversee the Bureau," *Texas Monthly*, November 16, 2021, at https://www.texasmonthly.com/news-politics/robert-santos-census-bureau/; *NPR News*, "How many people of color did the 2020 census miss? COVID makes it harder to tell," October 9, 2021; Tara Bahrampour, "First batch of data from troubled census set for release," *Washington Post*, April 26, 2021, p. A3; and Tara Bahrampour, "2020 Census may have undercounted Black Americans, new analyses say," *Washington Post*, October 13, 2021, at https://www.washingtonpost.com/dc-md-va/2021/10/13/2020-census-black-undercount/. A report by a group of nonprofit organizations with interests in the census and civil rights, looking ahead to 2030, detailed numerous ways in which it believed the 2020 census could have been improved. See New Venture Fund, *Looking to 2030: Findings and Recommendations from Census 2020 Partners and Funders*, October 2021, at https://censusproject.files.wordpress.com/2021/11/looking-to-census-2030-full-report-reduced.pdf.

[7] See CRS In Focus IF11486, *2020 Census Fieldwork Delayed by COVID-19*, by Jennifer D. Williams.

[8] D'Vera Cohn, "It's clear where college students are counted in the 2020 census, but not others who moved due to COVID-19," Pew Research Center, April 17, 2020, at https://www.pewresearch.org/fact-tank/2020/04/17/its-clear-where-college-students-are-counted-in-the-2020-census-but-not-others-who-moved-due-to-covid-19/; and Mike Schneider, "College towns plan to challenge results of 2020 census," *ABC News*, October 18, 2021, at

citizenship question on the census form, just as it was being readied for printing and posting online, reportedly raised concerns, including among some of the foreign-born population, especially noncitizens and people lacking legal status in the United States, and may have further complicated the already difficult task of persuading them to answer the census.[9]

Overall delays in census data collection and processing resulted in late releases of the data for apportioning seats in the House of Representatives and for within-state redistricting. The House apportionment numbers, which, under Title 13, *U.S. Code*, Section 141(b), were due no later than December 31, 2020, were released four months later, on April 26, 2021. They showed an apportionment population of 331.1 million. This population included residents of the 50 states as well as members of the U.S. military and federal civilian employees stationed abroad, plus their dependents living with them. The resident population of the 50 states and DC was reported as 331.4 million.[10] Under Title 13, Section 141(c), the tabulations of census data for use in redrawing state congressional and legislative districts had to be, but were not, delivered by March 31, 2021, to the states that had requested them. The 2020 census redistricting files in a "legacy format" were released on August 12, 2021, and the same data were released in a somewhat more usable format on September 16, 2021.[11]

This report presents selected assessments of 2020 census quality available thus far. Sources are the Census Bureau itself; the Government Accountability Office (GAO), which typically publishes reports and makes recommendations regarding the decennial census; the American Statistical Association (ASA); JASON, an independent science advisory group; and the National Academies of Sciences, Engineering, and Medicine's Committee on National Statistics (CNSTAT), which also generally examines decennial census quality.

Early Assessments by GAO

As the Census Bureau ended 2020 census data collection on October 15, 2020, it reported to GAO a national enumeration rate of 99.98% of housing units.[12] The occupants of 99.0 million housing units (67.0% of the total enumerated) responded on their own initiative—by internet, by phone, or on paper census forms—before the nonresponse follow-up field operation (NRFU) began; the occupants of an additional 48.6 million housing units (32.9% of the total enumerated, from preliminary data) were counted during NRFU.[13] In NRFU, GAO wrote, enumerators were to

https://abcnews.go.com/Lifestyle/wireStory/college-towns-plan-challenge-results-2020-census-80646102.

[9] Anita Kumar and Caitlin Oprysko, "Trump Abandons Effort to Add Citizenship Question to Census," *Politico*, July 11, 2019, at https://www.politico.com/story/2019/07/11/trump-expected-to-take-executive-action-to-add-citizenship-question-to-census-1405893; Jill Colvin, Mark Sherman, and Zeke Miller, "Trump abandons bid to include citizenship question on census," Associated Press, July 11, 2019, at https://apnews.com/article/immigration-donald-trump-ap-top-news-courts-supreme-courts-18db0fba2743496daeb27a92915bb260; Michael Wines, "2020 Census Won't Have Citizenship Question as Trump Administration Drops Effort," *New York Times*, July 2, 2019, at https://www.nytimes.com/2019/07/02/us/trump-census-citizenship-question.html; and Mike Schneider, "Census analysis finds undercount but not as bad as predicted," Associated Press, November 2, 2021, at https://apnews.com/article/coronavirus-pandemic-health-census-2020-think-tanks-congress-5a325059f0a7c7a0d2483b0130551ba9.

[10] U.S. Census Bureau, "2020 Census Apportionment Results," April 26, 2021, at https://www.census.gov/data/tables/2020/dec/2020-apportionment-data.html.

[11] U.S. Census Bureau, "Decennial Census P.L. 94-171 Redistricting Data," September 16, 2021, at https://www.census.gov/programs-surveys/decennial-census/about/rdo/summary-files.html#P1.

[12] U.S. Government Accountability Office, *2020 Census: The Bureau Concluded Field Work but Uncertainty about Data Quality, Accuracy, and Protection Remains*, GAO-21-206R, December 2020, p. 4.

[13] Ibid.

confirm vacant and nonexistent addresses and make "repeated visits to occupied homes to capture full information from each household." If an enumerator could not obtain information from a household, the enumerator was to leave a "notice of visit" form asking that someone in the household respond to the census online or by calling a Census Questionnaire Assistance Center.[14]

The bureau told GAO that by October 15, it had completed 99.93% of its NRFU workload, covering approximately 64.1 million households. The workload included housing units eventually determined to be occupied, vacant, or nonexistent; cases in which the bureau had completed interviews but returned to recheck the information it had collected; and cases in which the bureau had "high-quality administrative records," consisting of Social Security, Internal Revenue Service, or previous census records, that it could use for enumeration after visiting a household at least once.[15]

According to the bureau, the highest-quality census responses came from households that answered before NRFU or with assistance from enumerators during NRFU.[16] Preliminary data indicated that 116.1 million households (about 79%) were in this category, while the remaining 31.04 million households (approximately 21%) did not complete their questionnaires in the initial response phase or during NRFU.[17]

For the latter households, the bureau resorted to other means of enumeration.[18] Examples were proxy responses, used to obtain data for about 7.4 million occupied households (24.1%) in the NRFU workload; partial responses, from a yet-to-be-determined number of households; administrative records, for about 8.4 million households (14.0%) covered by NRFU; and imputation, for an unspecified number of households. A proxy response was one supplied by a person, like a neighbor, building manager, or landlord, presumably knowledgeable about the household's composition. Administrative records were mentioned above, and imputation generally involved inferring one household's characteristics from those of other households nearby.

A Subsequent GAO Assessment

A later GAO publication focused on data collection problems and workarounds in the 2020 census.[19] Although the Census Bureau reported completing more than 99% of its national nonresponse follow-up workload by October 15, 2020 (see above), partly through the use of technology, NRFU cases completed ranged from about 94% to 99% in 10 local areas. In the Shreveport, LA, area, which had the lowest completion rate (93.5%), a category 4 hurricane in August 2020 disrupted data collection for 82,863 housing units.[20] In some hard-to-count areas, the bureau "increased the use of travel teams to complete cases," sending about 26,000

[14] U.S. Government Accountability Office, *Decennial Census: Bureau Should Assess Significant Data Collection Challenges as It Undertakes Planning for 2030*, GAO-21-365, March 22, 2021, p. 3.

[15] U.S. Government Accountability Office, *2020 Census: The Bureau Concluded Field Work but Uncertainty about Data Quality, Accuracy, and Protection Remains*, GAO-21-206R, December 9, 2020, p. 4.

[16] Ibid., p. 5.

[17] Ibid.

[18] Ibid., pp. 5-8.

[19] U.S. Government Accountability Office, *Decennial Census: Bureau Should Assess Significant Data Collection Challenges as It Undertakes Planning for 2030*, GAO-21-365, March 2021.

[20] Ibid., p. 11.

enumerators to other places during NRFU.[21] Occasionally, the bureau offered enumerators financial incentives for working extra hours "while completing a minimum number of cases, and removed restrictions on overtime." In March 2021, noted GAO, the bureau "reported $98.4 million in expenses for enumerator awards, $795,000 of which was for travel awards."[22] In some areas "struggling to resolve cases or facing challenges to accomplish in-person interviews because of COVID-19 and natural disasters," enumerators switched to phone interviews. Bureau data indicated that "enumerators made more than 10 million calls and completed nearly 1.2 million cases by phone."[23] The bureau's regional leadership had discretion to authorize the phone interviews, and bureau officials reported to GAO satisfaction with the process "to make contact and complete census responses" during NRFU.[24] Nevertheless, as one census field supervisor told GAO, "phone enumerations were difficult to conduct remotely" because of NRFU "rules that required an in-person proxy rather than over the phone after three contact attempts."[25] GAO informed the bureau about this concern, and the bureau then "clarified for enumerators how they could use the handheld device to indicate that they could not complete an in-person proxy, or to complete the proxy interviews later when they returned to the field."[26]

Another difficulty encountered during NRFU was "the inability of supervisors to reassign open cases in a timely fashion. GAO found that census field supervisors did not have the authority to reassign cases and had to wait for the field manager to make those reassignments."[27] Bureau officials informed GAO that the bureau "would consider the reassignment of cases" as it moved toward 2030 census planning.[28]

GAO also questioned how adequate certain bureau procedures were for monitoring the quality of NRFU operations, "such as real-time monitoring of enumerator activities by supervisors and training assessments." GAO found that the bureau "did not have proper controls in place, allowing some enumerators to work without having passed the required training assessment." The bureau "agreed that additional controls were necessary."[29]

The group quarters enumeration was another matter of GAO concern:

> The Bureau planned to count individuals living in group quarters, such as skilled nursing and correctional facilities, between April 2, 2020, and June 5, 2020, but revised those dates to July 1, 2020, through September 3, 2020. The pandemic made it difficult to count group quarters. For example, Bureau staff found it challenging to locate a point of contact at some group quarters because facilities were closed due to the pandemic. Bureau officials told us that in December 2020 they decided to re-contact more than 24,000 out of approximately 272,000 group quarter facilities to collect data, and that imputation would be used to count individuals at the remaining facilities still reporting a zero population count.[30]

[21] Ibid.

[22] Ibid.

[23] Ibid., p. 12.

[24] Ibid., p. 13.

[25] Ibid.

[26] Ibid.

[27] Ibid., GAO Highlights.

[28] Ibid.

[29] Ibid.

[30] Ibid. GAO observed that residents of college and university housing were particularly challenging to count. When in-person student instruction stopped in March 2020 because of the pandemic, students left their campuses. The bureau reported receiving census data from 88% of student housing, but colleges and universities "did not always provide the

A further problem census field supervisors reported to GAO was difficulty accessing large, multi-unit buildings for NRFU. Enumerators "were often turned away" from these buildings "because of the pandemic," but access had been difficult even before it. In such instances, enumerators were left to collect information from alternative sources like administrative records or proxy data.[31]

American Statistical Association Report

GAO wrote in December 2020 that the American Statistical Association and the Census Scientific Advisory Committee "issued numerous recommendations in the fall of 2020, including for the Bureau to document what it knows in near real-time about the quality of the population counts that it provides to the President and to Congress."[32]

The ASA recommendations to which GAO referred are from a report by the association's 2020 Census Quality Indicators Task Force.[33] On it were two former OMB chief statisticians, three former Census Bureau directors, other former bureau officials, and academicians with statistical and census expertise.

At the time the ASA report was issued, in October 2020, the Census Bureau had "released information on the percentage of 'completed enumerations' by state with a goal of reaching 99 percent completion for each state." ASA pointed out, however, that the percentage of completed cases did not "suffice to draw conclusions about data quality." For example, "the tally of completed enumerations" covered "households counted through a proxy response from a neighbor, including cases in which the proxy could provide no information beyond a guess of the number of individuals living in the household. In fact, meeting enumeration goals for a truncated deadline" increased "the likelihood of operational shortcuts" possibly affecting "the quality of the count."

---

data" to the bureau. The reasons cited were concerns about students' privacy, insufficient points of contact for some fraternities and sororities, and some school administrators' perception that "completing the census was not a priority." Ibid., pp. 24-25.

[31] Ibid., p. 13.

A *Federal Register* notice invited public comment on a proposed 2020 Post-Census Group Quarters Review (2020 PCGQR), which is to provide "a mechanism for tribal, state, and local governmental units in the United States and Puerto Rico, or their designated representatives, to submit a request that the Census Bureau review their population counts" for group quarters "by block to correct error(s) affecting the inclusion of group quarters" and their residents in the 2020 census. The notice specified, "No new decennial information products will be created by the 2020 PCGQR. The Census Bureau will not revise any 2020 Census information products, such as the population counts delivered to the President for apportionment or the 2020 Census P.L. 94-171 Redistricting Data Files and Geographic Products." If, though, the bureau "determines that a submission identifies existing group quarters" and associated population counts, "then the Census Bureau will accept the records and update [population] counts for the governmental unit for use in Census Bureau programs, as appropriate. The Census Bureau will issue certified population count corrections, which governmental units can use for any purpose requiring their official Census counts. The Census Bureau will incorporate all 2020 PCGQR revisions into the intercensal population estimates and American Community Survey estimates starting in 2022 and will post the new counts" on its website. U.S. Department of Commerce, Census Bureau, "Agency Information Collection Activities; Submission to the Office of Management and Budget (OMB) for Review and Approval; Comment Request; 2020 Post-Census Group Quarters Review," 86 *Federal Register* 64896, November 19, 2021.

[32] U.S. Government Accountability Office, *2020 Census: Census Bureau Needs to Ensure Transparency Over Data Quality*, GAO-21-262T, December 3, 2020, pp. 6-7.

[33] American Statistical Association, *2020 Census Quality Indicators: A Report from the American Statistical Association*, October 2020, at https://www.amstat.org/asa/files/pdfs/POL-2020CensusQualityIndicators.pdf.

ASA noted, too, that by October 2020, the bureau had released "extensive response rate data" for "a number of geographic levels." The data, added the task force, "need to be analyzed to identify any lack of uniformity across geographic entities such as states, cities, counties, and rural areas, and population groups, which would indicate inequity of the counts across the U.S."[34]

ASA identified two types of quality indicators for field data collection in the census, notably the nonresponse follow-up operation. One type of indicator is based on paradata about the NRFU process: how many attempts an enumerator made "to enumerate a NRFU case," for instance, or how much time an enumerator spent on interviews. The other type is based on outcomes, like the number of households "enumerated by asking a neighbor" how many people were living in the unit and the number "enumerated using administrative records."[35]

ASA suggested that "because of the truncated timeframe" for field operations "and the effects of the pandemic and multiple natural disasters," the Census Bureau "make the quality assessment results available to the public at the census tract level," the third-lowest level of census geography. The public thus could ascertain in detail "the extent to which some areas" may have been "counted more accurately than others and determine the data's fitness for various uses."[36]

The task force called establishing these quality indicators "just the first step" and made five additional recommendations:

- "The indicators should be readily available and used expeditiously to assess the quality of the 2020 Census," preferably before release of the apportionment data and "most certainly" before the states received their redistricting data.[37]
- "Qualified external researchers should be granted access to the data to help conduct the analyses." Usual Census Bureau practice would be to conduct its evaluations internally. The unusual 2020 census circumstances, according to the ASA report, made outside expert assessments advisable to promote transparency, timeliness, and credibility. "Another important component of this research would be to allow local area experts such as state demographers to review the early tabulations and help the Census Bureau determine if unexpected discrepancies are, in fact, computer processing errors."[38]
- "Additional assessments should be conducted when more data become available." As the bureau did in several past censuses, it conducted a Post-Enumeration Survey (PES) that is expected to provide "a wealth of information" about 2020 census quality. The 2010 census coverage measurement program, for instance, "allowed for extensive analysis of quality," covering "duplication[s], undercounts, and erroneous enumerations." The task force noted that the PES results were to be available early in 2022 and possibly could be used to make corrections in the 2020 data that could affect the distribution of federal funds, especially to areas estimated to have been undercounted.[39]
- The bureau typically begins planning for the next census before completing the current census. The task force proposed that early 2030 census planning draw on

[34] Ibid., p. 2.

[35] Ibid.

[36] Ibid. See below, in contrast, JASON's comments about the need for aggregated data to protect respondents' privacy.

[37] Ibid., p. 13.

[38] Ibid.

[39] Ibid.

the lessons of 2020, "be conducted in public, and include extensive stakeholder input," particularly in view of "the concerns that have been expressed" about 2020 census quality. ASA wrote that among the challenges this census faced, which contributed to doubts about quality, were the COVID-19 pandemic, restructured census field operations, natural disasters, and litigation.[40]

- The task force advised that the bureau's authorizing legislation, Title 13, *U.S. Code*, be updated. "First," the task force wrote, "Title 13 needs to align better with recent Information Quality Act and OMB guidelines." The act "requires agencies to conduct pre-dissemination review of their information products," during which "each agency should consider the appropriate level of quality for each of the products that it disseminates based on the likely use of that information." Also, "agencies are required to produce measures of quality that accompany the release of important data," like the apportionment and redistricting data. "Adding these requirements to Title 13 would build confidence" in the bureau as it continues its work. Second, the task force advised that Title 13 be "examined to determine if further amendments to protect the integrity and independence of the Census Bureau and the decennial census are warranted."[41]

JASON Report

As the report by JASON noted, the Census Bureau engaged the independent science advisory group "to help assess" the bureau's "data quality processes and procedures, and to elicit recommendations on areas for improved communication." The report focused on 2020 census data quality, "compared with what would have been achieved in better circumstances," and especially on "the fitness of the data for the constitutional and statutory uses of the census."[42]

The process of readying the resident population counts, the apportionment data, and the redistricting data for release involved extensive census field operations and post-data collection processing. "Several intermediate internal data products" preceded the release of the apportionment counts.[43] They were

- the Decennial Response File 1 (DRF1), whose production required the Census Bureau to "determine the universe of housing units and group quarters," identify "unique persons," standardize responses, and assess the quality of cases for nonresponse follow-up by incorporating the results of field reinterviews;[44]

[40] Ibid., p. 14.

[41] Ibid.

[42] JASON, JASON Letter Report JSR-20-2N, *Assessment of 2020 Census Data Quality Processes*, February 8, 2021, p. 2. The report can be accessed at https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-census-data-quality-processes.pdf.

For background information about JASON, see Ann Finkbeiner, "JASON—a secretive group of Cold War science advisers—is fighting to survive in the 21st century," *ScienceInsider*, June 27, 2019, at https://www.science.org/content/article/jason-secretive-group-cold-war-science-advisers-fighting-survive-21st-century.

[43] JASON, JASON Letter Report JSR-20-2N, *Assessment of 2020 Census Data Quality Processes*, February 8, 2021, p. 4. The report can be accessed at https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/2020-census-data-quality-processes.pdf.

[44] Ibid., pp. 4-5.

- the Decennial Response File 2 (DRF2), whose production involved removing duplicate responses and thus "processing the entire nation at once rather than state-by-state as in DRF1";[45]
- the Census Unedited File (CUF), for which the bureau had to integrate administrative records and determine "the final population count for each address," by imputation if necessary;[46] and
- the state-level population counts, whose production involved "merging in the count of the Federally Affiliated Overseas population for each state," using "multiple independent methods" to ensure the accuracy of the apportionment numbers, and "comparing and reconciling final results."[47]

At the time the JASON report was written, the Census Bureau was still "working through its post-processing of the data" and had not "finalized the delivery dates for the constitutional and statutory data releases." As JASON pointed out after being briefed by the bureau on February 2, 2021, "several consequences of the pandemic and other events" increased "the importance of completing and checking all aspects of the processing milestones." One possible consequence cited, for example, was that the pandemic could have exacerbated the problem of duplicate census responses when college and university students were relocated away from campus and off-campus housing and back to their "primary residences."[48]

The bureau shifted its timeline for data delivery several times, first extending it to allow for adequate data processing and related quality checks, then compressing the timeline, and finally returning to the previously extended deadline of April 30, 2021, for delivery of the apportionment numbers.[49] JASON recommended that the Secretary of Commerce return to this deadline, "In accordance with Sec. 4 of President Biden's Executive Order ('Ensuring a Lawful and Accurate Enumeration and Apportionment Pursuant to the Decennial Census', January 20, 2021)."[50]

JASON recommended further that the bureau try to promote confidence in a "potential differential count of the population" by summarizing its "assessments of data quality across various geographies and for relevant demographic groups."[51] JASON stressed the importance of summary data. "One important consideration for any public release of information about data quality is the risk of disclosing sensitive information." Although "granular analyses" of detailed geographic or demographic information would be "useful internally," any data made public "from these analyses should be aggregated to avoid disclosure risk."[52]

The Census Bureau conducted a nationwide 2019 census test to gauge the effects on response rates of a possible question about the respondent's citizenship. JASON suggested comparing the 2019 (pre-COVID-19) response rates from the control group that did not get the citizenship question with response rates from the 2020 census (during COVID-19) to evaluate the pandemic's effects, both on "absolute" response rates and on "preferences between mail-in and

[45] Ibid., p. 5.

[46] Ibid.

[47] Ibid.

[48] Ibid.

[49] Ibid., pp. 5-6.

[50] Ibid., p. 6.

[51] Ibid., p. 8.

[52] Ibid.

internet responses."[53] Among the "limiting factors" JASON cited for such a comparison were that only areas with U.S. Postal Service mail delivery could receive the 2019 test forms; the bureau conducted nonresponse follow-up in the census but not the test; the bureau required identification numbers for internet responses to the test, but not the census; the bureau provided less language support for the test than the census; and the test and the census occurred in different months, when residence patterns may have differed.[54] JASON noted that the Census Bureau already planned to compare response rates from the 2019 census test and the 2020 census to evaluate "the efficacy of advertising" on responses. February 2022 was the tentative release date for the study's findings.[55]

JASON advised the bureau to "develop time series that clearly summarize changes in enumeration conditions and response rates" for its "in-person, online, and mail solicitations." According to JASON, time series "would highlight that although enumeration conditions varied, by adjusting operational strategies," the bureau "was able to avert deficiencies in enumeration that otherwise would have appeared."[56]

In preparation for the 2020 census, JASON observed, the bureau developed "software and hardware that made it possible not only to monitor the data collection process, but also to adapt deployment of resources in response to various issues."[57] As examples, the bureau developed a Survey Operational Control System and a Field Operational Control System that it used "to distribute workloads to enumerators." Enumerators uploaded their data with a Field Data Collection App, which served to "optimize labor assignments." These tools "allowed for a much more integrated approach to the NRFU process." For instance, when someone on the list for nonresponse follow-up answered the census before a NRFU visit occurred, "the NRFU list was immediately updated, thus improving efficiency." JASON recommended that the Census Bureau "collect and analyze the data from its control system tools" and summarize the data "to develop indicators of enumeration progress as a function of time that can be compared" to measures of progress in past censuses.[58]

JASON noted that another tool, the Census Review Analysis and Visualization Application (CRAVA), provided "detailed assessments of data quality at the county level for the responses to all questions asked" and, "for some responses," assessments "down to the census block level," the lowest level of census geography. The bureau, JASON wrote, should use the CRAVA-based reports "to show how data quality issues were identified and addressed over time" as the various data products (DRF1, DRF2, CUF, etc.) were generated.[59]

JASON also pointed out that in processing census data, the bureau "performs a number of checks for data completeness and consistency and detects potential inconsistencies" it calls "anomalies." These inconsistencies, JASON explained, "are not surprising or indicative of serious problems but are an expected and normal part of the data quality-control processes." JASON proposed that the bureau discontinue using the term "anomaly," which "may raise inappropriate alarm to non-experts, and communicate openly about the measures used to check for data consistency, the

[53] Ibid., p. 9.

[54] Ibid.

[55] Ibid.

[56] Ibid., p. 11.

[57] Ibid., p. 12.

[58] Ibid.

[59] Ibid., p. 13.

specific issues uncovered, and steps taken to address them." If possible, this information should be compared "with the number and type of issues" identified in the 2010 census.[60]

JASON acknowledged the bureau's "significant capacity for communication with many stakeholders, ranging from primary school students to expert data users," and mentioned favorably "the partnerships established with numerous organizations" during the 2020 census, including efforts to communicate with businesses and state and local governments.[61] Nevertheless, JASON found "a shortfall" in the bureau's communication efforts "regarding its enumeration processes and changes made to these processes for a range of reasons, including those resulting in improvements."[62] JASON advised the bureau to "develop tiered messaging," from several-minute videos to about an hour-long presentation, that increasingly quantifies "the myriad efforts undertaken to ensure" an accurate 2020 census count.[63]

Operational Quality Metrics

In line with ASA's and JASON's recommendations, the Census Bureau has made public on its website certain detailed "operational quality metrics" as 2020 census data have become available.[64] Examples are comparisons to alternative measures of the U.S. population with Demographic Analysis (DA), discussed below, or population estimates; how the bureau accounted for nonrespondents, including by imputation or, during NRFU, with administrative records, proxy responses by neighbors, or responses from household members whom the bureau contacted; and item nonresponse rates, analyzing nonresponses to particular census questions. The bureau released sub-state summaries of these metrics but, in keeping with data privacy concerns expressed by JASON, decided against releasing the metrics for lower geographic levels like counties or census tracts.[65]

The 2020 Census Coverage Measurement Program

The Census Bureau has noted its "long history of evaluating population coverage in decennial censuses," starting with the 1940 census. "Almost everything we know about the size of the coverage error, trends in census coverage and differences among subgroups of the population comes from the Census Bureau's own evaluation programs."[66] "Coverage error" refers to population undercounts or overcounts in the census.[67] The two principal means of evaluating coverage, discussed below, are Demographic Analysis and dual-system estimation by means of a Post-Enumeration Survey.

---

[60] Ibid., p. 14.

[61] Ibid., p. 15.

[62] Ibid.

[63] Ibid., p. 16.

[64] See U.S. Census Bureau, "2020 Decennial Census Data Quality, Operational Quality Metrics," at https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/data-quality.html#metrics.

[65] U.S. Census Bureau, "2020 Census Operational Quality Metrics: Sub-State Summaries," at https://www.census.gov/newsroom/blogs/random-samplings/2021/08/2020-census-operational-quality-metrics.html.

[66] U.S. Census Bureau, "Coverage Measurement," at https://www.census.gov/programs-surveys/decennial-census/about/coverage-measurement.2000.html.

[67] U.S. Census Bureau, "Methodology for the 2020 Demographic Analysis Estimates," at https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2020da_methodology.pdf, December 15, 2020.

Demographic Analysis

Demographic Analysis produces national-level estimates of the U.S. population that are derived largely, but not entirely, from administrative records. The federal government acquires these records, such as on births and deaths, for various reasons unconnected with the decennial census. DA is designed to be independent of the census.

As the bureau has explained, Demographic Analysis is used to produce "estimates of net coverage error, which are calculated as the percent difference between the DA estimates and the census counts."[68] The reference date for the present DA estimates is the official Census Day, April 1, 2020. The DA estimates of the U.S. population on this day, by age, sex, broad racial categories (Black, alone or in combination with other races, and non-Black, alone or in combination with other races), and Hispanic and non-Hispanic ethnicity, "were developed using a basic population accounting approach. The main source of data for the births and deaths is the National Vital Statistics System maintained by the National Center for Health Statistics. The foreign-born population for all birth cohorts was estimated primarily using a stock method and data from the American Community Survey. The estimates of the population born before 1945 (ages 75 and older) were developed using Medicare records."[69] The bureau "produced three sets of official estimates," and "a range of estimates (low, middle, and high)" for each set "to account for uncertainty in the data, methods, and assumptions used for the 2020 DA."[70]

The bureau released its first national tables from Demographic Analysis on December 15, 2020,[71] ahead of the April 26, 2021, release of the first 2020 census data: the state apportionment counts[72] and resident population counts for the states, DC, and Puerto Rico.[73] The bureau noted that "a report detailing the 2020 Demographic Analysis estimates of net overcounts and undercounts" was planned for later release.[74] The December 15, 2020, data gave the following range of estimates of U.S. population size on Census Day: low, 330,730,000; middle, 332,601,000; and high, 335,514,000. The actual census resident population count of 331,449,281 fell between the low and middle estimates.[75]

In addition, Demographic Analysis produced national-level low, middle, and high estimates of the population by age, sex, broad racial categories (Black alone/non-Black alone, Black alone or in combination/non-Black alone or in combination), and Hispanic/non-Hispanic ethnicity by single year of age and sex.[76] The bureau noted that the estimates could be produced only "in limited race

[68] U.S. Census Bureau, "Demographic Analysis (DA)," at https://www.census.gov/programs-surveys/decennial-census/about/coverage-measurement/da.html.

[69] U.S. Census Bureau, "Methodology for the 2020 Demographic Analysis Estimates," at https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2020da_methodology.pdf, December 15, 2020.

[70] Ibid.

[71] U.S. Census Bureau, "Census Bureau Releases 2020 Demographic Analysis Estimates," press release CB20-CN.133, December 15, 2020.

[72] U.S. Census Bureau, "Apportionment Population and Number of Representatives by State: 2020 Census," at https://www2.census.gov/programs-surveys/decennial/2020/data/apportionment/apportionment-2020-table01.pdf.

[73] U.S. Census Bureau, "Resident Population for the 50 States, the District of Columbia, and Puerto Rico: 2020 Census," at https://www2.census.gov/programs-surveys/decennial/2020/data/apportionment/apportionment-2020-table02.pdf.

[74] U.S. Census Bureau, "Census Bureau Releases 2020 Demographic Analysis Estimates," press release CB20-CN.133, December 15, 2020.

[75] Ibid.

[76] Ibid.

detail because they relied on historical records and measures of race that have changed over time." Further, DA estimates of the Hispanic population were not produced for all ages because the option for reporting Hispanic ethnicity "was not widely available on birth and death records until the 1990s."[77]

- The low, middle, and high percentages of the U.S. population estimated to be Black alone were, respectively, 13.5%, 13.7%, and 13.9%.[78]
- The percentages of the population estimated to be Black alone or in combination with other races were 14.9%, 15.1%, and 15.4%.[79]
- The percentages of the population under age 30 estimated to be Hispanic were 23.0%, 24.6%, and 26.0%.[80]

The Census Bureau has cautioned against prematurely comparing the DA estimates of race and ethnicity with data from the 2020 census or population estimates: "Many people want to evaluate how well specific race groups were counted in the 2020 Census by comparing the race data from the 2020 Census to the DA results or the population estimates. We caution against making these comparisons right now because the available benchmarks do not have the same race categories as the census."[81]

Post-Enumeration Survey

Like Demographic Analysis, the Post-Enumeration Survey is designed to be independent of the census and to indicate the extent of census accuracy. Unlike DA, though, the PES is intended to be a nationally representative sample survey whose results are matched with census results to estimate census errors both nationally and subnationally. More specifically, the bureau has explained, the 2020 census PES uses "dual-system estimation," in which the two systems are the PES and the census. The PES "independently interviews people, asks where they lived on April 1, 2020, and then matches that information to the census results." The PES "takes more than two years and involves enumerating housing units and people from scratch in about 10,000 blocks across the country." By matching "these housing units and people to the list of addresses and people in the census," the bureau can determine who was counted only in the census, only in the PES, and in both the census and the PES.[82] The PES should enable the Census Bureau to provide estimates of "net coverage error" as well as "components of coverage."

"Net coverage error," in the bureau's terminology, refers to "the difference between the census count and the PES estimate of the actual number of people in the U.S." Negative coverage error would mean that the census count was "too low and the census missed some people," so had an undercount. Positive coverage error, or an overcount, would mean that the census count was "too

[77] Ibid.

[78] Ibid.

[79] Ibid.

[80] Ibid.

[81] Eric Jensen and Sandra Johnson, "Using Demographic Benchmarks to Help Evaluate 2020 Census Results," U.S. Census Bureau, November 9, 2021, at https://www.census.gov/newsroom/blogs/random-samplings/2021/11/demographic-benchmarks-2020-census.html. See especially, at this site, mention of the 2020 Modified Race File and the Demographic and Housing Characteristics File.

[82] Timothy Kennel, "The Post-Enumeration Survey: Measuring Coverage Error," U.S. Census Bureau, December 16, 2021, at https://www.census.gov/newsroom/blogs/random-samplings/2021/12/post-enumeration-measuring-coverage-error.html.

high, indicating some people" may have been "counted more than once."[83] The bureau, additionally, checks the census numbers and the PES estimates for various groups according to their demographic characteristics, namely: groups by age, sex, race (Whites, with a breakout for non-Hispanic Whites; Blacks; Asians; American Indians or Alaska Natives, broken out by American Indians or Alaska Natives living on reservations, living on American Indian Areas off reservations, and living in the rest of the United States; Native Hawaiians or other Pacific Islanders; and people who identify as some other race), and Hispanic ethnicity. The bureau then compares "net coverage error rates" across demographic groups to determine differences in how the census counted them. "When a group has a larger or smaller net undercount than the country as a whole," the bureau calls it a "differential net undercount."[84]

"Components of coverage" that the bureau was to identify during PES fieldwork consist of correct enumerations, erroneous enumerations, and whole-person imputations. "Correct enumerations" in the 2020 census refer to U.S. residents counted as of April 1, 2020, where they should have been counted, in the United States at their usual residences. "Erroneous enumerations" include "duplicate records of people who were correctly counted in the census as well as people who were counted but should not have been. For example, they may have died before April 1, 2020," or been just visiting the United States. "Whole-person imputations" refer to "a statistical technique" used to fill in data missing from census records. The bureau is to count the number of census records requiring whole-person imputation and use the PES to estimate correct and erroneous enumerations.[85]

The Census Bureau has used some type of survey to estimate census coverage and accuracy since 1950.[86] The evaluation of the 1980 census was designated the Post-Enumeration Program. In the 1990 census, the Post-Enumeration Survey was known by that term. In 2000, it was called Accuracy and Coverage Evaluation and in 2010, Census Coverage Measurement.[87]

Fieldwork for the 2020 census Post-Enumeration Survey, again given that name, began in January 2020, when the bureau started developing an address list independent of the one used for the actual census.[88] The bureau announced on February 11, 2021,[89] that it was hiring temporary workers for the in-person follow-up phase of the PES, to have begun in June. During this phase, the bureau was to revisit previously interviewed housing units to confirm its records. The first results of the PES were tentatively scheduled for release at the end of December 2021 and more findings by the end of March 2022.[90] According to an October 9, 2021, NPR news report, however, in-person PES interviews were interrupted by the pandemic, so were to take place from November 2021 to February 2022.[91] In a November 16, 2021, press release, the bureau reported that follow-up would begin on November 29, 2021, and continue through March 2022, with

[83] Ibid.

[84] Ibid.

[85] Ibid.

[86] Ibid.

[87] U.S. Census Bureau, "Post-Enumeration Surveys," at https://www.census.gov/programs-surveys/decennial-census/about/coverage-measurement/pes.html.

[88] U.S. Census Bureau, "Census Bureau Hiring for Independent Survey That Measures Quality of 2020 Census," press release CB21-CN.12, February 11, 2021.

[89] Ibid.

[90] Ibid.

[91] *NPR News*, "How many people of color did the 2020 census miss? COVID makes it harder to tell," October 9, 2021.

release of the first PES results planned for "the first quarter of 2022" and additional results expected "in the summer of 2022."[92]

On March 10, 2022, the bureau released its national-level PES estimates,[93] with this overall assessment:

> "Today's results show statistical evidence that the quality of the 2020 Census total population count is consistent with that of recent censuses. This is notable, given the unprecedented challenges of 2020," said Director Robert L. Santos. "But the results also include some limitations—the 2020 Census undercounted many of the same population groups we have historically undercounted, and it overcounted others."[94]

The census undercounted Blacks, American Indians or Alaska Natives living on reservations, Hispanics, and people who reported being of some other race.

The census overcounted non-Hispanic Whites and Asians. Native Hawaiians or other Pacific Islanders were neither overcounted nor undercounted, the estimates indicate.

Among age groups, the 2020 census undercounted children up to age 17, especially children up to age 4. "Young children are persistently undercounted in the decennial census," the bureau noted.[95]

More specific PES results by race and Hispanic ethnicity follow below.[96]

- Blacks, alone or in combination with other racial groups, "had a statistically significant undercount of 3.30%. This is not statistically different from the 2.06% undercount in 2010."
- Hispanics "had a statistically significant undercount rate of 4.99%. This is statistically different from a 1.54% undercount in 2010."
- American Indians or Alaska Natives, alone or in combination with other groups, who lived on reservations, had "a statistically significant undercount rate of 5.64%," not "statistically different from a 4.88% undercount in 2010."
- "The non-Hispanic White alone population had a statistically significant overcount rate of 1.64%. This is statistically different from an overcount of 0.83% in 2010."
- Asians, alone or in combination with other groups, "had an overcount rate of 2.62%. This is statistically different from 0.00% in 2010."
- Native Hawaiians or other Pacific Islanders, alone or in combination with other groups, "had an estimated overcount rate of 1.28%. This rate is not different from an estimated 1.02% overcount rate in 2010." Neither rate is "statistically different from zero."

The PES estimates include components of coverage (mentioned above), namely "the number of correct census records, erroneous enumerations, whole-person imputations, and omissions."[97] The DA results do not include coverage estimates by component.

---

[92] U.S. Census Bureau, "Census Bureau's Final Post-Enumeration Survey Field Operation Set to Begin," press release CB21-CN.76, November 16, 2021.

[93] U.S. Census Bureau, "Census Bureau Releases Estimates of Undercount and Overcount in the 2020 Census," press release CB22-CN.02.

[94] Ibid.

[95] Ibid.

[96] Ibid.

[97] Ibid.

According to the PES, of the 323.2 million people living in housing units on April 1, 2020, an estimated

- "94.4% were counted correctly";
- "2.2% were counted erroneously (1.6% were duplicates and 0.6% were erroneous for other reasons)";
- "2.8% provided only a census count and had their demographic characteristics imputed, or statistically inserted"; and
- "0.6% needed more extensive imputation after all census follow-up efforts were attempted."

In addition, the PES estimates indicate 18.8 million 2020 census omissions. "Omissions," according to the Census Bureau, "are people who were not correctly counted in the census. Some of these people were missed by the census. However, omissions may have been accounted for in the census counts as whole-person imputations."[98]

PES and DA Estimates Compared

Limited comparisons of the PES and Demographic Analysis findings accompanied the Census Bureau's March 10 release of the PES estimates.[99]

The two sets of estimates produced results by age and sex. For ages 0-17, the PES shows a statistically significant undercount of 0.84%, and the middle DA series shows a 2.1% undercount. At ages 0-4, the PES indicates a statistically significant undercount of 2.79%, and DA shows a middle series undercount of 5.4%.[100]

The principal difference between the PES and DA results is at ages 18-29, where the PES shows statistically significant undercounts of 2.25% for males and 0.98% for females. DA, in contrast, shows overcounts for males and females, except males in the high series. The bureau pointed out that the DA estimates refer to the total population living in the United States on April 1, 2020, while the PES estimates are based on the household population living in the United States, excluding Remote Alaska areas. Further, a considerable number of people in this age category were living in college dormitories, which were not part of the PES sample. Census Bureau researchers, the bureau stated, "acknowledge this notable difference" between the PES and DA estimates and are trying to account for it.[101]

The PES and DA estimates of net coverage error for adults ages 30-49 show similar patterns. The PES indicates a statistically significant undercount of males in this age range, and DA also shows an undercount.[102]

The PES shows statistically significant overcounts of 0.55% for males and 2.63% for females ages 50 and older. DA, too, indicates overcounts of males and females ages 50 and older, except males in the high series.[103]

[98] Ibid.

[99] Ibid.

[100] Ibid.

[101] Ibid.

[102] Ibid.

[103] Ibid.

Review by the Committee on National Statistics

The National Academies of Sciences, Engineering, and Medicine's Committee on National Statistics has formed a panel "to review and evaluate the quality of the 2020 census." The review continues CNSTAT's long practice of independently assessing census data quality as well as the operations that produced the data. CNSTAT noted that this review is particularly needed in view of some "major changes to decades-old processes" and the "unprecedented" challenges to the census associated with COVID-19.[104]

The panel is examining the Census Bureau's information about the census data it collected, including indicators of data quality obtained during census operations; information from Demographic Analysis and the PES; analyses of administrative records; and relevant evaluations from the 2010 and 2000 censuses.[105]

The panel is to produce an interim report and, by April 2023, a final report with its conclusions and recommendations for the 2030 census.[106]

Author Information

Jennifer D. Williams
Specialist in American National Government

Disclaimer

[104] National Academies of Sciences, Engineering, and Medicine, Committee on National Statistics, Panel to Evaluate the Quality of the 2020 Census, at https://www.nationalacademies.org/our-work/panel-to-evaluate-the-quality-of-the-2020-census.

[105] Ibid.

[106] Ibid.