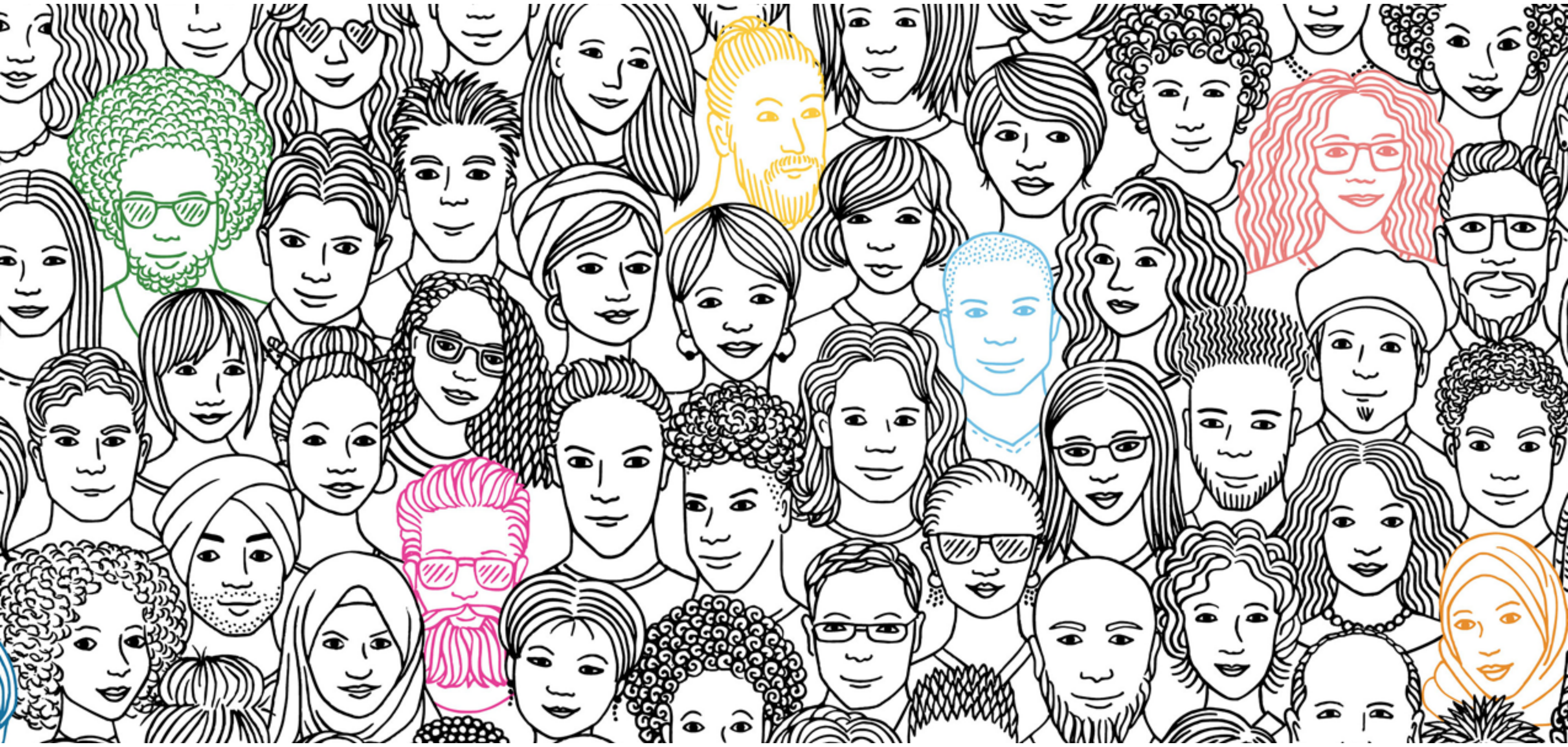
CENTER ON LABOR, HUMAN SERVICES, AND POPULATION



RESEARCH REPORT

Assessing Miscounts in the 2020 Census

Diana Elliott *Rob Santos* *Steven Martin* *Charmaine Runes*

June 2019

URBAN INSTITUTE · ELEVATE · THE · DEBATE

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is a leading research organization dedicated to developing evidence-based insights that improve people's lives and strengthen communities. For 50 years, Urban has been the trusted source for rigorous analysis of complex social and economic issues; strategic advice to policymakers, philanthropists, and practitioners; and new, promising ideas that expand opportunities for all. Our work inspires effective decisions that advance fairness and enhance the well-being of people and places.

Contents

Acknowledgments

This report was funded by the Bill and Melinda Gates Foundation. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at urban.org/fundingprinciples.

We are grateful to the many people and organizations who helped early on during the idea development for this project. Corrine Yu at the Leadership Conference on Civil and Human Rights, Bill O'Hare, and Terri Ann Lowenthal all provided helpful feedback to our team. We would also like to thank Mike Pergamit of the Urban Institute for his careful review and Liza Hagerman for her editorial work.

Assessing Miscount Risk in the 2020 Census

The decennial census is foundational to our democratic society. Census population counts guide appropriations and federal funding allocations, congressional redistricting, state and local budgets, and data-driven business and research decisions. Yet despite its importance, the 2020 Census faces unprecedented threats to its accuracy. Since 2010, decennial census preparations have faced (1) underfunding leading to scaled-back testing and outreach operations, (2) innovations that promise efficient implementation but remain undertested and underdeveloped, and (3) the last-minute introduction of a citizenship question. For this reason and others, the US Government Accountability Office (GAO) has added the 2020 Census to its "high risk" list of government activities in jeopardy in the coming years.[1]

The US Census has been in the news lately, given that the Supreme Court will decide if the citizenship question will be included before a June 2019 printing deadline. This court case and others challenging the addition of the citizenship question have heightened our collective attention to the importance that all participate and be counted in the census. While the citizenship question has garnered the most recent attention, it is not the only factor at play. A decade of underfunding, undertesting, and the introduction of new innovations to administer and impute the data suggest uncertainty about the future accuracy of census counts. Even if the citizenship question is not included in the final list of questions, current discourse about immigration could suppress participation.[2] Fortunately, there is still time to encourage participation through strategic outreach to the people and communities most likely to be missed.

At present, scant evidence exists about the effect of unfolding factors on the 2020 count, particularly for those in the population who are historically at risk of being missed. To understand how these factors could alter the 2020 count, we created assessments of the counts—overall, by state, and by demographic groups—under three scenarios, reflecting risks as low, medium, and high. We developed miscounts using the best available evidence about how different factors for the 2020 Census—including the discourse surrounding immigration and the possible inclusion of the citizenship question, diminished funding for testing, use of administrative records, and the introduction of the internet self-response (ISR) approach—could cumulatively affect the count. Considering these factors, we estimate that the count's overall accuracy could be lower than in 2010 and that some states and groups will be miscounted more than others in the 2020 Census.

This report highlights findings from our 2020 Census count assessments nationally as well as for different population subgroups and states under low-, medium-, and high-risk scenarios.[3] Key findings include the following:

- The overall accuracy of the national population count in 2020 could range from an undercount of 0.27 percent in the low-risk scenario to an undercount of 1.22 percent in the high-risk scenario. While these percentages may seem small, considering the overall US population, between nearly 900,000 and over 4 million people could be missed.
- If the 2020 performance of the census mirrored that of 2010, the national population could be undercounted by 0.27 percent because of demographic changes over the last decade.
- If the 2020 Census performs as the US Census Bureau expects, the national population could be undercounted by 0.84 percent.
- Some states may be more at risk for miscounts. For example, California has projected 2020 undercounts that could range from 0.95 to 1.49 to 1.98 percent by risk scenario (low, medium, and high risk, respectively).
- The miscounts may disproportionately affect some groups more than others. Black and Hispanic/Latinx-identified people[4] in the high-risk scenario could be undercounted nationally by 3.68 and 3.57 percent, respectively. White, non-Hispanic/Latinx people are at risk of being overcounted nationally by 0.03 percent in the high-risk scenario.
- Historically undercounted, children under age 5 are again at risk of being undercounted by up to 6.31 percent in the 2020 Census in the high-risk scenario.

BOX 1

Glossary

Throughout this report, we use technical terms—some that are also used by the US Census Bureau—which we describe below:

- **Administrative records:** These are data sources, typically pulled from other federal sources, that will be used to supplement address and resident information on the 2020 Census when there are information gaps or when households do not respond.
- **Census Coverage Measurement (CCM):** This program was conducted after the 2010 decennial census and consisted of a postenumeration survey and demographic analysis to understand how successful the census was in counting the American public.
- **Enumerators:** These are typically temporary staff, which the US Census Bureau hires in the year leading up to the census, who are tasked with visiting households that have not responded.
- **Federal Medical Assistance Percentage (FMAP):** This is the federal funding formula used to determine the percentage of each state's expenditures on medical programs that will be reimbursed by the federal government. It is a ratio of per capita state income to per capita total US income, and both depend on census counts.

- **Imputation:** This is the process of assigning data, through statistical procedures, when they are missing. This is one of the last steps in data processing before the census is finalized.
- **Internet self-response (ISR):** This option is available on the 2020 Census for residents to answer the questions online. The Census Bureau is prioritizing the “Internet First” mode for the 2020 Census.
- **Nonresponse follow-up (NRFU):** This is the period during decennial census operations when field staff, like enumerators, are sent to nonresponsive residences to conduct the count in-person. It follows a period when self-response from the American public is prioritized.

Potential Risks for the 2020 Census

As with any decennial census, issues arise during a decade that present new challenges for implementing it. In the decade leading up to the 2020 Census, the US Census Bureau had a less predictable funding stream and introduced newer technologies to help improve the count and reduce costs. Such innovations include an internet-based option to complete the census, the use of administrative records to enumerate households that fail to complete it, and an “adaptive design” approach to efficiently contact households in the field and know in real time if they have participated.

Meanwhile, events have unfolded that threaten the census’s accuracy. For example, key tests have been cancelled because of diminished funding during key ramp-up years this decade.[5] Additionally, the recent proposal to add the citizenship question to the census was so sudden and unanticipated that there was insufficient lead time to adequately test it. Further, the negative policy environment surrounding immigrants and the citizenship question is expected to suppress immigrant participation, regardless of whether the courts allow the citizenship question to be added. The push and pull of these factors could exacerbate the magnitude of miscounts. We discuss these factors and the potential risks they pose for the 2020 Census.

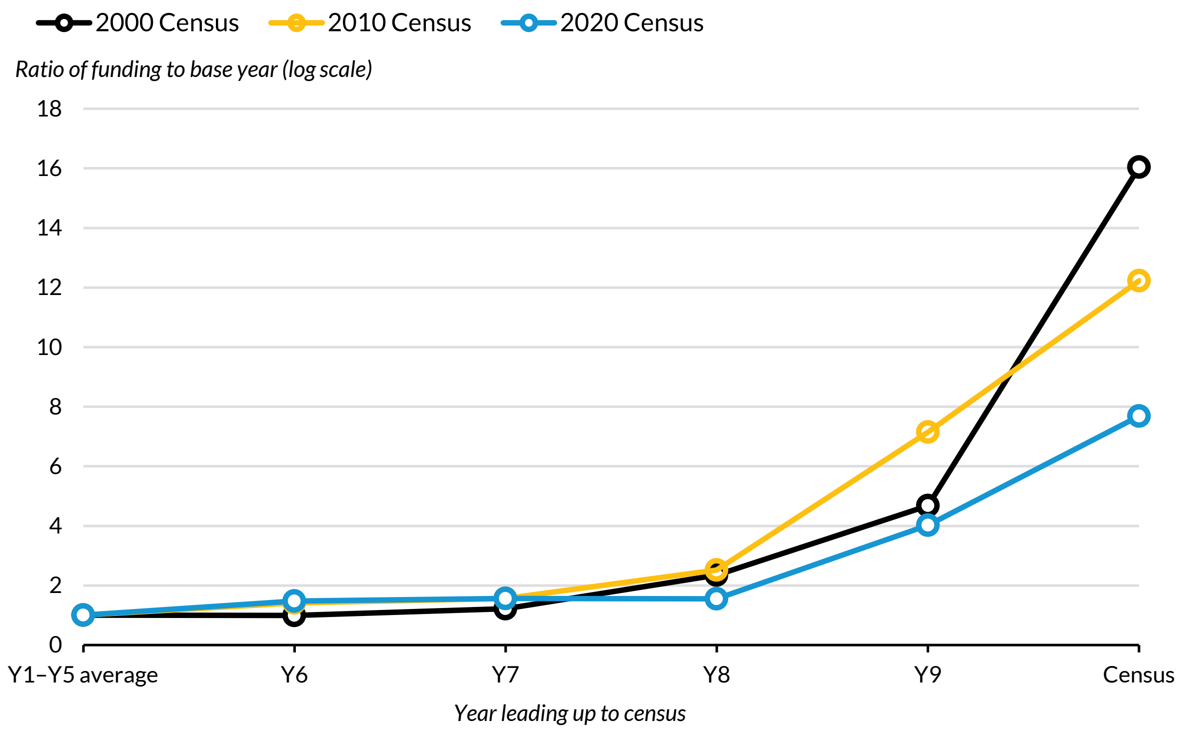
Factor: An Unpredictable Stream of Funding over the Decade

The logistics behind the decennial census are extraordinarily complicated. The US Census Bureau ramps up operations over a decade to canvass addresses, test questions and procedures, hire hundreds of thousands of workers, build partnerships, open field offices, implement new methodological and technological innovations, enumerate the country, and provide final counts to the federal government (US Census Bureau 2018b). During key testing years leading up to 2020, the US Census Bureau experienced budget shortfalls for the planning and rollout of the decennial survey (figure 1). To produce as accurate a count as possible, a predictable and adequate funding stream is necessary, which the 2020 Census has lacked.[6]

FIGURE 1

Census 2020 Funding Lags over the Decade

Authorized presidential budget for Census Bureau, 1991–2020



Source: "Budget Authority" Excel file available at "Public Budget Database," US Office of Management and Budget, accessed March 18, 2019, https://www.whitehouse.gov/omb/supplemental-materials/.

Note: Authorized budget reflects the amount of funding Congress allows a federal agency to spend.

The US Census Bureau has moved forward with the internet self-response approach (discussed in detail below), in part to improve self-response rates and reduce administrative and operational costs (US Census Bureau 2018a). However, it remains unclear if the projected cost reductions for this new technology will achieve its goals because these innovations have not been tested at scale.

The US Census Bureau also cancelled two of the three planned end-to-end tests in 2018—which were supposed to be dress rehearsals for the 2020 Census—because of inadequate funding.[7] Fortunately, the US Census Bureau has a communications and outreach budget comparable to what they had in 2010, albeit allocated in different ways (Goldenkoff 2018).[8] Investment in communications and outreach could help offset other budgetary cuts that the US Census Bureau made this decade. But we will not know if such investments will improve enumeration of the nation—especially the hard to count—until the 2020 Census has been completed.

Factor: Internet Self-Response

For the first time in the decennial census's history, there will be a universal option for households to complete their questionnaire online. The US Census Bureau anticipates that online responses will be the primary response mode for American households (US Census Bureau 2018b). While other modes of participating in the 2020 Census will be available—telephone, mail, and in-person follow-up—the "Internet First" approach is an operational priority, particularly in areas with reliable internet connectivity.

There are many benefits to the census when households "self-respond"—in other words, complete the questionnaire on their own without the need for an enumerator to collect data. The internet self-response (ISR) option could be an extremely important cost-savings mechanism for the 2020 Census; it will reduce costs related to printing and mailing, processing data, and fielding staff and operations to locate nonresponders (US Census Bureau 2018a). Further, ISR will produce higher-quality data earlier in the process; data checks will be built into the online option, ensuring that respondents answer the appropriate questions completely.

Unfortunately, self-responses to federal surveys have declined over time (Czajka and Beyler 2016). As self-response declines, the costs of conducting the census rise because of the greater workload and staffing needed to capture the data using human interaction. The size and scope of the decennial census is so immense that even small reductions in the share of self-responding households would escalate costs tremendously (US Census Bureau 2017a).

There is concern that self-response for the 2020 Census will be lower than what the US Census Bureau is planning. In 2015, those planning the 2020 Census assumed that 63.5 percent of households would self-respond to the census. This was then downgraded in 2017 to the current estimate of 60.5 percent, or just over three in five households (US Census Bureau 2017a). One of the reasons is the technology itself. The US Census Bureau has identified the ISR as a potential risk factor for the 2020 Census, citing a loss of confidence among the American public if the technology does not perform to expectations and/or if cybersecurity threats intervene (US Census Bureau 2018b).

The 2018 end-to-end test in Providence, Rhode Island, was the US Census Bureau's dress rehearsal for the decennial census. The 2018 end-to-end test had an overall self-response rate of 52.3 percent and an overall internet-response rate of 32.6 percent (US Census Bureau 2018b). Although the test exceeded the planned self-response target, it was still well below[9] the 60.5 percent self-response rate projected for the 2020 Census. The US Census Bureau's minimum estimated self-response rate for the 2020 Census is 55.5 percent (US Census Bureau 2017a), which could be a realistic outcome given recent events.

The success of ISR depends on *who* responds by internet, not just *how many* respond to it. Populations best served by an internet survey are those that are already predisposed to self-respond and be counted. In the 2010 Census, white homeowners were overcounted (for example, counted more than once at different addresses), while racial and ethnic minorities and renters were considered "hard-to-count" (Goldenkoff 2018). Experiments conducted for the American Community Survey's (ACS) implementation of an internet response option revealed that "advantaged homeowners" were the most likely to self-respond overall—and using the internet option—and those hardest to count may have actually had a reduced response rate overall (Baumgardner et al. 2014). This suggests that the ISR will be at best indirectly[10] helpful in boosting participation among a select population subset in the 2020 Census. To boost response rates among the hardest-to-count groups, the 2020 Census will spend $850 million on an outreach and communications campaign (Goldenkoff 2018). However, it is unclear if such outreach will be enough to improve enumeration among the hardest to count.

Nevertheless, the ISR—despite lower response rates than initially projected—will likely improve the efficiency and the accuracy of the census in the end. This innovation may reduce costs, which could then be used to better fund efforts to enumerate those who are hardest to count.

Factor: Using Administrative Records

Using administrative records is another innovation introduced to the 2020 Census, but it is unclear how they will improve its fairness and accuracy. Administrative records include government administrative data (for example, from federal and state agencies) and third-party data (for example, from private businesses) that are securely held and matched to US Census Bureau records to improve decennial census operations (McClure et al. 2017; US Census Bureau 2018b). While the sources of administrative records may vary, they all have one shared commonality: a person associated with an address at some point in time (Morris et al. 2016). Using administrative records has the advantage of cutting costs and potentially increasing the quality of the final 2020 Census data (McClure et al. 2017; Morris et al. 2016).

Using administrative records, largely from other federal agencies, in the 2020 Census may help in two important ways. First, they will be used to improve the identification of valid residential addresses—those that are neither businesses nor vacancies. Second, they will be used to reduce the caseload that requires enumerators to do "nonresponse follow-up" (NRFU) in the field, or the costly use of enumerators to locate households and count people who failed to complete the questionnaire on their own.[11]

Administrative records offer strong potential for improvements. If administrative records had been used in the 2010 Census to identify occupied residential addresses, the NRFU caseload would have been reduced considerably (Morris et al. 2016). This could present considerable cost savings for the 2020 Census (Rastogi and O'Hara 2012). Further, plans to use administrative records to impute data for the 2020 Census—or assign data to households and people who would otherwise be missing in the census—have the potential to be more accurate than using other statistical procedures (Rastogi and O'Hara 2012; Fernandez et al. 2018). But there are suggestions that the Census Bureau has not yet developed a complete model to impute data for missing households with administrative records (Kissam 2019). So using administrative records could improve accuracy in imputing entirely missing households if an effective method can be finalized in time.

Using administrative records is not a flawless process, however, and can introduce errors in the census data. For example, US Census Bureau researchers found in a test-case scenario that 8.5 percent of addresses deemed vacant by administrative records were legitimate and occupied residences (Morris et al. 2016). Conversely, many addresses categorized as occupied by administrative records turned out to be vacant; even an accurate net count from administrative records may produce additional error. In other words, there is some risk that administrative record use to identify valid addresses may be wrong at times and could unintentionally exclude household participation.

Also of note, not all Americans are represented fully in either census data or administrative records. Administrative records may be subject to the same population biases that make some groups hard to count in the census. Administrative records are not as robust for children as they are for adults (Rastogi and O'Hara 2012). For example, young children ages 4 and under are among the hardest-to-count groups, and in one study nearly 20 percent of children in the existing administrative records files were not present in the 2010 Census data (Fernandez et al. 2018). In such cases, administrative records may be used to improve "whole person imputation," or the assignment of data to people and households missed entirely in the 2020 Census count (US Census Bureau 2018b), but it is not yet clear what procedures the Census Bureau will use or how effective they will ultimately be.

Overall, administrative record use offers tremendous potential to reduce costs and improve quality to some extent. But it is not a perfect replacement for decennial enumeration and should be considered complementary to 2020 Census activities (Rastogi and O'Hara 2012).

Factor: Late Mandate to Add the Citizenship Question

The decennial census questionnaire is designed to be short and not too burdensome for the American public to complete, with the aim of eliciting high participation and response accuracy. The only questions asked on it are federally mandated. They include age, Hispanic/Latinx origin, race, relationship to others in the household, sex, resident tenure (whether an owner or renter), and select operational questions not used in tabulations but included to ensure accurate data (US Census Bureau 2018c). For the first time in 70 years, however, the Census Bureau was directed to add a question about citizenship to the 2020 Census.

Typically, questions are tested throughout the decade leading up to the decennial census to ensure that they are understandable to the public, accurately secure the information requested, and do not adversely affect public participation. At the very least, new questions are typically tested during the end-to-end test—the final dress rehearsal for the decennial census. Yet, the citizenship question was not submitted in time for testing; the first request to add the citizenship question was sent to the Census Bureau on December 15, 2017,[12] within only a few months of the deadline to submit final 2020 Census questions to Congress.[13] Although the citizenship question exists on the American Community Survey, the question is associated with poorer quality data and higher nonresponse relative to other comparable questions;[14] this will significantly influence the accuracy of a whole-population census.

Subsequent analyses revealed that the added citizenship question could increase 2020 Census operating costs because of suppressed participation.[15] Also, controversy over the citizenship question could discourage participation, even if the question is ultimately excluded. For example, the citizenship question was not included in the 2018 end-to-end test in Providence, Rhode Island, for the 2020 Census. However, there were still reports that despite the question's exclusion people did not want to answer the census because of immigration-related fears.[16] These same fears were documented by census researchers, who as early as September 2017 were alerting others within the Bureau of heightened concerns among immigrant respondents about participating in various surveys and tests.[17]

Hispanic/Latinx immigrants may be especially unlikely to respond. Recent studies have found that among Hispanic/Latinx-identified people the citizenship question could suppress census household participation and identification of household members among participating households (Baum et al.

2019; Kissam et al. 2019). One estimate suggests that the Hispanic/Latinx-identified population could see a differential undercount (i.e., compared with non-Hispanics) of at least 2 percent nationwide because of the citizenship question, with households either not responding or omitting members from their household roster; a similar 2-percent differential undercount is anticipated for the noncitizen population.[18] Consequently, Hispanic/Latinx people and households are likely to be underrepresented in the final counts, particularly if they identify as or reside with an immigrant. Even if the citizenship question is struck down by the courts, there likely will be residual negative affect on the Hispanic/Latinx and immigrant response rates in the 2020 Census.

Assessing Miscounts

It is impossible to know exactly how the factors described previously will affect the 2020 Census's accuracy. But with a few clear assumptions—rooted in the best available evidence—we developed informative models of what the population will be in 2020 and how well it will be counted in the decennial census. We present projected miscounts in this report—overall for population subgroups and by states—under three different scenarios reflecting low, medium, or high risk.

In this section, we describe first our method for projecting the populations of US states at census time; then our methods for simulating the 2020 Census's accuracy for the population. We present three risk scenarios and the evidence underlying our assumptions about the scenarios. Additional details about how the projections and risk scenarios were derived and estimated can be found in appendices A and B.

Projecting the Population on Census Day 2020

As a basis for our modeling, we began with a single projection of the actual population on April 1, 2020, or Census Day. We started with estimates of 2017 state populations from the US Census Bureau.[19] For every US state and the District of Columbia, we projected the population by racial and ethnic groups (non-Hispanic/Latinx white, black, American Indian and Alaska Native, Asian, Hawaiian and Pacific Islander, and Hispanic/Latinx of any race).[20] Then, within each racial and ethnic group in a state, we projected the population by age (single years). We then projected the population living in households owned or mortgaged versus rented.

Similar to the method the US Census Bureau uses to create national projections, our state-level population projections are created using a cohort-component method. The cohort-component method applies three components of demographic change—birth, survival, and migration rates—to a population and then ages that population (Colby and Ortman 2015). To illustrate the method, 27-year-olds today, who continue to live in the same place, should be counted as 28-year-olds one year from now, with a small but predictable fraction dying, moving elsewhere in the state, or moving out of state within that year. Appendix A provides additional details about the April 1, 2020, population projections.

Assessing the 2020 Census's Accuracy

After we modeled the US population on Census Day 2020, we assessed how accurately the census might count that population under three different scenarios—reflecting low, medium, or high risk—based on the assumptions underlying each one.

Our assessments rely on two basic principles. First, we deliberately use demographic categories in our projections that we know from past research on decennial censuses are counted with predictable levels of accuracy. The most reliably measured characteristics to understand the performance of past censuses are race and ethnicity, age group, and residence in an owned or rented household. For our high-risk scenario, we also project performance based on citizenship status.

Second, we draw upon publicly available source data, primarily from the US Census Bureau, to understand the accuracy of these groups' counts in past censuses, as well as to create our accuracy scenarios. Data on the accuracy of demographic groups' counts are drawn from the decennial quality assessments at the US Census Bureau, including both Census Coverage Measurement program (CCM) estimations and demographic analyses (DA) of the 2010 Census.[21] To understand the implications of different methodological and operational factors (such as internet self-response, administrative records usage, and planned protocols for nonresponse follow-up), we use evidence from tests in the run-up to the 2020 Census and official planning estimates about how people will respond to the 2020 census, conducted primarily by US Census Bureau researchers. We also incorporate assumptions based on public reports and pronouncements of recognized census experts—for example, from amicus briefs in court cases on the proposed census citizenship question.

In any census, all demographic groups will experience erroneous enumerations, such as multiple households counting the same person or a household counting a person who died before or was born after Census Day. The opposite also occurs with omissions, such as households where a person was left out of an otherwise completed census form or entire households that should have been counted but were missed. The Census Bureau also performs imputations to estimate who was missed, but this is not a perfect process either. The US Census typically has a net undercount, meaning that omissions tend to exceed erroneous counts, but that is not always the case. Further, based on population characteristics, some states are more likely to have overcounts than undercounts, while other states are likely to have undercounts below the national average.

We know from past censuses that some demographic groups are more likely to have under- or overcounts, and we use this information in our assessments. Groups that predictably have *net overcounts* include non-Hispanic/Latinx whites, people ages 5 to 17 or 50 years and older, people living in owned households, and people living in households that mail in census forms. Groups that predictably have *net undercounts* include black people, American Indians living on reservations, Hispanic/Latinx people, people ages 4 and under or 30 to 49 years old, people renting households, and people living in households that did not respond and had to be counted by enumerator follow-up. There are many other factors associated with who is hard to count in a census (see Goldenkoff 2018 for a detailed list of these factors), but the ones identified here are most salient for our analysis.

Given the evidence about (1) the predictability of census count accuracy for different demographic groups and (2) how different factors are likely to influence count accuracy in the 2020 Census environment, we created three scenarios that we designate as low, medium, and high risk. The risk levels are based on the level of accuracy each scenario is likely to produce—scenarios of the degree to which the census performance is successful and anticipated census planning parameters (e.g., expected self-response rates) are realized. Each scenario is described below, with additional details in appendix B.

SCENARIO 1: LOW RISK

If the 2020 Census performed as the 2010 Census did, what would be the outcome?

Since the 2010 Census was lauded as an operational success (Goldenkoff 2011), a similar performance for the 2020 Census could be viewed as the best that could be expected, considering the challenges faced in the decade leading up to the 2020 Census. Thus, we designate this scenario as low risk. In a sense, this scenario examines the effect of demographic change over the last decade, holding census performance constant between 2010 and 2020. In this time frame, US population age patterns, racial and ethnic distributions, and the proportion of renters have all changed, more in some states than others. From 2010 CCM evaluations, we know how often each of these groups mailed in their census forms and how accurate their counts were. We also know that nobody expects the 2020 Census to work out precisely as the 2010 performed; the Census Bureau is already counting on lower self-response (US Census Bureau 2018b). But having the basis for the low-risk scenario be the performance of the 2010 Census is a good way to show how much the US and its states have changed in ways that would make the population harder to count no matter what.

The results from the low-risk scenario show that **demographic changes alone would create a net undercount in 2020** relative to 2010. The US in 2020 has more racial and ethnic diversity than in 2010 and more renters relative to homeowners. Both factors will shift the 2020 Census toward larger net undercount, regardless of how well the census is executed. On the other hand, the US population will be older on average in 2020 than it was in 2010, and that will partly counteract the demographic trend toward larger undercount.

Under the low-risk scenario, the net undercount in 2020 would be 0.27 percent. By comparison, the census reported an official net *overcount* of 0.01 percent in 2010, although there was almost certainly a small net undercount because the CCM accuracy estimates did not fully capture the undercount of young children (O'Hare 2015). Overall, our net population undercount suggests that demographic changes alone will produce a less accurate count in 2020 relative to 2010, all else being equal.

As we will discuss, our findings show considerable variation between states and across demographic groups with respect to *who* is counted in this low-risk scenario.[22] These findings alone shed light on the prospects of a fair and accurate 2020 Census.

SCENARIO 2: MEDIUM RISK

If the 2020 Census proceeds as planned by the Census Bureau, and the operational changes work as expected, what will be the outcome?

The medium-risk scenario assumes the 2020 Census will be executed as planned. It adopts parameters from the 2020 Census Operational Plan including the presumed rates of self-response, the workload reductions associated with using administrative records, and the adaptive field protocol proposed for the NRFU process (US Census Bureau 2018b).

New operations will be employed in the 2020 Census that have not previously been used in a decennial census (US Census Bureau 2018b). This includes the internet self-response mode, which the US Census Bureau assumes will constitute most of the initial self-responses from households. Another innovation is using administrative records to help reduce the need for enumerators to locate people and households who do not respond in the first six weeks of the Census.[23] The medium-risk scenario assumes these and other aspects of the 2020 operational plan will occur as the US Census Bureau expects.

In broad terms, the medium-risk scenario makes these assumptions about 2020 Census operations:

- As in the low-risk scenario, census undercount rates for race, Hispanic/Latinx identification, age, and tenure are drawn from the 2010 Census performance assessments and are applied to demographic projections data for 2020.
- As expected in the 2020 Census Operational Plan, fewer households will self-respond (by internet, mail, or phone) than did so in 2010. Per the plan, we assume that 60.5 percent of households will self-respond within the first six weeks of Census Day by mail or internet, and

the remaining 39.5 percent will move to NRFU status (US Census Bureau 2018b). Because NRFU households are a primary source of undercounts resulting from lower quality (or no) data, this assumption is a principal driver of a larger net undercount projection for 2020.

- Demographic groups that were least likely to self-respond in the past will remain so again in 2020, proportionate to 2010.
- As expected in the 2020 Census Operational Plan, some NRFU households will be counted using matched administrative records instead of enumerators. This change will reduce the cost of the census and mitigate the undercount, but administrative records are not planned for use on much of the population (US Census Bureau 2018b). The US Census Bureau anticipates a 5-percent NRFU workload reduction from using administrative records. This translates to only 8.2 percent of NRFU-occupied households that do not self-respond who are then counted using administrative records.
- Counts made using administrative records will have a net undercount of zero, but this average will mask demographic variation in miscounts. Older age groups whose administrative records often produce duplications will be overcounted, while young children who often have no administrative records will be undercounted (Fernandez, Shattuck, and Noon 2018).

We find that if the 2020 Census plays out as the US Census Bureau has planned, there will be a net undercount of 0.84 percent in this medium-risk scenario. This means that the **expected increase in nonparticipation will be partly offset by administrative records, but the count will be more inaccurate than in the low-risk scenario**.

SCENARIO 3: HIGH RISK

If 2020 Census operational changes perform below expectations, and discourse surrounding immigration and the citizenship question further suppresses participation, what will be the outcome?

This scenario assumes that decennial operations will perform below expectations in several ways. First, we assume that combined self-response (from internet, mail, or phone) will be only 55.5 percent—the pessimistic lower bound of the Census Bureau's predicted response level (US Census Bureau 2018b). Then we reduce self-response for the Hispanic/Latinx population an additional 5.8 percentage points, which survey experts have warned could happen if the citizenship question is included in the 2020 Census (Brown et al. 2018). Even if the citizenship question is not included on the 2020 Census, we know from US Census Bureau researchers that there is an increased climate of fear and hesitation to participate among Hispanic/Latinx and immigrant residents, which makes this estimate plausible either way.[24] For these same reasons, we assume that noncitizens will systematically be missing from household rosters, even when households respond. Experts predict that such noncitizen omission (i.e., nonparticipation) could exacerbate the census differential undercount for noncitizens by 0.5 to 2.0

percentage points.[25] In this scenario, we assume the midpoint of these values, or 1.25 percentage points.

Under this scenario, we project a total population net undercount of 1.22 percent.[26] This suggests that **operational challenges and the discourse surrounding immigration would further reduce the 2020 count's accuracy beyond that of the other risk scenarios**.

This scenario is high risk but plausible for several reasons. We know that self-response to the census and other government surveys has declined over time (Czajka and Beyler 2016). Those hardest to count include people living in rural areas, those in complex households, and those with language barriers and/or privacy concerns (Goldenkoff 2018), many of whom may not be well counted in 2020. For the 2020 Census, there are added concerns that the government's position on immigration and the citizenship question will suppress the count among some already hard-to-count groups.[27] These factors could culminate in lower accuracy for the 2020 Census.[28]

Findings

Each of the three risk scenarios we assessed suggest that there will be a net undercount of the population in the 2020 Census.

Under the low-risk scenario—where we assume the 2020 Census will perform comparably to the 2010 Census—the net undercount in the 2020 Census would be 0.27 percent. This suggests that demographic changes alone will produce a less accurate count in 2020 relative to 2010, all else being equal. Under the medium-risk scenario—where we assume that the 2020 Census will perform as the US Census Bureau expects in its planning documents—the net undercount in the 2020 Census would be 0.84 percent. This suggests that operational changes will contribute to a higher risk of an undercount, above and beyond demographic changes. Finally, under the high-risk scenario—where we assume the 2020 Census will perform at the lowest expectation of the US Census Bureau and discourse about immigration and the citizenship question will further suppress participation—the net undercount in the 2020 Census would be 1.22 percent. Thus, the range of expected undercounts in the 2020 Census could be as low as 0.27 percent or as high as 1.22 percent under these scenarios (table 1).[29] In fact, 1.22 percent may be a conservative upper bound for risk if other factors disrupt census operations or the American public's confidence in the process (see the section entitled "Other Factors for Consideration").

While net undercounts are of concern to the accuracy of the census, they mask demographic- and state-level variation that suggests some groups will be miscounted at higher rates than others. For example, we find that the black population would be undercounted by as little as 2.43 percent in the low-risk scenario to as much as 3.68 percent in the high-risk scenario. Similarly, the Hispanic/Latinx-identified population would be undercounted by as little as 2.01 percent to as much as 3.57 percent in the high-risk scenario. Young children stand out as the group most at risk of being undercounted, ranging from 4.61 percent in the low-risk scenario to 6.31 percent in the high-risk scenario (table 1).

In contrast, the white, non-Hispanic/Latinx population would be overcounted by 0.03 percent in the high-risk scenario. Similarly, those 50 and older would be overcounted by 0.54 percent in the high-risk scenario. These groups have typically been overcounted in the past, and these findings suggest that this will not change in 2020 and may actually be magnified (table 1).

TABLE 1

Black and Hispanic/Latinx People and Young Children Are at Greater Risk of Being Undercounted

Percent undercount (-) and overcount (+), overall and by demographic groups, for 2020 assessed miscounts

	2010 actual	April 1, 2020, projections	Low	Medium	High
Overall	**308,745,500**	**332,092,300**	**-0.27%**	**-0.84%**	**-1.22%**
Race					
Black	38,929,300	46,934,300	-2.43%	-3.24%	-3.68%
White, non-Hispanic	223,553,300	202,182,900	0.74%	0.30%	0.03%
Asian or Pacific Islander	15,214,300	22,488,000	-0.49%	-0.97%	-1.36%
American Indian or Alaska Native	2,932,200	4,834,400	-0.55%	-1.39%	-2.12%
Hispanic/Latinx-identified	50,477,600	62,076,900	-2.01%	-2.84%	-3.57%
Age					
4 and under	20,201,400	20,466,500	-4.61%	-5.69%	-6.31%
5–17	53,980,100	53,568,800	0.60%	-0.10%	-0.56%
18–29	51,773,900	54,035,700	-0.57%	-1.10%	-1.47%
30–49	83,741,300	85,320,700	-1.72%	-2.31%	-2.71%
50 and older	99,048,800	118,700,600	1.26%	0.83%	0.54%

Source: Urban Institute calculations. See appendix B for additional details.

Notes: Undercounts are expressed as negative percentages, and overcounts are expressed as positive ones. Population counts have been rounded to the nearest hundred.

State-level estimates also show considerable variation (table 2). California, Texas, and Nevada are all states at risk of being undercounted. In the low-risk scenario, there could be a nearly 1 percent net undercount in California (0.95 percent). This only increases in California under the medium- (1.49 percent) and high-risk scenarios (1.98 percent). Similarly, Texas ranges from a 0.78 percent undercount in the low-risk scenario to 1.49 percent in the medium-risk scenario to 1.96 percent in the high-risk scenario. Nevada ranges from 0.76 percent in the low-risk scenario to 1.29 percent in the medium-risk scenario to 1.73 percent in the high-risk scenario. For these three states, our scenarios suggest that undercounts could be more pronounced than for other states. Meanwhile, states like Maine, New Hampshire, Vermont, and West Virginia have notable overcounts in the low- and medium-risk scenarios, related to demographic factors like older populations, more homeowners, and fewer people of color, which are associated with higher likelihoods of being overcounted.

TABLE 2

Nearly 2 Percent of California and Texas Residents Could Be Undercounted in the High-Risk Scenario

Percent undercount (-) and overcount (+), overall and by state, for 2020 assessed miscounts

	2010 actual	April 1, 2020, projections	Low	Medium	High
Overall	**308,745,500**	**332,092,300**	**-0.27%**	**-0.84%**	**-1.22%**
Alabama	4,779,700	4,919,400	-0.13%	-0.73%	-1.01%
Alaska	710,200	735,500	0.01%	-0.58%	-0.96%
Arizona	6,392,000	7,307,700	-0.33%	-0.95%	-1.40%
Arkansas	2,915,900	3,052,800	-0.06%	-0.61%	-0.91%
California	37,254,000	40,048,100	-0.95%	-1.49%	-1.98%
Colorado	5,029,200	5,827,500	-0.09%	-0.67%	-1.08%
Connecticut	3,574,100	3,592,200	0.01%	-0.53%	-0.92%
Delaware	897,900	987,300	-0.01%	-0.65%	-0.96%
District of Columbia	601,700	719,800	-1.74%	-2.27%	-2.68%
Florida	18,801,300	21,856,200	-0.44%	-1.04%	-1.48%
Georgia	9,687,700	10,748,800	-0.64%	-1.25%	-1.65%
Hawaii	1,360,300	1,410,400	-0.37%	-0.92%	-1.28%
Idaho	1,567,582	1,825,097	0.35%	-0.20%	-0.69%
Illinois	12,830,600	12,708,900	-0.18%	-0.78%	-1.14%
Indiana	6,483,800	6,772,400	0.26%	-0.30%	-0.59%
Iowa	3,046,400	3,197,300	0.55%	0.00%	-0.36%
Kansas	2,853,100	2,935,100	0.15%	-0.40%	-0.72%
Kentucky	4,339,400	4,518,100	0.28%	-0.22%	-0.52%
Louisiana	4,533,400	4,682,600	-0.44%	-1.05%	-1.33%
Maine	1,328,400	1,357,900	0.87%	0.36%	-0.04%
Maryland	5,773,600	6,121,600	-0.47%	-1.11%	-1.47%
Massachusetts	6,547,600	6,969,200	0.04%	-0.46%	-0.91%
Michigan	9,883,600	10,058,300	0.30%	-0.27%	-0.52%
Minnesota	5,303,900	5,733,800	0.42%	-0.14%	-0.52%
Mississippi	2,967,300	2,982,400	-0.42%	-1.06%	-1.31%
Missouri	5,988,900	6,191,200	0.23%	-0.30%	-0.56%
Montana	989,400	1,086,300	0.56%	0.05%	-0.31%
Nebraska	1,826,300	1,959,300	0.18%	-0.36%	-0.73%
Nevada	2,700,600	3,154,300	-0.76%	-1.29%	-1.73%

	2010 actual	April 1, 2020, projections	Low	Medium	High
New Hampshire	1,316,500	1,367,400	0.71%	0.21%	-0.18%
New Jersey	8,791,900	9,068,800	-0.33%	-0.89%	-1.29%
New Mexico	2,059,200	2,091,800	-0.45%	-1.25%	-1.76%
New York	19,378,100	19,857,900	-0.68%	-1.14%	-1.58%
North Carolina	9,535,500	10,607,500	-0.24%	-0.82%	-1.13%
North Dakota	672,600	758,300	0.35%	-0.17%	-0.60%
Ohio	11,536,500	11,783,900	0.17%	-0.35%	-0.62%
Oklahoma	3,751,400	3,963,500	-0.02%	-0.59%	-0.95%
Oregon	3,831,100	4,306,100	0.09%	-0.42%	-0.82%
Pennsylvania	12,702,400	12,875,300	0.27%	-0.28%	-0.58%
Rhode Island	1,052,600	1,066,900	-0.02%	-0.53%	-0.96%
South Carolina	4,625,400	5,207,800	-0.14%	-0.75%	-1.03%
South Dakota	814,200	894,900	0.49%	-0.06%	-0.50%
Tennessee	6,346,100	6,914,100	-0.03%	-0.57%	-0.84%
Texas	25,145,600	29,369,000	-0.78%	-1.49%	-1.96%
Utah	2,763,900	3,268,500	0.20%	-0.39%	-0.88%
Vermont	625,700	626,600	0.77%	0.28%	0.09%
Virginia	8,001,000	8,626,700	-0.20%	-0.78%	-1.09%
Washington	6,724,500	7,750,700	-0.05%	-0.56%	-0.97%
West Virginia	1,853,000	1,787,000	0.70%	0.20%	-0.02%
Wisconsin	5,687,000	5,875,300	0.32%	-0.19%	-0.46%
Wyoming	563,600	565,200	0.45%	-0.09%	-0.36%

Source: Urban Institute calculations. See appendix B for additional details.
Notes: Undercounts are expressed as negative percentages, and overcounts are expressed as positive ones. The estimated miscounts for Alaska (low risk) and Iowa (medium risk) are rounded to zero. Population counts have been rounded to the nearest hundred.

More concerning, however, is the considerable within-state variation that reveals some residents are more likely to be counted than others. Looking at demographic variation between the 10 most populous states, black, Hispanic, and young children are at higher risk of being miscounted than other groups (table 3). Using California as our example, in the high-risk scenario, black and Hispanic/Latinx-identified residents could be undercounted at much higher rates (3.87 and 3.65 percent, respectively) than white, non-Hispanic/Latinx residents (0.22 percent). Similarly, the undercount of young children in California in the high-risk scenario would be 7.05 percent relative to those ages 50 and older, whose count would be minimally affected with a 0.07 percent undercount.

Variation within states matters because funding is generally allocated to states based on different populations' and age groups' needs. Within states, funding may be allocated according to population counts in different communities. If young children, for example, are missed at higher rates, then communities with high shares of children ages 4 and under could miss out on receiving their fair share of funding within states (discussed in detail in the next section).

TABLE 3

Some Demographic Groups in Populous States Are at Risk of Being Undercounted

Ranges of percent undercount (-) and overcount (+), by demographics, for the 10 most populous states

	Total	Black	White NonHisp	Asian/ HPI	AIAN	Hispanic /Latinx	Ages 4 and under	Ages 5–17	Ages 18–29	Ages 30–49	Ages 50+
California (low)	-0.95%	-2.65%	0.41%	-0.57%	-0.86%	-2.18%	-5.26%	-0.12%	-1.16%	-2.35%	0.71%
California (high)	-1.98%	-3.87%	-0.22%	-1.38%	-2.30%	-3.65%	-7.05%	-1.37%	-2.16%	-3.41%	-0.07%
Florida (low)	-0.44%	-2.45%	0.85%	-0.29%	-0.55%	-1.86%	-4.96%	0.20%	-0.87%	-2.11%	1.09%
Florida (high)	-1.48%	-3.72%	0.13%	-1.11%	-2.05%	-3.43%	-6.75%	-1.05%	-1.88%	-3.22%	0.28%
Georgia (low)	-0.64%	-2.42%	0.69%	-0.35%	-0.81%	-2.08%	-5.00%	0.25%	-0.91%	-2.03%	0.97%
Georgia (high)	-1.65%	-3.80%	0.03%	-1.26%	-2.43%	-3.67%	-6.74%	-0.94%	-1.86%	-3.07%	0.20%
Illinois (low)	-0.18%	-2.34%	0.82%	-0.30%	-0.41%	-1.82%	-4.44%	0.71%	-0.49%	-1.59%	1.33%
Illinois (high)	-1.14%	-3.70%	0.15%	-1.27%	-1.96%	-3.50%	-6.17%	-0.47%	-1.41%	-2.60%	0.61%
Michigan (low)	0.30%	-2.28%	0.96%	-0.31%	-0.15%	-1.65%	-4.05%	1.20%	-0.04%	-1.15%	1.66%
Michigan (high)	-0.52%	-3.47%	0.28%	-1.17%	-1.82%	-3.42%	-5.62%	0.17%	-0.81%	-2.01%	1.03%
New York (low)	-0.68%	-2.72%	0.49%	-0.74%	-0.93%	-2.43%	-5.05%	0.26%	-0.87%	-2.18%	0.79%
New York (high)	-1.58%	-3.95%	-0.23%	-1.49%	-2.45%	-3.62%	-6.68%	-0.81%	-1.71%	-3.11%	0.08%
North Carolina (low)	-0.24%	-2.32%	0.77%	-0.54%	-0.35%	-2.03%	-4.68%	0.57%	-0.57%	-1.66%	1.27%
North Carolina (high)	-1.13%	-3.51%	0.12%	-1.38%	-2.02%	-3.66%	-6.33%	-0.52%	-1.40%	-2.59%	0.60%
Ohio (low)	0.17%	-2.47%	0.76%	-0.60%	-0.55%	-2.02%	-4.18%	1.04%	-0.17%	-1.26%	1.57%
Ohio (high)	-0.62%	-3.70%	0.08%	-1.53%	-2.23%	-3.56%	-5.70%	0.07%	-0.89%	-2.08%	0.97%
Pennsylvania (low)	0.27%	-2.24%	0.91%	-0.36%	-0.43%	-1.88%	-4.11%	1.09%	-0.07%	-1.20%	1.64%
Pennsylvania (high)	-0.58%	-3.51%	0.22%	-1.55%	-2.05%	-3.51%	-5.73%	0.03%	-0.89%	-2.11%	1.00%
Texas (low)	-0.78%	-2.47%	0.67%	-0.39%	-0.52%	-1.83%	-5.03%	0.20%	-0.99%	-2.09%	0.92%
Texas (high)	-1.96%	-3.77%	-0.02%	-1.28%	-2.10%	-3.54%	-6.97%	-1.20%	-2.12%	-3.29%	0.02%

Source: Urban Institute calculations. See appendix B for additional details.

Notes: Undercounts are expressed as negative percentages, and overcounts are expressed as positive ones. Abbreviations in the table are explained as follows: Non-Hispanic/Latinx (NonHisp); Hawaiian and Pacific Islander (HPI); American Indian and Alaska Native (AIAN).

Implications for Funding and Apportionment

Census 2020 miscounts will have significant ramifications for federal funding, congressional apportionment, infrastructure investments, and community development. In this section, we describe some of the implications for funding allocation across the states, as well as implications for state apportionment that ultimately drives redistricting and representation in Congress. Through these examples, we demonstrate the lasting effect of miscounts for the decade following a decennial census and why it is so important that the count be accurate and fair.

Federal Funding Allocations

Over 300 federal programs rely on decennial census counts or census-derived counts to distribute financial assistance equitably (Reamer 2018). In fiscal year 2016 alone, around $880 billion went to states, service providers, and other recipients through federal grants, payments, and loans (Reamer 2018). Over the course of a decade, trillions of dollars in federal funding flow based on census population counts.

Most of the federal programs that use census counts determine state funding through formulas that use eligibility criteria based on demographic characteristics. This means that the demographics of the population (age, for example) and their accuracy in the census are critical for funding allocations. This underscores the importance of both a fair *and* accurate count of all residents at the state and local levels. As discussed in the previous section, our projected undercounts suggest that some states, like California, could be deeply affected by a lack of fairness and accuracy in two ways. First, the state would be at risk of losing funding over the next decade because of differential miscounting when compared with other states that could reap more than their fair share. Second, within the state, communities with higher shares of young children—a group likely to be undercounted at higher rates—would receive less than their fair share when compared with communities with higher proportions of residents ages 50 and older—a group likely to be more accurately counted.

To illustrate with vintage 2015 data from Reamer 2018, consider five programs under the Department of Health and Human Services (HHS): Medicaid; the Children's Health Insurance Program (CHIP); Title IV-E Foster Care; Title IV-E Adoption Assistance; and the Child Care and Development Fund (CCDF). These programs are the only "dollar-for-dollar" federally funded state programs, which means that the federal government matches what the state spends according to a certain set rate each year.[30] Together, these programs accounted for almost half of all federal grants to states and about 13 percent of all state budgets, around $286 billion in FY 2015 (Reamer 2018).

These programs use the Federal Medical Assistance Percentage (FMAP) to determine funding allocations—a formula that specifies the percentage of each state's expenditures that will be reimbursed by the federal government. The FMAP is a ratio of per capita state income to per capita total US income. Both the numerator and denominator depend on population estimates derived from decennial census counts.[31] The lower the FMAP ratio, the higher the federal reimbursement percentage will be (subject to explicit exceptions, of course).

Inaccuracies in decennial census counts will affect each state's per capita income determination, and, therefore, their FMAP calculation for a decade. For instance, per capita state income would be incorrectly calculated to be lower than it truly is if that state had a pronounced undercount in the decennial census. Similarly, a state with a pronounced overcount would stand to receive more than their fair share in per capita calculations. This imbalance would persist throughout the decade, as population estimates used in per capita calculations are based on decennial counts, which are only incrementally updated throughout the decade until the next census.

Reamer (2018 illustrated the effect of an undercount on FMAP-related funding. He estimated that a one-percentage-point undercount increase in Texas's 2010 decennial population count would trigger a forfeiture of over $291 million in federal reimbursement funding in FY 2015.

Another illustration of miscounts' effect on federal funding relates to low-income families with children. Children in need are the beneficiaries of many of the largest federal programs using decennial counts and census-derived data. Unfortunately, young children (under age 5) historically have been at risk of being undercounted (US Census Bureau 2017b), and as we show in our data, they are at risk of being undercounted at high rates in our various scenarios in 2020 too. To show the effect of state undercounts on young children, Reamer (2018) estimated that a 1 percent larger undercount in the 2010 Census would have resulted in a $8.2 million decrease in Children's Health Insurance Program (CHIP) funding for Texas and almost $4.2 million for Florida. This illustrates that undercounting young children has a dramatic effect on state funding—in some cases at the magnitude of millions of dollars.

Apportionment and Representation

The decennial enumeration not only affects federal financial assistance to states, service providers, and households, but it also determines the number of seats each state gets in the House of Representatives, using the Method of Equal Proportions.[32] Article 1, Section 2, of the US Constitution describes this as the census's primary mandate. Each state gets at least one congressional seat (although the District of Columbia does not), and then every seat after that is apportioned based on a priority value. Priority values are calculated by applying a multiplier to the total state population, as enumerated in the census.

A potential undercount in some states in the 2020 Census would matter considerably. Without the benefit of the data in this report, there is suggestion that states are already at risk of losing seats. For example, preliminary research from Election Data Services used 2018 population estimates to find that states like Arizona, North Carolina, Florida, and Texas could gain at least one additional seat, while states like Alabama, Michigan, and West Virginia might lose a seat given trends already at work.[33] Given the findings we present here, the estimates of seats gained and lost could change even more if the high-risk scenario is realized.

In addition to being used to determine the allocation of seats in the US House of Representatives, census data are also used to define state legislative districts, school districts, and voting precincts (US Census Bureau 2017a). State legislative districts may be particularly subject to within-state variations. Again, the community-by-community count accuracy matters tremendously. For example, if one community happens to have an overrepresentation of groups that are at risk of being undercounted in the 2020 Census (for example, higher shares of young children and black or Hispanic/Latinx residents), then they could miss out on their fair share of funding relative to communities in the same state with higher rates of older, white non-Hispanic/Latinx residents.

BOX 2

Limitations of This Analysis

Like all such ventures, the analyses in this research report have limitations that are worth noting:

- While the projections in this report reflect the *total US population*, our miscount calculations use parameters that were developed from research on the *household* population.
- We do not make projections for the group quarters population (e.g., people residing in rooming houses; institutionalized populations such as those in prisons and nursing homes) or the emergency and transitional shelter populations.[34] These populations comprised 2.7 percent of the total population count in 2010 (Smith, Holmberg, and Jones-Puthoff 2012). Implicitly, this analysis assumes that effects on the group quarters population will be similar to those of the household population.
- We do not include Puerto Rico, other US territories, or the overseas populations (e.g., military and foreign service) in the demographic projections.
- We tried to be realistic but ultimately the risk scenarios we adopt (i.e., low, medium, high) are necessarily subjective. We assume that the 2020 Census operations will be completed successfully as planned.
- The miscount assessments we provide are quantitative but are not meant to be used for statistical inference. There are no margins of error or significance levels available. Our goal is to provide some sense of the 2020 Census's possible outcomes under a few scenarios that we subjectively deemed realistic.
- We adopt what we believe is a robust approach in developing scenarios, but we may have missed an important risk factor that will affect the 2020 Census more than the factors we considered.

Other Factors for Consideration

While this research report addresses how the citizenship question, lack of timely funding, use of administrative records, and launch of an online response option might affect the decennial enumeration, it cannot consider every potential risk in the above miscount estimates. Current unknowns such as self-response rates, changing attitudes toward technology and cybersecurity, and unpredictable events—natural or geopolitical—could all profoundly influence the final 2020 Census counts.

The decennial census largely relies on voluntary participation, despite being mandatory. In previous decades, the US Census Bureau used a mail-out-mail-back method with paper questionnaires, adding a reminder via postcard a week after the questionnaire was sent in 2000.[35] In 2010, the final mail response rate for all addresses (including those vacant and unoccupied) was 66.5 percent (Letourneau 2012). Census officials hope that the rollout of the online option will optimize self-response for the upcoming census. However, self-response across federal surveys has been on the decline for several years (Czajka and Beyler 2016). Fewer than 7 in 10 households intend to fill out the upcoming decennial survey (McGeeney 2019). Regardless of the outcome, it is possible that the final 2020 self-response rate will differ from what is currently anticipated.

In the wake of multiple data breaches, another factor to consider is the rising public concern around data confidentiality and cybersecurity. Census officials have not been able to fully test the security of the internet self-response approach because it is new to the 2020 Census. The US Census Bureau consequently identified the internet self-response mode and cybersecurity threats as potential risks in their operational plan (US Census Bureau 2018b). It is unclear if the census infrastructure will be able to support the simultaneous online participation of millions of households across the country or the safeguard against potential cybersecurity threats.[36]

Finally, there is always the possibility that an event arises before, on, or after April 1, 2020, that disrupts enumeration. If the infrastructure for an area of the country is affected by a disruption—for example, widespread natural disaster (tornadoes, flooding, earthquake)—this would affect mail and internet response as well as plans for nonresponse follow-up. For example, Hurricane Katrina changed demographic patterns[37] and, consequently, the 2010 Census rollout in affected areas.

Conclusion

Overall, it appears that an undercount in the 2020 Census is inevitable. The only question is by how much. Further, demographic differences will likely lead to differential undercounts, and, consequently, concerns over fairness. Our projections show that even under the lowest-risk scenario—where we assume that the 2020 Census will perform exactly as the 2010 Census did—the national population count will be less accurate.

Demographic and structural changes over the past decade suggest that the population in 2020 will be harder to enumerate. Groups known to be hard to count—including complex households, renters, young children, immigrants, and people of color—will represent a larger share of the population in 2020 than they did in 2010. Nationwide, response rates to surveys, even those mandated by the federal government, have declined. These trends make the challenge of enumerating the country more complicated.

To counter the challenge of enumerating a nation that is harder to count and increasingly less responsive, the 2020 Census will introduce operational changes like internet self-response to boost responses and the use of administrative records to fill in information for the missing population. Not only are these new additions insufficiently tested in a decennial census environment, but the best evidence suggests they will disproportionately improve the count of those who are already easiest to count, leaving the hard-to-count population a lingering challenge. In fact, our medium-risk scenario—where we assume the 2020 Census will perform exactly as anticipated—projects a less accurate count than that of 2010 precisely because of these changes.

Finally, political discourse about immigration and the citizenship question have created a potential chill among some groups in the country, including those who are Hispanic/Latinx-identified and immigrants. When we consider how the count could be affected by a potential chill to these groups' participation in our high-risk scenario, we project that the count will be even more inaccurate.

The concern is not only that the overall count will be less accurate, but that it will be less fair too. Fairness means that the population count truly reflects the rich diversity of the nation's population, rather than undercounting some demographic groups while overcounting others. This matters because decisions about federal funding and representation are based on the census counts. When entire communities are underrepresented in the count, they do not receive their rightful political voice or fair share of funding.

There is still time to ensure that representation in the 2020 Census is fair and accurate. By investing in outreach and engaging communities in a culturally sensitive way, there is time to ensure that representation will be better. In a democratic society, a fair and accurate count is critical and relies on all of us to participate and complete the census in 2020.

Appendix A. Methodology for Census 2020 Population Projections

The Census 2020 population projections are by state, age, and race/ethnicity for April 1, 2020 (Census Day). These projections provide the foundation for assessing 2020 undercounts in these categories compared with past US Census miscount rates.

These state-level projections are created using a cohort-component method, which applies three components of demographic change—birth, survival, and migration rates—to a population and then ages that population. This is comparable to the method that the US Census Bureau uses (Colby and Ortman 2015).[38] The idea is that 27-year-olds today, who continue to live in the same place, should be counted as 28-year-olds one year from now, with a small but predictable fraction dying, moving in state, or moving out of state within that year. The following sections describe how we produced the projections to Census Day.

Definitions of Demographic Categories Used

Race and Ethnicity

For this study, we applied a "bridged race" approach to specifying race categories that correspond with the race specifications used in the 2010 CCM study—the source of our data on historical miscounts in the census.[39] To be consistent with racial and ethnic categories published in the CCM study, we produce population projections for the total population and for the following categories: white non-Hispanic/Latinx, Hispanic/Latinx (all races), black (Hispanic/Latinx or not), American Indian and Alaska Native (Hispanic/Latinx or not), Asian (Hispanic/Latinx or not), and Hawaiian/Pacific Islander (Hispanic/Latinx or not).

Age

Age is projected by single years from birth through age 84; a single category includes those ages 85 and older.

Geography

Geography is projected at the state level.

Methodology for Creating Projections

Data

We use US Census Bureau population estimates (vintage 2017)—using midyear population estimates for 2015, 2016, and 2017—to make our state-level population projections (US Census Bureau 2018d). We use the population estimates for each state and the District of Columbia, for each year from birth to age 84 and for those ages 85 and older, and for each of the following racial and ethnic categories:[40]

- Hispanic/Latinx ethnicity, white only race selected
- Hispanic/Latinx ethnicity, black only race selected
- Hispanic/Latinx ethnicity, American Indian and Alaska Native only race selected
- Hispanic/Latinx ethnicity, Asian only race selected
- Hispanic/Latinx ethnicity, Hawaiian and Pacific Islander only race selected
- Hispanic/Latinx ethnicity, two or more race categories selected
- non-Hispanic/Latinx ethnicity, white only race selected
- non-Hispanic/Latinx ethnicity, black only race selected
- non-Hispanic/Latinx ethnicity, American Indian and Alaska Native only race selected
- non-Hispanic/Latinx ethnicity, Asian only race selected
- non-Hispanic/Latinx ethnicity, Hawaiian and Pacific Islander only race selected
- non-Hispanic/Latinx ethnicity, two or more race categories selected

The data used to calibrate our projections at the national level are the US Census Bureau's 2017 national population projections for the total United States, for five-year age categories, and for each of the following racial and ethnic categories (using the midyear population estimates for 2020)[41]:

- All Races, Hispanic/Latinx, or non-Hispanic/Latinx
- White, Hispanic/Latinx, or non-Hispanic/Latinx
- Black, Hispanic/Latinx, or non-Hispanic/Latinx
- American Indian and Alaska Native, Hispanic/Latinx, or non-Hispanic/Latinx

- Asian, Hispanic/Latinx, or non-Hispanic/Latinx
- Hawaiian and Pacific Islander, Hispanic/Latinx, or non-Hispanic/Latinx
- Two or more race categories selected, Hispanic/Latinx, or non-Hispanic/Latinx.[42]

Based on these categories, we produced a set of bridged race categories to correspond with the race categories in the CCM study.

Projection Procedure

Populations are projected separately based on the population estimates for 50 states and the District of Columbia and for the race and ethnicity categories specified above, for a total of 612 separate projections. The procedure for each projection by age is a modification of the Hamilton-Perry cohort procedure (Hamilton and Perry 1962; Swanson, Schlottmann, and Schmitt 2010). The following steps describe how we used vintage 2017 projections data from the US Census Bureau to create population projections for April 1, 2020. The process is quite technical and is broken down into the following steps:

- **Step 1**: For age $a = 0$, the population projections for midyear 2018, midyear 2019, and midyear 2020 are each assigned to be the population estimate for midyear 2017.

  $P_{(0,2018)} = P_{(0,2019)} = P_{(0,2020)} = P_{(0,2017)}$

- **Step 2**: For ages $a = 1$ through 84, the population a for 2018 is projected as the population estimate for one year younger $a\text{-}1$ in 2017, adjusted by the estimate age a in 2017 minus the estimate for age $a\text{-}1$ in 2016. The procedure is then repeated for 2019 and 2020.

  $P_{(a,2018)} = P_{(a-1,2017)} + (P_{(a,2017)} - P_{(a-1,2016)})$

  $P_{(a,2019)} = P_{(a-1,2018)} + (P_{(a,2017)} - P_{(a-1,2016)})$

  $P_{(a,2020)} = P_{(a-1,2019)} + (P_{(a,2017)} - P_{(a-1,2016)})$

- **Step 3:** For ages $a = 85+$, the population for 2018 is projected as the population estimate for ages 85+ in 2015, extrapolated to 2018 using the 2016–2017 trend. (Any negative population projections are set to 0.)

  $P_{(85+,2018)} = P_{(85+,2015)} + 3*(P_{(85+,2017)} - P_{(85+,2016)})$

- **Step 4:** Before we begin to "rake" the data or, in other words, adjust the data to known population totals, we create "prerake" projections for those ages 4 and under. The "prerake" midyear 2020 population projections are summed for each age from birth to 4 for all states and racial and ethnic groups. The US Census Bureau midyear 2020 population projection for ages 4 and under is then divided by the sum to produce a calibration ratio for that age group.[43]

- **Step 5:** Each prerake population projection for each age from birth to 4 (3,060 total cells, or state by race by ethnicity by age) is then multiplied by the ages-4-and-under calibration ratio to produce the first set of population projections we call "rake 1."
- **Step 6:** We then rake the data a second time. To do so, steps 4 and 5 are repeated for each five-year age category to produce "rake 1" population projections for *all* ages.
- **Step 7:** The "rake 1" midyear 2020 population projections are then subjected to a second rake process, "rake 2," which replicates steps 4, 5, and 6—this time for race and ethnicity for ages 4 and younger, 5 to 17, 18 to 34, 35 to 64, 65 to 84, and 85and older—from the US Census Bureau's projections data.[44] This step produces "rake 2" population projections.
- **Step 8:** The "rake 2" population projections are then summed across the whole population and calibrated against the US population projection for the total population in midyear 2020 to produce "rake 3" population projections.
- **Step 9:** The "rake 3" population projections for midyear (July 1, 2020) are then adjusted backwards to April 1, 2020 (Census Day), by subtracting from each "rake 3" population cell one-quarter of the difference between the "prerake 2020" and "prerake 2019" populations for that cell. All cell populations are then rounded to the nearest whole person.

Finally, to accommodate early census undercount analyses from a time when only one race response was possible, we created a set of "bridged race" projections. To create bridged race projections, each state's projected population with "two or more races" was partitioned into all possible combinations of race identifications, using the proportions in the American Community Survey five-year estimates for each state for 2012–2016 (Ruggles et al. 2018). Each state's projected separate "multiple race" responses were then allocated into proportions for various single races, based on bridged race allocation ratios (Liebler et al. 2008). Because bridged race estimates are subject to error proportionate to age differences in multiracial responses within each state, we performed a sensitivity check to compare distributions of multiracial permutations within each state at all ages against the distribution within each state at ages 4 and under compared with ages 5 to 85 and older. The age differences in the distributions of responses within the multiracial category varied by a few percentage points on average, but states were similar in age variation.[45]

Appendix B. Methodology for the Census 2020 Miscount Assessments

The Census 2020 miscount assessments are provided by state, race/ethnicity within each state, and age within each state. These assessments allow users to examine population undercounts under a range of assumptions about how the census is administered and how persons in the US respond to the decennial census.

Methodology: Three Scenarios Motivating the Miscount Assessments

Although it is impossible to precisely forecast the final 2020 Census counts, a practicable approach is one that posits a few plausible scenarios spanning a range of potential threats and opportunities in the enumeration process. To this end, we articulate three scenarios of increasing risk levels to the accuracy of the 2020 Census counts overall for the nation as well as for demographic subgroups. We refer to the three scenarios as low, medium, and high risk because they reflect increasing threats to the ability of the 2020 Census to accurately count all persons in the US. Below we discuss the basic assumptions underlying each scenario.

Overview

The miscount assessments for this project are synthetically derived and relatively straightforward. We take a projection of the 2020 US population by state and subject it to a 2020 Decennial Census miscount model that dictates which demographic subpopulations will be miscounted at certain rates, where

- positive miscount rates above zero denote a net undercount—positive because that is the most common miscount,
- miscount rates equal to zero designate an accurate count, and
- negative miscount rates below zero signify a net overcount.

For each state cell by race/ethnicity by age group, we adjust the projected 2020 population for that cell by its corresponding modeled 2020 Census miscount rate to produce the synthetic, model-based hypothetical population count for that cell. We employ three models that reflect varying risks of miscounting. They are described below. But first, we note the base population from which all miscounts are derived.

Base Population: 2020 Population Projection

All risk scenario models rely on the existence of a 2020 population projection. This was the first item that was addressed in the development of model-based miscount estimates.

The projection of the 2020 US total population (i.e., not a census-enumerated population) for April 1, 2020, uses the demographic projection methodology outlined in appendix A. The projection involves starting with the US Census Bureau's 2017 population estimates and then aging the population to 2020, while adjusting for migration and births and deaths. The projections are broken down by state and within state for cross-classified race/ethnicity by age groups. We also use the distribution of persons living in owned versus rented households in each state as reported in the 2017 American Community Survey (ACS). It is important to note that the demographic composition of the US is shifting over time. Specifically, non-Hispanic/Latinx whites will account for a smaller share of the US population in 2020 relative to 2010, while Hispanics, blacks, and other minority groups will have larger shares.

Low-Risk Scenario

Our low-risk census scenario combines the demographics of the 2020 population with the performance (in terms of counting) accuracy of the 2010 Census. We assume the 2020 Decennial Census counts are as accurate as 2010 Census counts for each racial and ethnic group, each age group, and each tenure status as measured by the 2010 CCM program,[46] supplemented with the demographic analysis that demonstrated the undercount of children under age 5.

In essence, our low-risk scenario asks "What if the 2010 Census performance accuracy was achieved with the 2020 population?"

LOGIC OF THE LOW-RISK APPROACH

The objective in the low-risk scenario is to develop a set of estimated miscount adjustments that can be applied to the 2020 projected population separately by state, race/ethnicity, age, and resident housing tenure (i.e., renter, owner). Since no such adjustment matrix exists, it must be built from existing research data. The data sources used were the CCM program miscount rates for the categories listed above (Mule 2012). We also used the demographic analysis (O'Hare 2015) that showed a historic undercount of children under age 5, as well as the 2017 ACS distributions of residents in owned or rented households by race, ethnicity, and age, as discussed below. The challenge is that CCM miscount rates are only available as marginal rates; that is, they are available for individual race categories, Hispanic/Latinx status, and age groups, but not for combined race by Hispanic/Latinx cells or race by Hispanic/Latinx by age cells. Our approach was to develop an additive sequential series of miscount adjustments that have had the interactive effects removed so miscount rates are not "double counted." For instance, the non-Hispanic/Latinx white and the over-50 populations both have net overcounts. Simply adding the miscount rate of non-Hispanic/Latinx whites to the miscount rate of persons over age 50 would overstate the miscount for non-Hispanic/Latinx whites over 50. This is because, in part, the reason for the overcount in the over-50 population is that non-Hispanic/Latinx whites are overrepresented in that age group. We correct for that problem.

The sequence with which the additive miscount effects were developed was

- race/ethnicity first,
- then the incremental effect of age (after removing the age and race/ethnicity combined miscount effect), and
- then the incremental effect of resident housing tenure after removing the above two effects.

The overall miscount adjustment was obtained by simply adding the adjustments above and applying them to the corresponding (race/ethnicity by age group by tenure by state) 2020 population cell total for each state. Below we describe how each miscount rate was developed.

RACE-ETHNICITY MISCOUNT RATES

Race/ethnicity miscount rates were not developed in the CCM program (Mule 2012). Instead, net miscount rates were provided separately for race groups and for ethnic groups (i.e., Hispanic/Latinx status). We developed race by ethnicity miscount rates by taking the simple average of the two net miscount rates of each specific racial and ethnic category. To illustrate, the miscount rate for Hispanic/Latinx black persons was obtained by taking the average of the net undercount for blacks (A) and the net undercount for Hispanics (B), so for this subpopulation we simply assigned a net miscount rate of C = (A + B)/2.

Next, we adjusted the resulting race/ethnicity miscount rates to account for the undercounting of those ages 4 and under, as found through demographic analysis (DA) (O'Hare 2015). Had we not done this, the CCM rate would have been incorrect because it did not fully recognize the magnitude of this component of the undercount in its CCM miscount estimates. A revised, adjusted race/ethnicity miscount rate was derived by adding a correction to the CCM miscount rate. The correction was obtained as follows. If the CCM ages 4 and under DA miscount rate is (D), the difference between it and the CCM rate was taken (E = C – D). The correction (F) is obtained by taking the difference between the 2010 CCM and DA miscount rates for those ages 4 and under (E) and multiplying it by the fraction (p) of the race/ethnicity group that was ages 4 and under in 2010, or F = (E x p). The adjusted miscount rate C* is obtained by adding the original CCM miscount rate C and the correction C* = (F + C). Mathematically, this adjustment can be readily summarized as follows:

p = fraction of ages 4 and under

q = (1 – p) = residual (i.e., 5+-year-old population)

C = the CCM-based miscount rate.

But if the CCM rate underrepresents the undercount of those ages 4 and under, then C should be lower. So, let

D = Ages 4 and under miscount rate from DA where D < C (i.e., D is smaller)

E = D – C, the difference between the DA miscount estimate for those ages 4 and under and the current estimate for a given race/ethnicity group.

The "corrected" miscount rate C* should be

$C^* = (D \times p) + (C \times q)$.

But

$C^* = (D \times p) + (C \times q) + (C \times p) - (C \times p)$

$= (D \times p) - (C \times p) + (C \times q) + (C \times p)$

$= (D - C) \times p + C = (E \times p) + C = F + C$.

Using this approach, CCM net undercount estimates were calculated for the non-Hispanic/Latinx white, black, American Indian/Alaska Native, Asian, Hawaiian/Pacific Islander, and Hispanic/Latinx populations. However, after the developing the miscount estimates for these subgroups, we collapsed the results for the Asian and HPI populations out of caution that the separate counts were too small to be reliable.

INCREMENTAL AGE MISCOUNT RATES

To obtain the incremental miscount effect of age, we began with the 2010 CCM estimates for net miscounts for five age groups—ages 4 and younger; 5 to 17; 18 to 29; 30 to 49; and 50 or older (Mule 2012)—but substitute the more accurate DA estimate for the net undercount for the ages-4-and-under group (O'Hare 2015). Then we take the race/ethnicity group percentage distributions in each age group and corresponding miscount rates (from above) and develop a "predicted" estimate of the specific miscount rate solely on the basis of the race/ethnicity composition of that age group (G). We then take the age-specific CCM miscount rate (with the ages 4 and under DA substitution)—call it H—and subtract the predicted race/ethnicity miscount rate (G) to produce the age group's incremental miscount rate with the race/ethnicity miscount rate removed: (J = H – G). This is performed for each age group. For instance, the 50+ age group has a higher proportion of non-Hispanic/Latinx whites than the general population, and because the non-Hispanic/Latinx white population has a net overcount, the 50+ age group would be expected to have some overcount simply as a result of its racial and ethnic composition. Our approach removed the miscount component from the 50+ group because of race/ethnicity.

INCREMENTAL HOUSEHOLD TENURE MISCOUNT RATES

The CCM found notable differences in miscount rates by resident household tenure (i.e., person resides in a rental versus an owned or mortgaged home). However, household tenure, race/ethnicity, and age were strongly correlated in 2010, so we needed to develop an incremental resident tenure adjustment of the net miscount beyond that because of race/ethnicity and age. Our approach was to simply use the method adopted for age (above) but applied twice—once to remove the race/ethnicity component and the other to remove the incremental age component.

We began with the 2010 CCM estimates for net undercounts for persons in rented households and net overcounts for persons in owned households (Mule 2012). For each tenure category, we calculated the miscount rate attributable to race/ethnicity using the race/ethnicity population distribution and miscount rates (K). We then subtracted the race/ethnicity miscount effect (K) from the tenure owner/renter effect (L) to produce the incremental miscount rate for a given tenure status, adjusted by race/ethnicity (M = L – K). Next came the adjustment to remove the incremental age miscount effect. For each tenure group, we used the age population distribution and corresponding incremental age miscount rates from above (J) to calculate the miscount rate attributable to age (N). We then took the race/ethnicity-adjusted tenure miscount rate (M) and subtracted the incremental age miscount rate (N) to produce the net incremental tenure miscount rate adjusted for race/ethnicity and age (P = M – N = L – K – N). For example, persons living in owned households have a higher proportion of non-Hispanic/Latinx whites than the general population, as well as a higher proportion ages 50 and older. Because the non-Hispanic/Latinx white population has a net overcount and so does the 50 and older population net of its racial and ethnic distribution, the population of persons living in owned households would be expected to have some overcount simply as a result of its race/ethnicity and age composition. We calculated an effective tenure undercount for 2010 as the observed tenure-specific undercount or overcount, minus the expected tenure-specific undercount because of the racial and ethnic composition of each tenure category, minus the expected tenure-specific undercount estimate because of the age composition of each tenure category (adjusted for race/ethnicity).

We did not attempt to make estimates of net undercounts based on living status in group quarters, which constitutes approximately 2.5 percent of the US population in 2010 as well as in 2020. We believe this to be a minor limitation of the adopted approach.

STATE

Our miscount estimates for each US state in 2020 are based on the projected race/ethnicity, age, and tenure distribution of that state's population in 2020, multiplied by the corresponding net miscount rates attributable to race/ethnicity, age, and tenure distributions as described above.

Note that while our low undercount scenario produces undercount estimates by state, those estimates are *not* derived from 2010 CCM undercount rates by state, which had too much sampling uncertainty to be used in our analysis. The procedure we use is instead a "synthetic" estimate of the undercount based on each state's demographic characteristics alone. Synthetic state undercount estimates based on state racial and ethnic distributions have been calculated and used in previous analyses of undercounts.[47] Our synthetic state undercount estimates are conceptually the same but include additional adjustments for age and tenure distributions within each state.

Note also that our 2020 estimates of the undercount are based on a mixture of DA for ages 4 and under and the CCM postenumeration survey data for all other ages. As such, it is not appropriate to compare our low-risk scenario directly with the official census undercount for 2010, which is based solely on CCM data. If an adjustment was made to the 2010 US Census undercount by using DA estimates instead of PA estimates for ages 4 and under only, the net undercount estimate for 2010 would increase from -0.01% based on CCM alone to about +.014% based on CCM and DA combined.

Medium-Risk Scenario

Our medium-risk scenario takes the low-risk scenario parameters and makes two key changes. First, we adjust our estimated miscount rates to reflect US Census Bureau expectations that 39.5 percent of US households do not self-respond to the census before launching the nonresponse follow-up effort (US Census Bureau 2017a).[48] Second, we adjust the miscount rates to reflect the use of administrative records (AR) in the NRFU to impute persons in predicted occupied households that fail to respond after an enumerator visit. According to the census operational plan, this use of AR would reduce the NRFU workload by 5 percent, and an additional 8 percent workload reduction would come from AR-predicted vacant households (that then would require no enumerator visit and be closed out as "vacant") (US Census Bureau 2018b).[49]

In essence, our medium-risk scenario asks "What if the 2020 Decennial Census goes pretty much as planned by the US Census Bureau?"

The miscount rates from the medium-risk scenario require three sets of data:

- the final estimated miscounts from the low-risk scenario, described above,
- an estimated national overall net miscount rate because of the Census Bureau's adoption of a 60.5 percent self-response target (lower than that obtained in 2010) at the onset of the NRFU effort, and
- an adjustment reflecting the use of AR to impute predicted occupied housing units in the NRFU when they failed to respond after one enumerator visit.

The first set of data already exists by virtue of the low-risk miscount scenario. The second parameter needed is the net miscount adjustment from adopting a lower self-response rate than what was obtained in the 2010 Census. We note that this adjustment is motivated by the observed correlation between self-response and net undercount (O'Hare 2018).[50] Since the actual 2010 decennial self-response was higher (Letourneau 2012) than the planned 2020 self-response (60.5%), an adjustment in the overall net miscount rate was developed relative to that in the low-risk scenario.

To begin, we used data from the 2010 NRFU Assessment (Walker et al. 2012) to distribute NRFU households to race/ethnicity, age, and tenure categories in proportion to the relative nonresponse levels for each group estimated in the 2010 Census. Hence, relative to 2010, we project how the volume of nonresponse follow-up addresses would increase to 39.5 percent overall and increase more for blacks than for non-Hispanic/Latinx whites, more for young children than for older adults, and more for renters than for homeowners.

Next, we decomposed the demographic miscount rates from the low-risk scenario into separate rates for persons by NRFU status (i.e., non-NRFU vs NRFU household), based on analysis of 2010 Census data and subject to the algebraic constraint that the combined undercounts for the NRFU and non-NRFU subpopulations must sum to the overall undercount estimate from our low-risk scenario. Then we adjusted the estimated undercount rates for each race/ethnicity category, age group, and tenure status by a quantity equal to the additional proportion of the population in NRFU households for that demographic category (since it will increase in 2020 relative to what was experienced in 2010) and multiplied by the estimated difference in undercount rates between persons in non-NRFU and NRFU households.

Throughout this scenario, we use the simplifying assumption that NRFU miscount rates at the population level correspond with NRFU miscount rates at the household level.

To account for the expected effects of the census's use of AR we use the observed distribution[51] of follow-up visits to NRFU households in 2010 to estimate that if a 5 percent workload reduction is achieved, then 8.2 percent of occupied NRFU households and assumedly 8.2 percent of the population (i.e., persons) in NRFU households will be resolved through the use of AR data with an assumed undercount rate of exactly zero. That is, we assume that AR imputations of the predicted occupied housing units in the NRFU that fail to respond after one enumerator visit will be error free.

Our medium-risk undercount estimates for each US state in 2020 are based on the projected race/ethnicity, age, and tenure distribution of that state's population in 2020; multiplied by the 2010 net undercount rates for race/ethnicity, and for age and tenure distributions as described above; adjusted for the census's predicted increase in nonresponse; then offset by the presumed accurate count from the NRFU-occupied housing units imputed via AR.

High-Risk Scenario

Our high-risk scenario is a replication of our medium- and low-risk scenarios, but with three additional changes. We assign a higher overall proportion to NRFU status, based on the Census Bureau's high variant for possible NRFU in 2020. We also make two adjustments to the net undercount that have been deemed likely by Mathiowetz[52] if the census contains a citizenship question as currently proposed by the US Department of Commerce. These two adjustments are an additional proportion of Hispanic/Latinx households in NRFU status and a small increase in the percentage of the US noncitizen population that is omitted from household rosters in mailed-in census forms.

In essence, our "high undercount" scenario asks "What would happen if the Census Bureau encounters nonresponse rates at the high end of their expectations and if the inclusion of a citizenship question brings the changes in response behavior that have been predicted?"

The initial setup for the high-risk scenario is as in the medium-risk scenario. Then we increase the predicted NRFU for the US population from the census's expected level of 39.5 percent to the census's high variant of 44.5 percent, distributing the additional nonresponse households by demographic characteristics using the same procedures as in the medium-risk scenario.

Then we further increase NRFU for the households with Hispanic/Latinx residents by an additional 5.8 percentage points, as per Brown and colleagues (2018). This adjustment reflects additional suppression of census initial response by the Hispanic/Latinx population if a citizenship question is included in the census.

Finally, we raise overall net undercount by an additional 1.25 percentage points for the US noncitizen population, as per Mathiowetz.[53] This undercount is assigned to each state, according to the fraction of noncitizens in each racial and ethnic group in each state in the 2017 American Community Survey. This adjustment reflects the omission of noncitizens from household rosters of census forms turned in by citizens who have noncitizens residing in their households.

Notes

[1] GAO (US Government Accountability Office), "GAO High Risk List: 2020 Decennial Census," GAO, March 2019, https://www.gao.gov/highrisk/2020_decennial_census/why_did_study. The GAO defines high risk as government areas that are considered "at high risk for waste, fraud, abuse, or mismanagement, or in need of transformation." For a short description, see "GAO's 2019 High-Risk List Update," US Government Accountability Office WatchBlog, March 6, 2019, https://blog.gao.gov/2019/03/06/gaos-2019-high-risk-list-update/.

[2] Center for Survey Measurement, "Respondent Confidentiality Concerns," memorandum for the associate directorate for research and methodology, September 20, 2017, https://www2.census.gov/cac/nac/meetings/2017-11/Memo-Regarding-Respondent-Confidentiality-Concerns.pdf.

[3] For the full data and interactive to see how various states and subgroups may be affected in the 2020 Census, see Diana Elliott, Robert Santos, Steven Martin, and Charmaine Runes, "2020 Census: Who's At Risk of Being Miscounted?" Urban Institute, June 2019, https://apps.urban.org/features/2020-census.

[4] The term "Hispanic/Latinx" is used throughout this report to reflect the different ways in which people self-identify. The US Census Bureau uses the term "Hispanic."

[5] Center on Poverty and Inequality–Georgetown Law, "2018 Census E2E Test Issue Brief #1: Importance and Risks," June 1, 2018, http://www.georgetownpoverty.org/wp-content/uploads/2018/06/GCPI-ESOI-E2E-Test-Memo-20180531.pdf.

[6] Diana Elliott and Charmaine Runes, "An Underfunded 2020 Census Puts an Accurate Count at Risk," *Urban Wire* (blog), March 18, 2019, https://www.urban.org/urban-wire/underfunded-2020-census-puts-accurate-count-risk.

[7] Vanita Gupta, Arturo Vargas, and Terri Ao Minnis to Sheleen Dumas, "Leadership Conference Comments on the Census Bureau's End-to-End Test-Peak Operations," The Leadership Conference on Civil & Human Rights, August 13, 2017, https://civilrights.org/resource/leadership-conference-comments-census-bureaus-end-end-test-peak-operations/.

[8] The 2020 budget for communications and outreach is $851 million compared with $828 million in 2010 (inflation-adjusted to 2017 dollars) (Goldenkoff 2018).

[9] Although the 2018 End-to-End Test adopted a significantly lower self-response target that was exceeded, the fact that a final dress rehearsal achieves a substantially lower self-response rate than was planned for 2020 raises concern.

[10] ISR is the least costly participation mode available in the upcoming 2020 Census. To the extent that ISR achieves or preferably exceeds the planned level of participation, funds will be available for NRFU enumerators to visit hard-to-count households and communities. If the ISR rate falls below the planned 2020 rate, then funds otherwise planned for the "harder to count" will have to be spread to cover the additional cost of enumerating the otherwise "easier-to-count" households that could have responded via ISR but did not.

[11] "2020 Census" Notice, 83 Fed. Reg. 248 (December 28, 2018), 67213–67222, https://www.govinfo.gov/content/pkg/FR-2018-12-28/pdf/2018-28164.pdf.

[12] Hansi Lo Wang, "How the 2020 Census Citizenship Question Ended Up in Court," *NPR Morning Edition*, November 4, 2018, https://www.npr.org/2018/11/04/661932989/how-the-2020-census-citizenship-question-ended-up-in-court.

[13] US Census Bureau, "Census Bureau Submits Planned Questions for 2020 Census to Congress," news release CB18-55, https://www.census.gov/newsroom/press-releases/2018/2020-question.html.

[14] John M. Abowd, "Technical Review of the Department of Justice Request to Add Citizenship Question to the 2020 Census," memorandum for Wilbur L. Ross, Jr., Secretary of Commerce, January 19, 2018, https://miracoalition.org/images/Documents/Census-Bureau-Jan2018memo-citizenship-question.pdf.

[15] John M. Abowd, "Technical Review of the Department of Justice Request to Add Citizenship Question to the 2020 Census."

[16] Michael Wines, "Census Officials Say Rhode Island Rehearsal Is Going Well. Not Everyone Agrees," *New York Times*, April 28, 2018, https://www.nytimes.com/2018/04/28/us/2020-census-test-rhode-island.html.

[17] Center for Survey Measurement, "Respondent Confidentiality Concerns."

[18] "Memorandum of Law in Opposition to Defendants' Motion for Summary Judgment" at 16, Kravitz et al. vs. US Department of Commerce, US District Court for the District of Maryland (no. 8:18-cv -01041-GJH) (filed November 27, 2018). https://www.brennancenter.org/sites/default/files/legal-work/MemoppdefsMSJ_2018-11-27.pdf.

[19] US Census Bureau, "State Population by Characteristics: 2010–2017," updated February 12, 2019, https://www.census.gov/data/tables/2017/demo/popest/state-detail.html#tables.

[20] We do not include Puerto Rico, other US territories, or the overseas populations (e.g., military and foreign service) in these projections.

[21] US Census Bureau, "Decennial Census of Population and Housing: Coverage Measurement," https://www.census.gov/programs-surveys/decennial-census/about/coverage-measurement.html.

[22] The state standard deviation in the low-risk scenario is 0.472. The state standard deviation is a measure of the typical unweighted difference between a state's undercount and the US undercount. Increasing the net undercount will not increase the state standard deviation if all states have the same change in undercount; it could even decrease if the undercount is increased differentially in states that previously had overcounts.

[23] "2020 Census" Notice, 83 Fed. Reg. 248 (December 28, 2018).

[24] Center for Survey Measurement, "Respondent Confidentiality Concerns."

[25] "Memorandum of Law in Opposition to Defendants' Motion for Summary Judgment" at 16, Kravitz et al. vs. US Department of Commerce.

[26] State standard deviation in the medium-risk scenario: 0.632% (this is a measure of the typical unweighted difference between a state's undercount and the US undercount.)

[27] Center for Survey Measurement, "Respondent Confidentiality Concerns"; John M. Abowd, "Technical Review of the Department of Justice Request to Add Citizenship Question to the 2020 Census."

[28] We also note that these risk scenarios and projections do not explicitly quantify effects on the group quarters (including institutionalized) populations. Implicitly, this assumes that whatever factors we adopted apply the same to both the household and group quarters populations.

[29] These findings should be interpreted with caution because they are based on the best available evidence, assumptions, and projections at the time of publication. They should not be thought of as being precise point estimates. Further, many variables could change these outcomes in the next year, for better or worse.

[30] "Federal Medical Assistance Percentage (FMAP) for Medicaid and Multiplier," Henry J. Kaiser Family Foundation, https://www.kff.org/medicaid/state-indicator/federal-matching-rate-and-

multiplier/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D.

31 Bureau of Economic Analysis (BEA) develops per capita income that ultimately relies on population estimates from the Census Bureau for the denominators of per capita income. The population estimates are derived using decennial census totals that are then updated for births, deaths, and migration. For details, see US Census Bureau, "Methodology for the United States Population Estimates: Vintage 2018: Nation, States, Counties, and Puerto Rico—April 1, 2010 and July 1, 2018," accessed March 1, 2019, https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2010-2018/2018-natstcopr-meth.pdf.

32 US Census Bureau, "Congressional Apportionment: Computing Apportionment," accessed March, 1, 2019, https://www.census.gov/population/apportionment/about/computing.html.

33 Election Data Services, "Arizona Gains Rhode Island's Seat with New 2018 Census Estimates; But Greater Change Likely by 2020," news release, December 19, 2018, https://www.electiondataservices.com/wp-content/uploads/2018/12/NR_Appor18wTablesMaps-20181219.pdf.

34 See the following table for information on the 2017 group quarters population: US Census Bureau, "Population in Group Quarters: 2017 American Community Survey 1-Year Estimates," accessed March 1, 2019, https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_1YR_B26001&prodType=table. See Smith, Holmberg, and Jones-Puthoff (2012) for information on the emergency and transitional shelter populations

35 Robin A. Pennington, Michael Bentley, and Gina Walejko, "An Overview of the 2020 Census Design: Joint Statistical Meetings 2018, Session 417: Redesigning Federal Surveys," US Census Bureau, accessed March 1, 2019, https://www.census.gov/content/dam/Census/newsroom/press-kits/2018/jsm/jsm-presentation-2020-census-design.

36 Lapowsky, Issie, "The Challenge of America's First Online Census," *Wired*, February 6, 2019, https://www.wired.com/story/us-census-2020-goes-digital/.

37 US Census Bureau, "Facts for Features: Hurricane Katrina 10th Anniversary: Aug. 29, 2015," news release CB15-FF.16, July 29, 2015, https://www.census.gov/newsroom/facts-for-features/2015/cb15-ff16.html.

38 For technical documentation related to how the US Census Bureau produces projections, see "2012 National Population Projections Methodology and Assumptions," updated November 14, 2016, https://www.census.gov/programs-surveys/popproj/technical-documentation/2012-national-population-projections.html.

39 US Census Bureau, "Decennial Census of Population and Housing: Coverage Measurement."

40 Methodology for the population estimates was accessed December 10, 2018 at US Census Bureau, "State Population by Characteristics: 2010–2017." Population estimates data were downloaded December 10, 2018, from https://www2.census.gov/programs-surveys/popest/datasets/2010-2017/state/asrh/sc-est2017-alldata6.csv.

41 See the US Census Bureau's "Projected 5-Year Age Groups and Sex Composition: Main Projections Series for the United States, 2017–2060" and "Race and Hispanic Origin by Selected Age Groups: Main Projections Series for the United States, 2017–2060" here: "2017 National Population Projections Tables," updated September 6, 2018, https://www.census.gov/data/tables/2017/demo/popproj/2017-summary-tables.html.

42 Projections data were downloaded from US Census Bureau data files on December 10, 2018, from https://www2.census.gov/programs-surveys/popproj/tables/2017/2017-summary-tables/. File "np-2017-t3.xlsx" contains projections by five-year age categories, and file "np-2017-t6.xlsx" contains projections by racial and ethnic categories, within broad groups by age.

[43] We use file "np-2017-t3.xlsx," which contains projections by five-year age categories and is located here: https://www2.census.gov/programs-surveys/popproj/tables/2017/2017-summary-tables/.

[44] We use file "np-2017-t6.xlsx," located here: https://www2.census.gov/programs-surveys/popproj/tables/2017/2017-summary-tables/.

[45] Bridged race estimates may also be subject to error proportionate to period, state, and age differences in how multiple-race respondents would *choose* a single race if asked to do so. This cannot be measured.

[46] The Census Coverage Measurement program (CCM) was a postcensus enumeration survey conducted to measure the extent of miscounts overall, by state and for some demographic subpopulations.

[47] See Section 5, page 114 of *US Census Monitoring Board: Presidential Members—Final Report to Congress*, September 1, 2001, at http://govinfo.library.unt.edu/cmb/cmbp/reports/final_report/fin_sec5_effect.pdf.

[48] See page 15, table 1 in US Census Bureau, "2020 Census Life-cycle Cost Estimate Executive Summary," December 21, 2017, https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/planning-docs/cost-estimate.html.

[49] See page 11, table 1 in US Census Bureau, "2020 Census Detailed Operational Plan for: 18. Nonresponse Followup Operation (NRFU)," April 16, 2018, https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/planning-docs/NRFU-detailed-op-plan.html.

[50] "Memorandum of Law in Opposition to Defendants' Motion for Summary Judgment" at 16, Kravitz et al. vs. US Department of Commerce.

[51] See page 61, table 18 in Walker et al. (2012) for distributions of calls to NRFU households by occupancy status to complete an interview.

[52] "Memorandum of Law in Opposition to Defendants' Motion for Summary Judgment" at 16, Kravitz et al. vs. US Department of Commerce.

[53] "Memorandum of Law in Opposition to Defendants' Motion for Summary Judgment" at 16, Kravitz et al. vs. US Department of Commerce.

References

Baum, Matthew A., Bryce J. Dietrich, Rebecca Goldstein, and Maya Sen. 2019. "Estimating the Effect of Asking About Citizenship on the U.S. Census: Results from a Randomized Controlled Trial." Cambridge, MA: Harvard Kennedy School Shorenstein Center on Media, Politics, and Public Policy.

Baumgardner, Stephanie K., Deborah H. Griffin, and David A. Raglin. 2014. *The Effects of Adding an Internet Response Option to the American Community Survey.* 2014 American Community Survey Research and Evaluation Report Memorandum Series ACS14-RER-21.

Brown, J. David, Misty L. Heggeness, Suzanne M. Dorinski, Lawrence Warren, and Moises Yi. 2018. "Understanding the Quality of Alternative Citizenship Data Sources for the 2020 Census." Hillcrest Heights, MD: Center for Economic Studies, US Census Bureau.

Colby, Sandra, and Jennifer Ortman. 2015. "Projections of the Size and Composition of the U.S. Population: 2014 to 2060: Population Estimates and Projections." Current Population Reports P25-1143. Washington, DC: US Census Bureau.

Czajka, John L., and Amy Beyler. 2016. "Background Paper: Declining Response Rates in Federal Surveys: Trends and Implications." *Final Report—Volume I.* Washington, DC: Mathematica Policy Research.

Fernandez, Leticia, Rachel Shattuck, and James Noon. 2018. "The Use of Administrative Records and the American Community Survey to Study the Characteristics of Undercounted Young Children in the 2010 Census." CARRA Working Paper Series. Washington, DC: Center for Administrative Records Research and Application.

Goldenkoff, Robert. 2018. *2020 Census: Actions Needed to Address Challenges to Enumerating Hard-to-Count Groups.* Washington, DC: Government Accountability Office.

Goldenkoff, Robert. 2011. *2010 CENSUS Lessons Highlight the Need for Fundamental Reforms.* Washington, DC: Government Accountability Office.

Hamilton, C. Horace, and Josef Perry. 1962. "A Short Method for Projecting Population by Age from One Decennial Census to Another." *Social Forces* 41 (2): 163–70. https://academic.oup.com/sf/article-abstract/41/2/163/2227276?redirectedFrom=fulltext.

Kissam, Edward. 2019. *A Cascade Model: How Latino Immigrants' Lowered Response Lead to Differential Undercount in Census 2020.* Sacramento, CA: San Joaquin Valley Health Fund.

Kissam, Edward, Richard Mines, Cindy Quezada, Jo Ann Intili, and Gail Wadsworth. 2019. *San Joaquin Valley Latino Immigrants: Implications of Survey Findings for Census 2020.* Sacramento, CA: San Joaquin Valley Health Fund.

Letourneau, Earl. 2012. *2010 Census Mail Response/Return Rates Assessment Report.* 2010 Census Planning Memoranda Series. https://census.gov/content/dam/Census/library/publications/2012/dec/2010_cpex_198.pdf.

Liebler, Carolyn A., and Andrew Halpern-Manners. 2008. "A Practical Approach to Using Multiple-Race Response Data: A Bridging Method for Public Use Microdata." *Demography* 45 (1): 143–55.

McClure, Dave, Robert Santos, and Shiva Kooragayala. 2017. "Administrative Records in the 2020 US Census: Civil Rights Considerations and Opportunities." Washington, DC: Urban Institute.

McGeeney, Kyley. 2019. "2020 Census Barriers, Attitudes, and Motivators Study Survey Report." Washington, DC: US Census Bureau.

Morris, Darcy Steeg, Andrew Keller, and Brian Clark. 2016. "An Approach for Using Administrative Records to Reduce Contacts in the 2020 Decennial Census." *Statistical Journal of the IAOS* 32 (2): 177–88.

Mule, Thomas. *Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States.* DSSD 2010 Census Coverage Measurement Memorandum Series #2010-G-01, May 22, 2012.

O'Hare, William. 2015. *The Undercount of Young Children in the US Decennial Census.* New York: Springer.

O'Hare, William. 2018. "The Likely Impact of the Addition of a Citizenship Question in the 2020 Census on Census Self-Response, Net Undercount, and Omissions Rates." Unpublished manuscript.

Rastogi, Sonya, and Amy O' Hara. 2012. *2010 Census Match Study Report.* Washington, DC: US Census Bureau.

Reamer, Andrew. 2018. *Estimating Fiscal Costs of a Census Undercount to States.* Counting for Dollars 2020: The Role of the Decennial Census in the Geographic Distribution of Federal Funds report series. Washington, DC: The George Washington Institute of Public Policy.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. 2018. *IPUMS USA: Version 8.0 [Dataset].* Minneapolis: IPUMS.

Smith, Amy Symens, Charles Holmberg, and Marcella Jones-Puthoff. 2012. *The Emergency and Transitional Shelter Population: 2010.* 2010 Census Special Reports. Washington, DC: US Census Bureau.

Swanson, David, Alan Schlottmann, and Bob Schmidt. 2010. "Forecasting the Population of Census Tracts by Age and Sex: An Example of the Hamilton–Perry Method in Action." *Population Research and Policy Review* 29 (1): 47–63.

US Census Bureau. 2018a. *2020 Census Detailed Operational Plan for: 12. Internet Self-Response Operation (ISR): A New Design for the 21st Century: Version 1.0.* Washington, DC: US Census Bureau.

US Census Bureau. 2018b. *2020 Census Operational Plan: A New Design for the 21st Century: Version 4.0.* Washington, DC: US Census Bureau.

US Census Bureau. 2018c. *Questions Planned for the 2020 Census and the American Community Survey: A New Design for the 21st Century: Version 1.0.* Washington, DC: US Census Bureau.

US Census Bureau. 2018d. "Methodology for the United States Population Estimates: Vintage 2017: Nation, States, Counties, and Puerto Rico—April 1, 2010 to July 1, 2017." Washington, DC: US Census Bureau.

US Census Bureau. 2017a. *2020 Census Life-cycle Cost Estimate Executive Summary: A New Design for the 21st Century: Version 1.0.* Washington, DC: US Census Bureau.

US Census Bureau. 2017b. *Investigating the 2010 Undercount of Young Children—Analysis of Census Coverage Measurement Results: A New Design for the 21st Century: Version 1.0.* Washington, DC: US Census Bureau.

Walker, Shelley, Susanna Winder, Geoff Jackson, and Sarah Heimel. 2012. "2010 Census Nonresponse Followup Assessment." Washington, DC: US Census Bureau.

About the Authors

Diana Elliott is a senior research associate in the Center on Labor, Human Services, and Population at the Urban Institute. She is a mixed methods researcher whose work spans local level studies of financial well-being to national level work using credit and large government data sets to better understand social policy. Prior to Urban, she was the research manager for Pew's financial security and economic mobility work and was a family demographer at the US Census Bureau. Elliott holds a PhD in sociology from the University of Maryland, College Park.

Robert Santos is Vice President and Chief Methodologist at the Urban Institute. He is an expert in social science/policy research and statistical design. Rob is an elected American Statistical Association Fellow as well as a recipient of ASA's highest recognition for distinguished service and leadership, the Founder's Award. He will be the 116th President of the American Statistical Association for 2021, and was a past president of the American Association for Public Opinion Research and a past vice-president of the ASA.

Steven Martin is a senior research associate at the Urban Institute with a specialization in family demography. He also does work on numerous topics related to quantitative analysis of policy outcomes such as analyses of event histories of interdependent events, time use, quality of life measures, attitudinal measures, and a variety of other work with survey, census, and vital statistics data. Dr. Martin has also published on data quality, including the analysis of predictors of response error for different types of respondent recall under different survey formats. Martin holds a PhD in sociology from the University of Wisconsin.

Charmaine Runes is a research analyst at the Urban Institute's Center on Labor, Human Services, and Policy. She is a mixed methods researcher, focusing on multigenerational antipoverty interventions that support and empower disadvantaged youth and low-income working families. She also works on projects that elevate community-based participatory research methods and the role of structural racism in public policy. Runes graduated with honors from Macalester College with a BA in economics.

Statement of Independence

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.

URBAN INSTITUTE · ELEVATE · THE · DEBATE

500 L'Enfant Plaza SW
Washington, DC 20024

www.urban.org