Hard-to-Count Scores and Patterns of 2020 Census Response Rates: A Case Study of California

J. Gregory Robinson, Independent Researcher, Edward Kissam, Werner-Kohnstamm Family Fund, William O'Hare, O'Hare Data and Demographic Services LLC

Introduction

In a seminal study of patterns of response rates in the 2020 census, O'Hare and Lee (2021) document the variations in 2020 response rates across demographic groups—and the correlation of the 2020 patterns with the differentials of previous censuses.[1] In a concluding section they note "this study sets the stage for more detailed data analysis of tracts within a state or one metropolitan area to see if the national patterns found here are seen uniformly across the county". This paper utilizes the rich demographic and socioeconomic data in the Census Bureau's Planning Database (PDB) to perform this analysis for all tracts in California.

This paper lays out the development of the variables in the PDB (guided by analyses of response levels in prior censuses as well as by the ethnographic studies surrounding the 1990 census). The PDB assembles a range of housing, demographic, and socioeconomic characteristics that are correlated with non-response and census and survey coverage. Importantly, the HTC Score is a summary indicator by which small-area geographies (e.g., census tracts) can be sorted on a hard-to-count continuum based on local community characteristics related to enumeration difficulty.[2]

Analysis of Census 2020 self-response in relation to selected community characteristics and the composite California HTC score developed by the state's Department of Finance provides valuable insights both for assessing quality of the 2020 census data and for future planning. This opportunity stems, in part, from California being very diverse in terms of race/ethnicity and socioeconomic characteristics but, also, because the state relied on HTC scores to guide a robust "Get Out The Count" campaign to improve census response in historically low-response areas. California is also a state with large concentrated urban populations as well as rural areas and uneven availability of Internet, an important issue since online census response was a new mode of data collection in Census 2020.

Organization of the Analysis

We begin with a brief overview of the conceptual tools used to analyze census response rate, the original Census Bureau HTC scores, the California HTC index, and the Planning Database, a broad compendium of data of potential use for Bureau planners as well as a broad range of census stakeholders seeking to understand and improve census response.

We then discuss the relationship between response rate and final census coverage after non-response followup (NRFU) and subsequent data processing have been completed. Trends in census response rates from 1990-2010 are reviewed in order to provide historical context for our observations about 2020 response. This analysis contributes to debates and discussion about the extent to which varying patterns of Census 2020 response are similar to those in the past and the extent to which the introduction of the online response option and the unforeseen modifications to Census 2020 operations required due to the COVID-19 pandemic affect response. We then go on to review overall California patterns with a focus on the striking differences in self-response level between Hispanic-majority tracts and Non-Hispanic-majority tracts. We also discuss differences in level of response between several important sub-state regions. These analyses demonstrate the limitations to state-level and national-level analyses of Census 2020 response since the sub-state data show dramatic differences in response.

We also analyze the relationship between Internet connectivity and Census 2020 self-response and implications for differential undercount and for efforts to increase "digital equity" as part of planning for Census 2030.

We conclude by highlighting some implications of our analysis both for assessing the reliability of Census 2020 data as a basis for equitable allocation of census-driven federal, state, and philanthropic funding, as well as for business and local government planning. We also highlight some implications for Census 2030 planning.

The Census Planning Database and Hard-To-Count (HTC) Scores

What Is the Planning Database?

Decennial census planning has relied for several decades on the Planning Database (PDB) which now incorporates data drawn both from the American Community Survey (ACS) and the previous census. The PDB is publicly accessible and is an important organizational acknowledgement of the Census Bureau's commitment not only to internal research for systematic improvement of its survey research but also to its' partnership with external census stakeholders in efforts to maximize census response.

What Are Hard To Count (HTC) Scores?

The Census Bureau has conducted research on response patterns for many decades. An important data element in the Planning Database is a hard-to-count index which was first developed for use in the 2000 decennial census and began incorporating ACS data (which had replaced the earlier "long form" census data with detail data on demographic and other community characteristics) following the 2010 census[3].

The original vision in developing the HTC framework was that it would be possible to adapt it to incorporate what might possible be different factors affecting response levels in different sorts of neighborhoods, census tracts, and communities in order to optimize local planning.[4] However, there has subsequently been more attention given to a generic hard-to-count index called the low-response score (LRS) which was developed by the Census Bureau based on analysis of national response patterns.[5]

HTC scores are used in an index to assess how difficult a tract is expected to be to enumerate and, thereby, to guide interventions to overcome anticipated problems. While HTC scores are often thought of as an indicator of decennial census response in a tract they actually are relevant in assessing overall difficulty of enumeration, including challenges faced in conducting non-response followup (NRFU) since the ethnographic research suggests that the same type of factors that affect self-response may also affect response to enumerators in NRFU.[6] It is reasonable to expect HTC scores to be a useful tool in analyzing and operational planning to increase ACS response.

Hard-to-countness is most often thought about in relation to the longstanding phenomenon of differential undercount of racial minorities and has been documented via post-enumeration survey (PES) methodology for decades. However, the algorithms used to construct a HTC score do not rely on race/ethnicity but on a number of "structural" community characteristics associated with low response such as: high proportions of renter households, high levels of poverty, proportion of heads of household without a high school education, proportion of multi-unit housing structures, crowded housing. HTC indices can also incorporate variables to describe census "treatment"/mode of data collection for different geographic areas—since the research is clear that level of response is linked to census system factors as well as respondent motivation and community context.[7]

Our analysis here is based on the California HTC score that was designed to be similar to the original Census Bureau HTC score but modified to incorporate characteristics believed to be particularly important in determining census response in California. Percent of limited-English households, percent of households headed by a foreign-born person, and percent of households without a broadband subscription were added to the variables in the original HTC index developed by the Census Burau.[8] The California HTC tract-level scores were constructed relying on 14 variables, most of them drawn from the most recent ACS 5-year data prior to 2020, although one—broadband connectivity—was drawn from California Public Utility Commission data. The California HTC tract scores and the resulting index of hard-to-countness, notably, do not incorporate data on racial/ethnic characteristics of tracts.[9]

Using the HTC scores, tracts can be systematically sorted on a "HTC continuum". Our analysis of patterns of HTC scores and self-response rates looks at quintiles or deciles (from hardest to easiest to count) and corresponding tract-level self-response rates (from lowest to highest level of response). Because it needed to be developed well before the 2020 Census took place the California HTC index, as well as our analysis, rely on ACS 2013-2017 data in the Census Bureau's planning database and use Census 2010 tract geographic designations.[10] The range of California HTC scores is from 1 to 124 with the highest scores representing the hardest-to-count tracts.[11]

California Diversity and Relationship Between HTC Score and 2020 Self-Response

California is the largest and most diverse state in the U.S. with a "minority-majority" population (40% Hispanic, 35% Non-Hispanic White,16% Asian, 6% Black, American Indian/Alaska Native) of close to 40 million.[12] It also has the most immigrants of any state in the nation, 10.5 million, about one quarter of the state's total population.[13] It is socio-politically and geographically diverse with 58 counties, including ten major urban areas, large swathes of rural agricultural production in counties that are composed mostly of urban clusters, as well as predominantly remote rural counties with mountainous or desert terrain. It is reasonable to expect that hard-to-countness and resulting Census 2020 self-response will vary greatly among California's 8,000 or so populated tracts.

Table 1 below shows the differing characteristics of California's harder and easier to count tracts, the neighborhoods they are composed of, the communities they make up and how these characteristics are related to 2020 census self-response[14] Our analysis looks at Census 2020 response in relation to HTC quintile (with Quintile 1 representing the hardest-to-count tracts and Quintile 5 representing the easiest-to-count tracts).

Table 1
Characteristics of California Tracts By HTC Quintile

HTC Quintile	Average HTC Index	2020 Census Self-Response Rate	%Renter Units	%Not High School Grad	%Ling-uistic Isolation	%In Poverty	% of HHs with Internet Access
California Average	***40***	***70.3***	***45.5***	***17.4***	***9.2***	***15.0***	***85.1***
Quintile 1	86	60.0	73.9	37.0	21.1	29.9	72.5
Quintile 2	56	65.8	58.8	22.1	11.1	18.8	82.4
Quintile 3	37	69.1	46.6	15.0	7.3	12.9	86.3
Quintile 4	22	74.1	32.5	10.3	5.4	9.0	89.6
Quintile 5	8	81.1	21.1	6.6	3.5	5.9	92.6

Source: U.S. Bureau of the Census, 2019 (Planning Database), 2020 Self-Response Rates (October 16 date), State of California, 2020 (Hard-to-Count Index)

As can be clearly seen in **Table 1**, there is, as expected, a strong negative association between HTC score and Census 2020 self-response rate, despite major investments by California state government and philanthropy in efforts to maximize census response in these more socioeconomically disadvantaged neighborhoods and communities.

Our analysis here highlights the strikingly different characteristics of the populations at the opposite ends of the HTC scale. The populations living in the "hard-to-enumerate" Quintile 1 tracts have much higher levels of rentership, language isolation, and poverty than the "easier-to-enumerate" Quintile 1 tracts, and lower levels of education and internet access.

It is important to note that all of these factors associated with differences in self-response rate are "structural"; that is they reflect the varying socioeconomic characteristics of local areas but, also, mode of census data collection since tracts with higher levels of Internet connectivity were offered "Internet First" and lower ones were offered "Internet Choice".[15] As can be seen in the table, various structural characteristics associated with higher or lower self-response are co-variant, but because they are not perfectly correlated, the HTC score which incorporates several provides a better predictor of tract-level self-response than any single characteristic.

Linguistic isolation is, essentially, a proxy for concentrations of immigrants because, although some US-born individuals are limited in English, they represent a very small proportion of the overall limited-English households.

The proportion of households in a tract with less than a high school education is, inevitably, co-variant with prevalence of poverty since educational attainment strongly affects earning power. However, educational attainment is, also, a proxy for concentration of immigrants. Although youth of all race/ethnic groups drop out of high school, high school completion rates in the U.S. and in California have been rising for the past several decades. The vast majority of California heads of household lacking a high school education are Mexican and Central American immigrants, although there are smaller populations of immigrants from other countries who also were unable to secure a high school education.[16]

The self-response rate gap between HTC Quintile 1 (the lowest-responding tracts) and HTC Quintile 5 (the highest-responding tracts) is substantial: 21.1%. This gap is useful to keep in mind in assessing California state-level self-response rate in relation to other states across the country. While there was very high level of self-response in some communities, there was also very low Census 2020 response in others.

There is solid empirical evidence that California's investment in "Get Out the Count" efforts did mitigate low response in the hardest to count tracts[17]; however, our analysis shows that serious disparities remained. Although California's overall 2020 census response rate was 69.6%, the range of self-response rates for quintiles of tracts within California (from 60% to 81%) is comparable to the range of state-level self-response rates (from 54.7% in Alaska to 75.1% in Minnesota).[18]

Interpreting Observed Patterns of Self-Response in Census 2020: The Distinction Between Response and Return Rates

The best metric for analyzing census response at any geographic level is the percent of responses from occupied housing units. Two metrics representing respondent participation in the census have been routinely reported historically—mail response rates and mail return rates. Response rates are based on the number of questionnaires returned divided by the number of forms/online response invitations sent out to all addresses/MAF IDs, while return rates are based on the number of forms divided by the number of occupied units. Consequently, response rates are always lower than return rates because vacant units are still included in the denominator and the Census 2020 reports of tract-level self-response rate (online or by mail) are comparable to prior censuses' mail response rates.

While return rates are preferrable as measures of respondent participation, only response rates are available until much later in the census cycle. For the 2020 census, self-response rates were published on a daily basis during the enumeration period from March until October, 2020 (the October 16 data are used in this study). The release of tract, state, and national return rates for 2020 has not been scheduled. Consequently, the primary uncertainty related to our current analysis of response patterns is that there may eventually turn out to be a difference in vacancy rates between quintiles of tracts on the HTC continuum. However, in California, the vacancy rates do not differ significantly across the five HTC quintiles and did not change appreciably from 2010 to 2020, so the pattern of change detected by the response rates used here should mirror the pattern of change measured by the return rates.

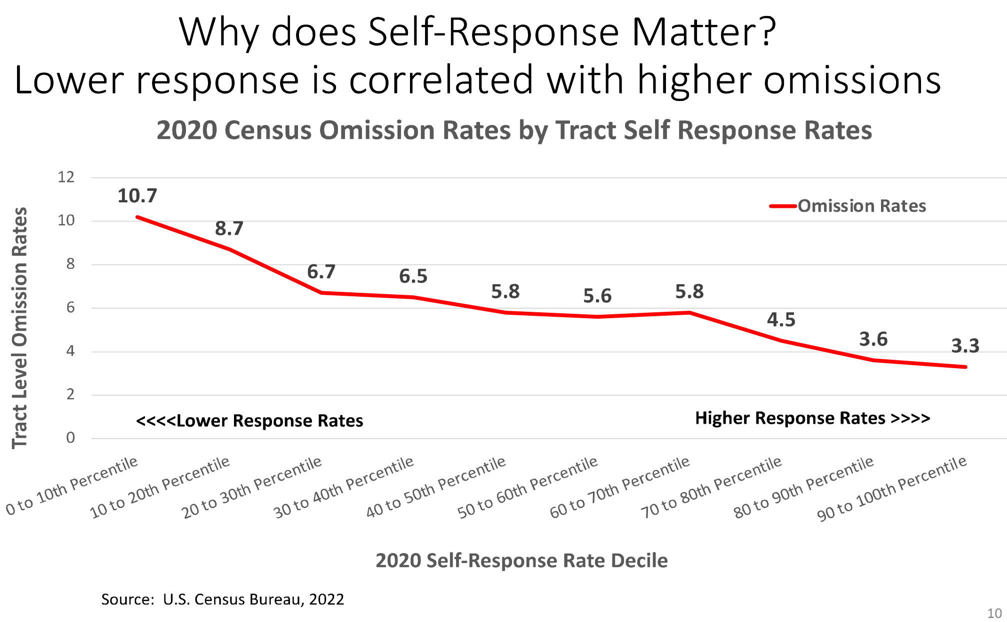
The Association Between Census 2020 Self-Response and Census Omission

Figure 1 below shows there is a consistent relationship between level of self-response in Census 2020 and census omissions in final tabulations of census data.[19] In the final Census 2020 tabulation, households were more often omitted in the lowest-response tracts than in the highest-response tracts. This is measured via the Post-Enumeration Survey (PES).

Non-response followup (NRFU) is a ubiquitous component of survey research and has been an important and expensive census operation for decades. While the highest possible level of response is desirable in any survey and in the decennial census, the key issue in assessing the utility of tabulations for the purpose of census-driven allocation of funding is whether census non-response is random or whether it leads to differential undercount of racial/ethnic minorities or, for that matter, any distinct and disadvantaged sub-population and the extent to which NRFU can mitigate the problem of systematic non-response.

Basically, **Figure 1** shows that NRFU in 2020 did not successfully "cure" the problem of low self-response with the likely result being significant differential undercount—associated with low response level in the high HTC/low-response tracts. This then "translates" patterns of self-response into patterns of differential net undercount, with higher net undercoverage in high HTC/low-response tracts and lower net undercoverage (or net overcounts) in low HTC/high-response areas.

Figure 1



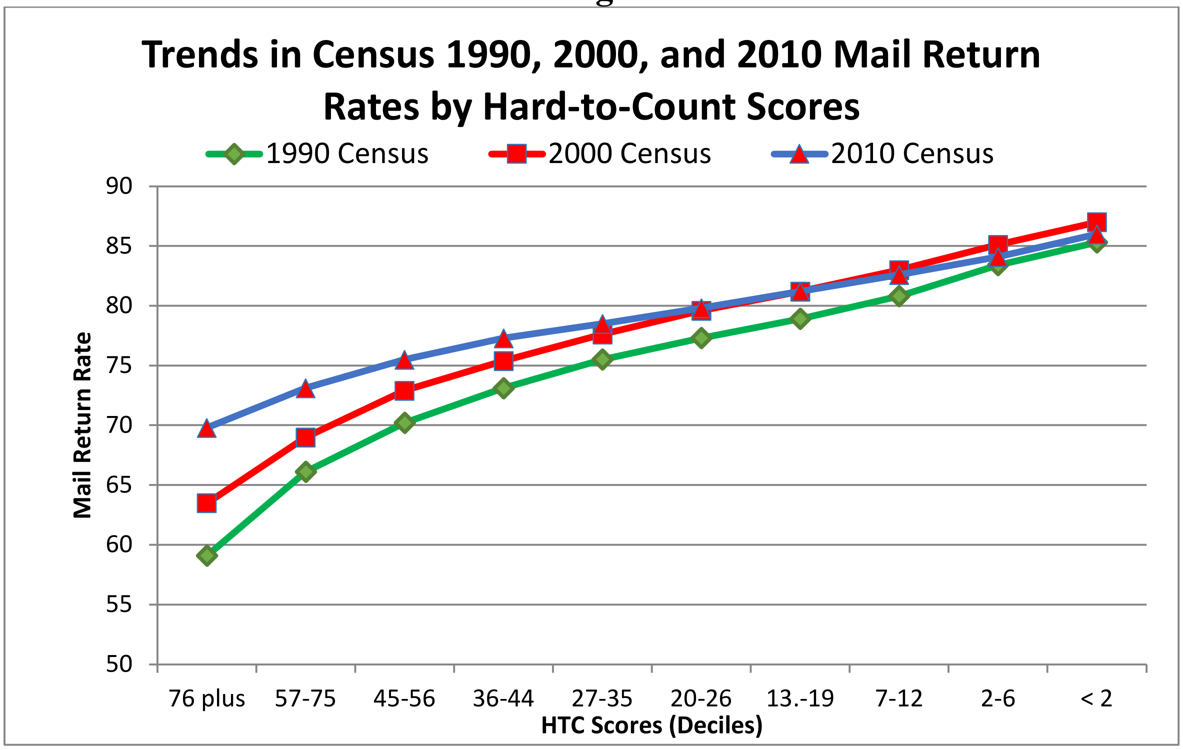
The strong association of census response rate and net undercount patterns will be explored more in a later section of this paper.

Census 2020 in Historical Context: Progress Toward A More Equitable Census

Historically, the quality of decennial censuses has improved from decade to decade since 1990. **Figure 2** below shows the patterns of census mail return rates from 1990-2010.
Overall return rates increased fairly steadily from 1990 to 2010 (from 75.3% to 79.3%)[20] Most important to recognize for census equity is that the gap between the hardest-to-count census tracts and the easiest-to-count has narrowed significantly.

The Y axis on left of **Figure 2** gives the mail return rate—ranging from 50 to 90 percent; the x-axis below represents all tracts sorted in HTC deciles, from the <u>decile with the highest HTC scores</u>— (the left most decile) to that with the lowest HTC score— (right most decile).
We see the clear inverse association in this figure—low return rates in tracts with high HTC scores (left side in **Figure 2**) and relatively high return rates in areas with very low HTC scores (right side). This pattern holds for each census from 1990 to 2010.

Figure 2



Source: Bruce, Robinson, and Devine, 2012

The key point to note here is the <u>change</u> in the pattern of the response trend lines over the 20-year period. Return rates improved by 10 percentage points from 1990 to 2010 in the "Hardest-to-count" decile of tracts while the return rates improved by less than 1 percentage point in the "easier-to-enumerate" tracts on the right.

It has been widely reported that Census 2020 achieved the highest national self-response rate ever. However, the question which arises from looking at the decade-to-decade trends in **Figure 2** is what happened with respect to progress in closing the response gap between the hardest-to-count and the easiest-to-count tracts. This is a substantive issue for census equity since the differences in census quality across groups are the most important metric for equitable apportionment and allocation of funding.

The current question: What happened to response rate patterns of change in the 2020 census on the Hard-to-count continuum? Our case study for California begins to answer this question.

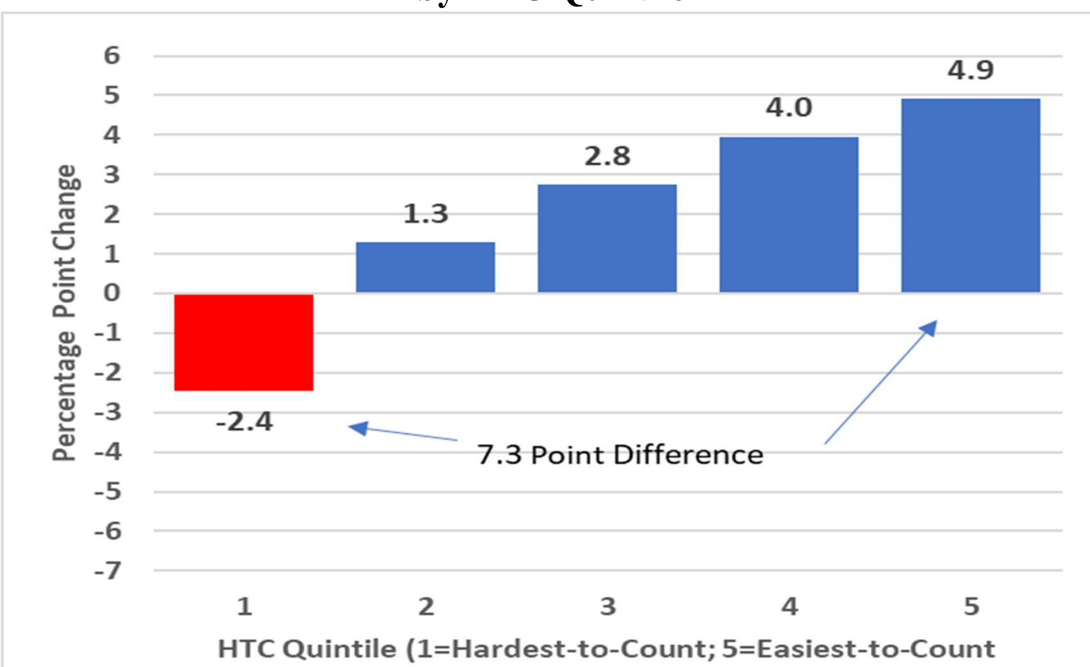
California Self Response in Census 2020 Compared To Self-Response in Census 2010

Although state-level tabulation of census response shows that overall census response in California increased in Census 2020 (by 1.4 percentage points from 68.2% in 2010 to 69.6% in 2020), **Figure 3** below shows that the increase was uneven and that response improved much more in the easiest-to-count tracts than in the hardest-to-count ones—despite the very sensible "Get Out the Count" campaign focus on promoting self-response in the hardest-to-count tracts.

There is a gap of 7.3 percentage points in the self-response change between hardest and easiest to count tracts in California. As can be seen in **Figure 3** below, the 2020 self-response rates have declined from 2010 in HTC Quintile 1 (by 2.4 percentage points), while increasing appreciably in the "easiest to enumerate" quintile—by almost 5 points.

This pattern is reversed from that observed at the national level from 1990-2010. The 2020 pattern contrasts markedly from that observed for the 1990-2010 period shown in **Figure 2**, where response rate increased the most in the hard-to-count tracts and much less in the easier-to-count tracts.

Figure 3
Change in California Self-Response Rates from 2010 to 2020
by HTC Quintile



Since level of self-response in a census tract is correlated with census coverage (as shown in **Figure 1**), census equity is actually better assessed by examining the gap in self-response between harder-to-count and easier-to-count tracts than by looking at aggregate self-response rate at the state level.

California's "Get Out The Count" (GOTC) campaign, coupled with the introduction of the online response option, may well have bolstered overall census response for the state. It mitigated worries that 2020 would result in loss of federal representation due to low response--since state-level aggregate self-response did increase. But it is important to recognize that, at the same time, despite a well-funded and well-targeted campaign there continued to be lower response in the more disadvantaged areas. A recent study of GOTC campaign impact in the southern San Joaquin Valley (Lee and Ito 2023)[21], a particularly hard-to-count sub-state region, showed that the campaign had mitigated but not eliminated the gap in response between easier-to-count and harder-to-count tracts. It appears that the online response option worked well in areas with high levels of broadband connectivity and increased their already-high response rates further.

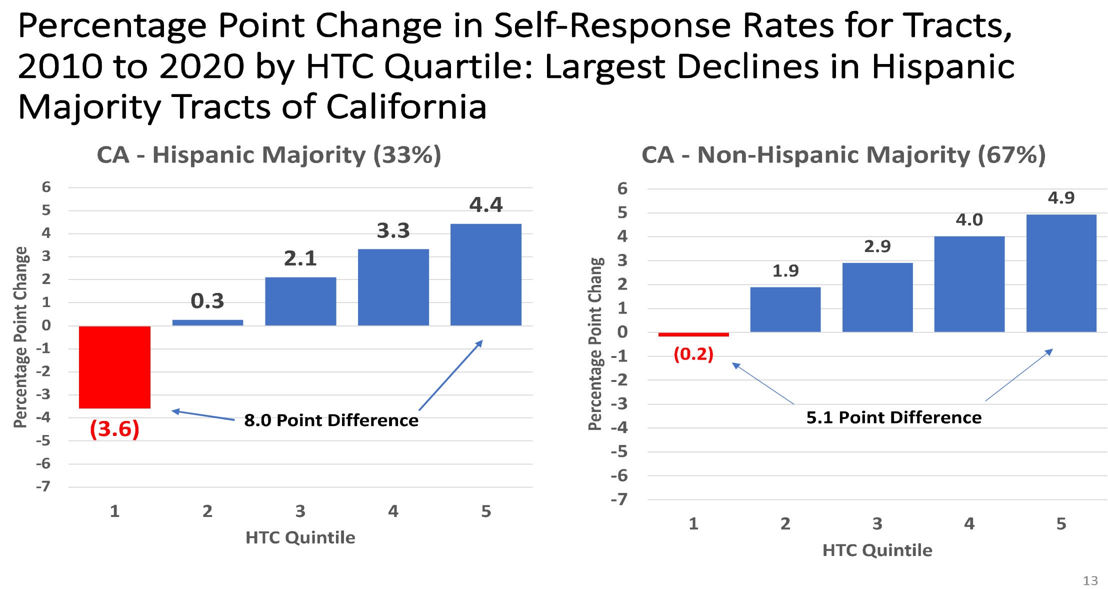
California 2020 Self-Response In Hispanic-majority and Non-Hispanic Majority Tracts

An obvious question that arises from the observation that self-response improved less in the harder-to-count than in the easier-to-count tracts is the extent to which the 2020 self-response patterns are likely to have improved or eroded equity for Hispanics—California's largest racial/ethnic population but one where socioeconomic and sociopolitical inequities persist, particularly for the low-income hard-to-count neighborhoods and communities with concentrations of Hispanic immigrants.

In **Figure 4** below, we examine the change from 2010 to 2020 in response rates for Hispanic and Non-Hispanic majority tracts in each quintile of tracts as ranked by HTC score. Hispanic majority tracts (in the left-hand panel) are those where the Hispanic population comprises more than 50 percent of the total tract population; all other tracts are referred to as Non-Hispanic majority tracts (in the right-hand panel).

Here, in looking at self-response patterns through the dual lens of race/ethnicity and socioeconomic advantage or disadvantage, there is, as for the state overall, a large difference in the change in the 2020 census response rates from 2010, but the gap is even wider for Hispanic majority tracts. For Hispanics (left-hand figure), the response rate in Quintile 1 **fell 3.6 percentage points** (but fell "only" 0.2 percentage points for Non-Hispanic majority tracts).

Figure 4



These differences by race/ethnic origin shape the patterns of response rate change along the HTC continuum illustrated in **Figure 4**. Since HTC score is based on multiple indicators of social disadvantage such as poverty, low educational attainment, and limited English, this highlights the persistent correlation between low census response and social disadvantage. The self-response patterns depicted in **Figure 4** indicate that inequity in allocation of federal and state funding resulting from lower response in the HTC Quintile 1 tracts, neighborhoods, and communities is most burdensome for jurisdictions with concentrations of low-income Hispanic households.

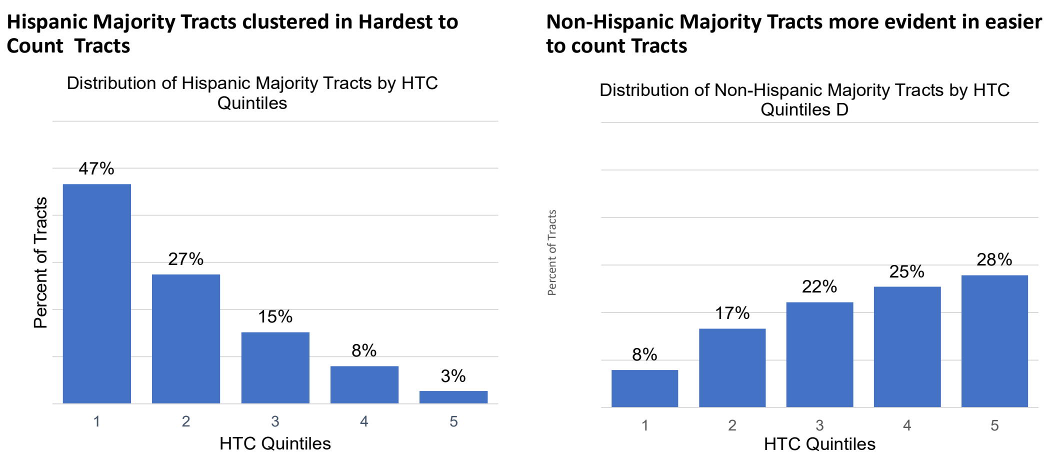
How Do Differences in Self-Response Among Hispanic Majority Tracts Affect Communities in California with Concentrations of Hispanics?

Our analysis is useful in highlighting the fact that census response levels vary significantly within groups of tracts, neighborhoods, and communities with similar race/ethnic profiles. It is then useful to consider, also, how this affects overall geographic patterns of response at the community, county, and state level.[22] **Figure 5** below shows that 47% of the Hispanic-majority tracts are in the highest HTC quintile 1 while only 3% live in the "easiest- to-count" quintile 5. For the Non-Hispanic population, the distribution is the opposite -- the distribution of tracts in the HTC quintile 1 is low (8%), while 20% reside in HTC Quintile 5 tracts.

Although Census 2020 self-response in California (69.6%) is higher than the national average (67.0%), the fact that three-quarters of Hispanic-majority tracts fall into the lower two quintiles of self-response while only 11% live in tracts in the highest two quintiles are cause for concern—highlighting the point that individual census response and socioeconomic opportunities stemming in part from census-data driven funding allocation formulae for key social, educational, and health investments are closely linked to and better understood as facets of community context than in terms of individual census respondent behavior.

Figure 5

Distribution of California Census Tracts by HTC Quintiles
and Hispanic/Non-Hispanic Majority Population



Substate Geographic Variations in Census Response

Our analyses have so far focused on the general dynamics of neighborhood and community context as factors that strongly determine self-response in Census 2020 at the state level.

However, it is important, also, to recognize that there are also differences in census response from one sub-state region to another—especially in large and diverse states such as California. To the extent that these differences appear, it must be recognized that they provide an indication of the geographical dimension of differential undercount and can result in inequities in eventual allocation of funding and political power.

It should be presumed, that, although a core set of community characteristics are consistently associated with response level, the precise dynamics of the interplay among specific factors associated with lower or higher census response varies substantially from community to community.

We focus here on two southern California counties: Los Angeles County and San Diego County.[23] These two counties, with a combined population of about 13 million, make up one-third of the total California population total.

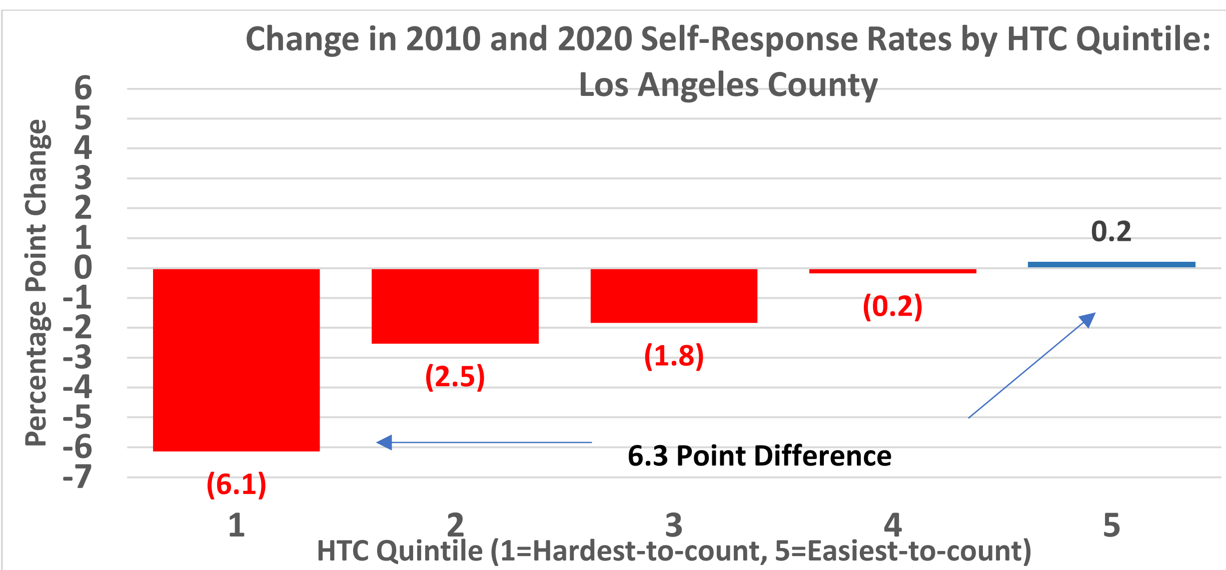
Because these counties are geographically large, highly-populated, and diverse socioeconomically and in racial/ethnic profile, fine-grained sub-county patterns of census response are, practically speaking, important to census stakeholders such as local municipalities. There are, for example, 88 municipalities in Los Angeles County and 18 incorporated cities in San Diego County.[24] School districts are also important stakeholders—since federal funding for compensatory education is distributed on the basis of census-derived data; there are more than 75 school districts in Los Angeles County and 42 in San Diego County.[25]

Trends and Patterns of Response in Los Angeles County: Census 2010 to Census 2020

We look, first, at Los Angeles County, the most populated county in California with a population of over 10 million living in more than 2,000 census tracts. Like the state as a whole, Los Angeles County is racially, socioeconomically, and geographically diverse (it includes a substantial rural area in addition to the very large urban area in the county).

What can be seen in this analysis is not good news—as can be clearly seen in **Figure 6** below. There were larger response rate declines from 2010 to 2020 than statewide, and in multiple HTC quintiles. The 2020 response rate fell 6.1 percentage points in HTC Quintile 1 tracts in Los Angeles County—more than double the state level change in Quintile 1 tracts. And the overall response rate fell by 3.9% for the entire county (69.0% to 65.1%).

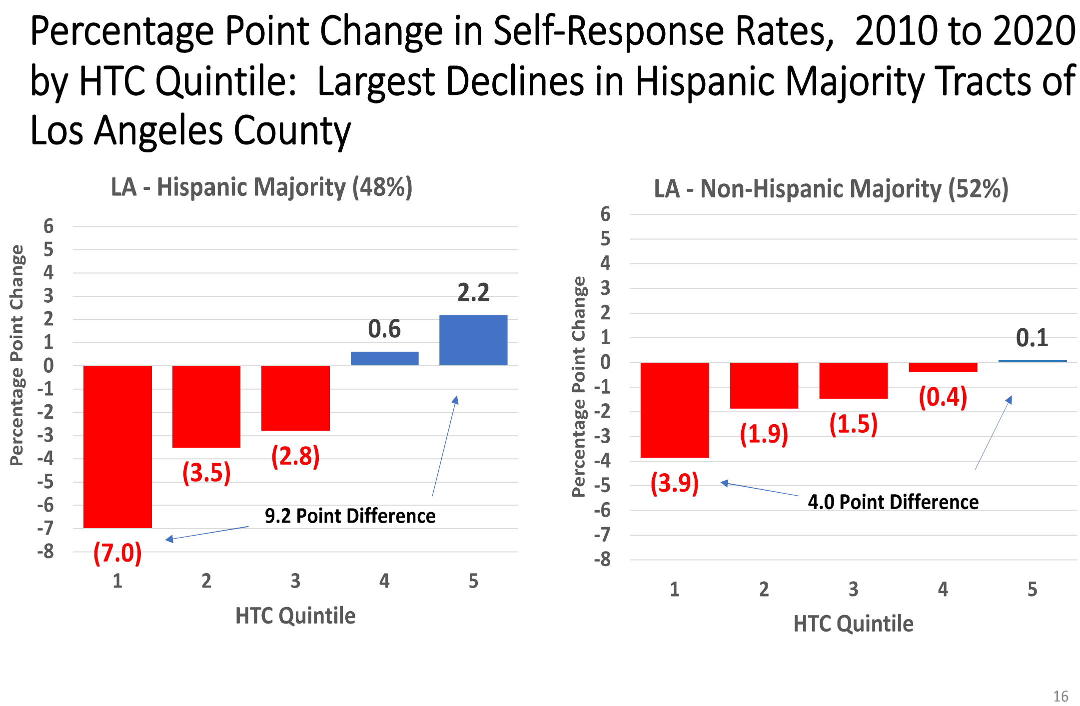
Figure 6



As might be expected, the contrast in Los Angeles County self-response in Hispanic-majority and Non-Hispanic-majority tracts echoes the pattern observed for the state as a whole.

Figure 7 below shows the striking difference in 2010-2020 response declines for these two types of tracts.

Figure 7



In addition to showing declines in self-response from 2010 to 2020 for most Los Angeles County tracts, especially the steep drop in the Hispanic-majority HTC Quintile 1 tracts—it shows a dramatic self-response gap between the easiest-to-count and hardest-to-count among the Hispanic-majority tracts.[26]

This is an important finding because it shows the need to understand census response level to be more closely linked to the "structural" factors captured in the HTC index than to race/ethnicity per se. As shown in **Figure 7**, the "self-response gap" between the easiest-to-count and hardest-to-count Hispanic-majority tracts is more than twice that found in the non-Hispanic-majority tracts (9.2% gap. Vs. 4.0% gap).

The full analysis of factors affecting level of census response within Hispanic neighborhoods in Los Angeles County is beyond the scope of this paper but the overall research on the county's Hispanic population strongly suggests that these differences in census response stem, in large measure, from the socioeconomic and sociopolitical status of successive waves of Mexican and Central American immigrants who have come to Los Angeles.

Trends and Patterns of Response in San Diego County: Census 2010 to Census 2020

San Diego County is a California county that, like Los Angeles County, has a large urbanized core population. While both counties include rural areas, San Diego differs from Los Angeles County in having rural areas with a mix of labor-intensive agricultural production and affluent suburban settlement. San Diego County is also smaller, having a population of about 3.3 million, making it about one-third the size of Los Angeles County. San Diego County, like Los Angeles County is racially diverse. However, it has a higher proportion of non-Hispanic Whites (44%), a lower proportion of Hispanics (34.8%) and slightly fewer Asians (12.9%). It also has a large and diverse foreign-born population (22.7% of the county total) but fewer than Los Angeles County (where about 34% of the population is foreign-born). Overall, San Diego County has a lower proportion of HTC tracts.[27]

Figure 8 below shows that the changes in San Diego County census response from 2010 to 2020 are very different than those observed in Los Angeles County. Whereas in Los Angeles County the 2020 census response rate declined from 2010 in multiple quintiles (Figures 6-7)—2010 to 2020 response rate increases in San Diego County occur across the HTC continuum, though the gains are much smaller in the hardest-to-count tracts.

Note the increase in response rates was more than 7 points in two quintiles, which boosted the overall increase for the county (a gain of 5.9 percentage points overall, versus the 3.9-point decline for Los Angeles County).

Figure 8

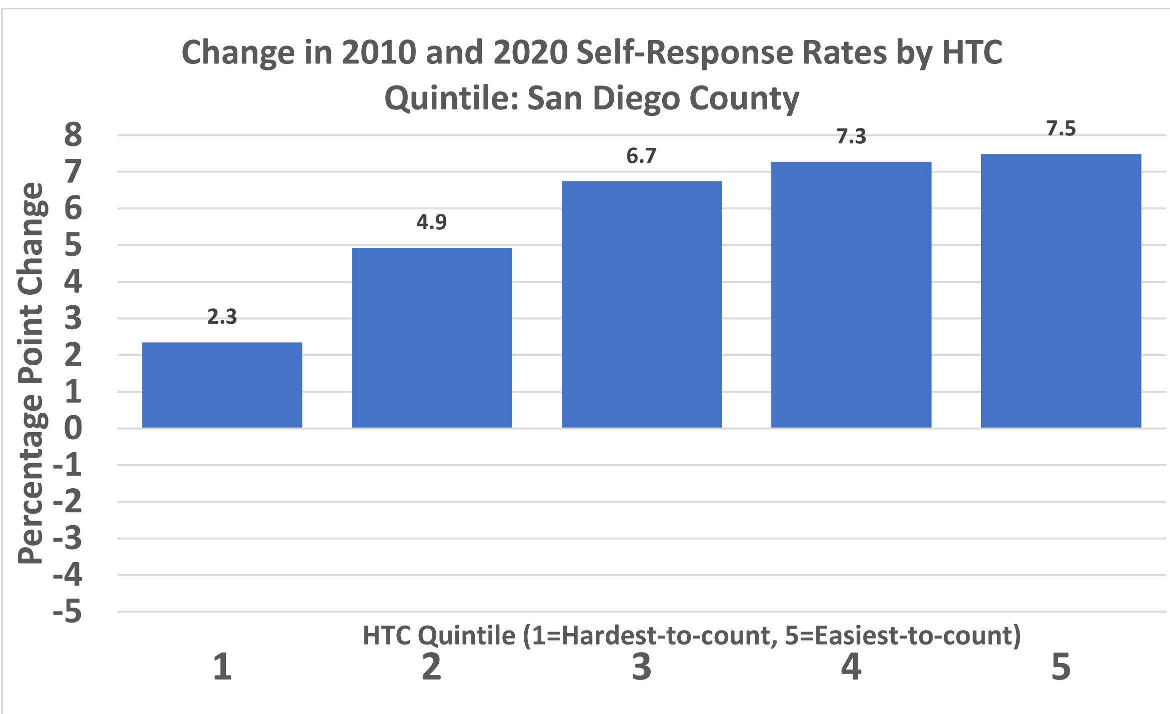
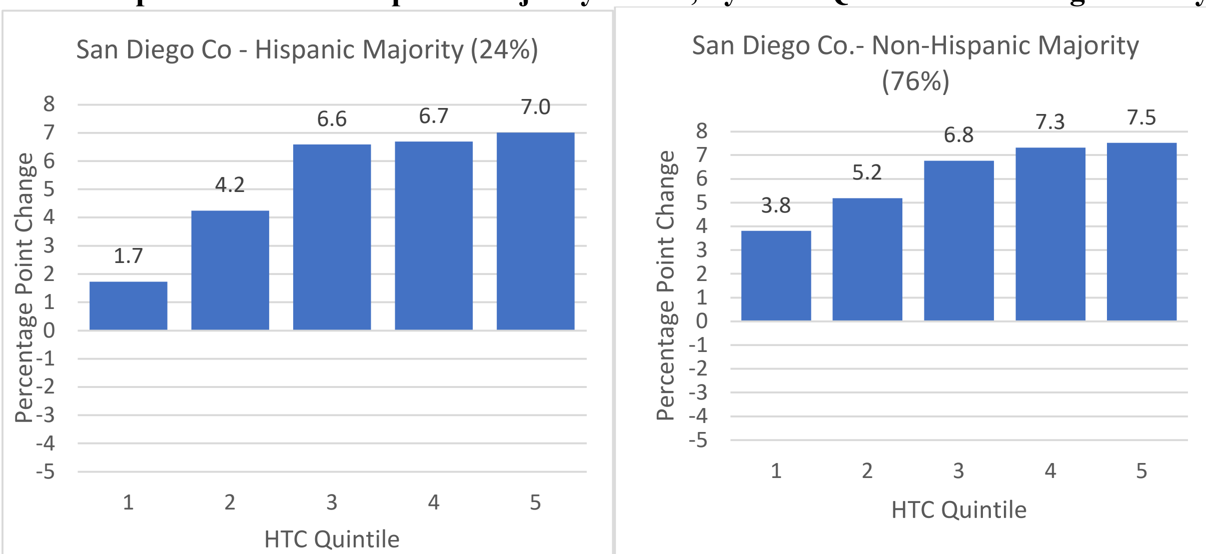


Figure 9 shows the percentage point change in self-response rates from 2010 to 2020 across the HTC continuum is similar for Hispanics and Non-Hispanic majority tracts in San Diego, those the HTC Quintile 1 gains are smaller for Hispanics.

Figure 9

Change in 2010 and 2020 Self -Response Rates
for Hispanic and Non-Hispanic Majority Tracts, by HTC Quintile: San Diego County



The contrasts between Los Angeles and San Diego counties with respect to the dynamics of Census 2020 self-response as compared to Census 2010 suggest that similar variations exist for many other sub-state areas. The current analysis is simply a beginning toward improved understanding of the complex interactions between different dimensions of community context and census data collection.[28]

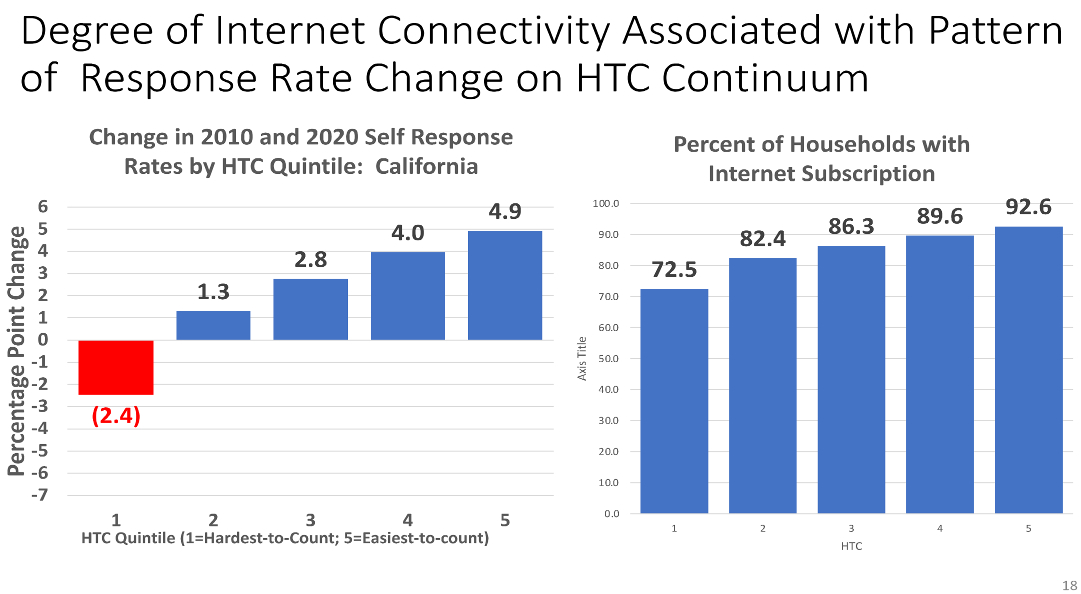
Patterns of Internet Connectivity and Census 2020 Response

Our analysis has now examined the association between community context as visualized based on ACS-derived indicators incorporated into the Hard-to-Count continuum and Census 2020 response. We now examine the impact of the major change in decennial census data collection—the shift toward online response as the preferred census response modality (although mail response and phone response continued to be available).

An important observation that comes out of this study of response rate patterns from a Hard-to-Count perspective is the strong association of Internet connectivity and the change in self-response rates from 2010 to 2020.

Response rate change by HTC quintile for California is displayed in the left-hand figure of **Figure 10** below; the higher levels of internet connectivity moving from harder-to-count to easier-to-count tracts is viewed on the right.

Figure 10



The decline in 2020 census response rates from 2010 seen for HTC Quintile 1 is associated with low levels of internet access; the relatively large response rate gains for "easy-to-enumerate" Quintile 5 is found in areas with very high connectivity. This relationship between census response and Internet access across the HTC scale is seen in all the areas evaluated here, especially for Hispanic majority tracts.

Although multiple facets of community context enter into tract-level self-response it is clear that patterns of Internet access make a difference. In Los Angeles County, for example, there is a strong positive correlation between Internet access and self-response level (Pearson correlation=.442, $p < .001$). In Fresno County, in the rural communities where Internet access is more uneven the positive correlation between tract-level broadband access and self-response is even more dramatic (Pearson correlation=.651, $p < .001$).

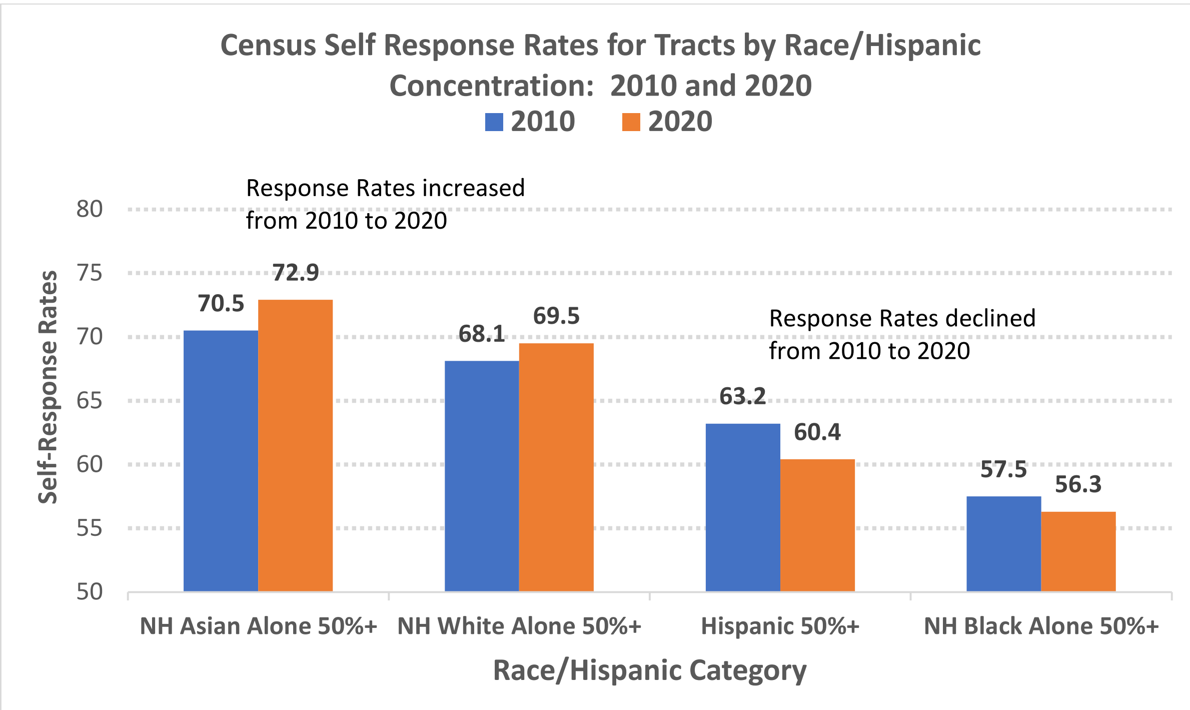
Association of Patterns of Change in Census Response Rates and Net Undercount Rates at the National Level, 2010 to 2020

Census Self-Response Rates

While the overall national response rate increased slightly from 66.5 percent in 2010 to 67.0 percent in 2020, this change was very different from that observed in Hispanic origin majority tracts in California. It needs to be remembered that our analysis here examines the relationship between community characteristics and census response, not the related but distinct patterns of individual household response.

The national response rates for tracts with higher concentrations of Non-Hispanic Asians and Non-Hispanic Whites increased from 2010 to 2020, while the rates for Non-Hispanic Blacks and Hispanics declined. Thus, the differentials in response rates in communities with different racial/ethnic profiles has widened See **Figure 11** below. This overall pattern of change is similar to the change in response rates in California tracts on the HTC continuum.

Figure 11
National-Level Changes in Census Response Rate from 2010 and 2020



Source: 2010 response rates from 2010 census operations data contained in the 2019 Planning Database at: https://www.census.gov/topics/research/guidance/planning-databases/2019.html; 2020 response rates from April 2021 Working Paper by William O'Hare and Jae June Lee (Figure 2, pg. 16) at: https://www.georgetownpoverty.org/issues/who-responded-in-the-2020-census/

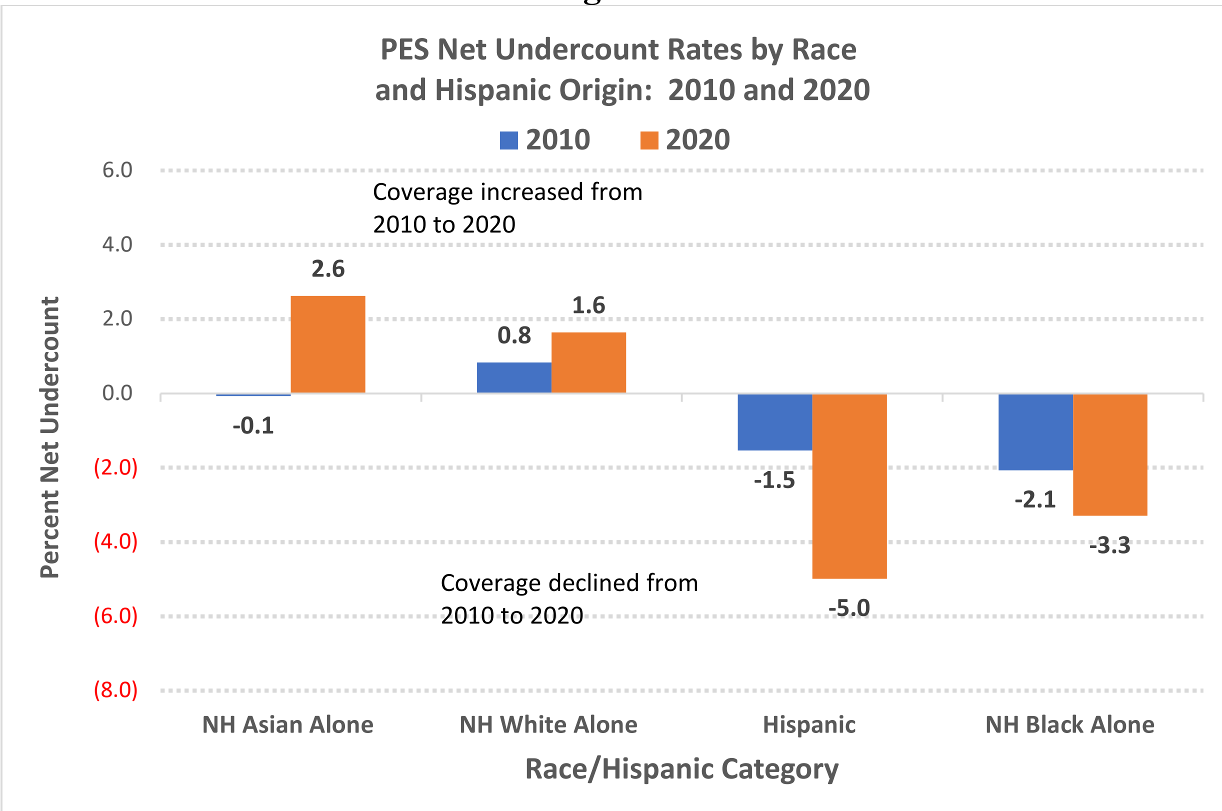
Net Undercount Rates

The widening of these gaps in level of census response at the national level are consistent with PES-based estimates of the eventual differential undercount for these racial/ethnic populations

The overall national net undercount as measured by the Post-Enumeration Survey (PES) of 0.24 percent for 2020 was not statistically different from the negligible rate of 0.01 percent in 2010. However, the national pattern hides the appreciable differences for demographic groups.

These 2010-2020 trends in coverage are shown in **Figure 12** on the next page..

Figure 12



Source: U.S. Bureau of the Census, 2022 at: https://www2.census.gov/programs-surveys/decennial/coverage-measurement/pes/national-census-coverage-estimates-by-demographic-characteristics.pdf

The PES-based net undercount rates for Non-Hispanic Asians and Non-Hispanic Whites declined from 2010 to 2020 (in fact, both these populations were "overcounted" in 2020), while the undercount rates for Non-Hispanic Blacks and Hispanic increased. Thus, the differentials in coverage (or net undercount) rates across the demographic groups widened. And the pattern of change for the coverage rates mirrors the pattern of change seen in our analysis of changes in tract-level response rates in California (**Figures 3-4, 6-8**), the Census Bureau's analysis of response omission rates in relation to decile of self-response (**Figure 1**), and the overall response rates for communities with concentrations of different racial/ethnic groups (**Figures 11-12**).

Analysis of the historical relationship between mail return rate and PES-estimated undercount in the 1990-2010 census shows that, although the relationship between lower response and ultimate differential undercount has varied somewhat from census to census, the PES-based estimate of lower response for disadvantaged racial/ethnic groups is always associated with differential undercount. For example, in 1990, the estimated national net undercount rate of Hispanics was 3.4 percentage points higher than the overall U.S. total and the national mail return rate of

Hispanics was 11.3 points lower than the overall mail return rate. The ratio of differential undercount to census response was lower in 2000 and 2010.

Determining the ratio of census response to undercount in 2020 has important practical implications because it provides a basis to then examine small-area differences in response rate and to infer the corresponding geographic disparities in enumeration since the response rate for each tract in the U.S has been made publicly available.

While the limitations of NRFU are most evident when examining the relationship between national-level response rate to PES-estimated differential undercount, it is likely that the adequacy of NRFU varies not only in relation to non-response workload[29] but, also, in relation to other community characteristics such as local social norms that discourage participation in proxy interviews[30] and proportion of undocumented immigrants in a tract (since a PIK—Protected Identification Key, typically, a Social Security Number—is required to link an administrative record to a non-responding household in order to protect privacy).

Similar patterns of ACS response are problematic in a different way, because they potentially result in sample bias due to some subpopulations being less likely to respond to the ACS than others and methodological limitations to assuring that ACS weighting overcomes that problem. We discuss the practical implications of this “double whammy” of correlated variations of response to the decennial census and to the ACS for equitable allocation of census-driven federal and state funding in a forthcoming paper.[31]

We have not focused here on the PES-based estimates of the relationship between tenure (homeowner or renter-headed household) and census coverage but, there too, there is an obvious and persistent negative correlation between percent of renter-headed households in a tract, tract-level census response and resulting coverage: a 1.91% gap between owner-headed and renter-headed households.[32]

Conclusions

Our analysis of Census 2020 tract-level self-response data shows that, despite California efforts to promote census response in harder-to-count tracts, local level of response was well-correlated with Hard-to-Count scores along the continuum. Looking only at national and state levels of self-response obscures the actual patterns of self-response in smaller area geographies, where there continue to be deep pockets of low self-response in HTC tracts and the likelihood of resulting inequity.

Our examination of changes of census response rate in 2020 in the historical context of trends from 1990 through 2010 shows that the steady pace of improvement in response rates in hard-to-count tracts that was achieved nationally from 1990 to 2010 reversed between 2010 and 2020—at least in our California case study.

It is surely possible to point to the negative impact the COVID-19 pandemic had on decennial census operations as an important factor in this reversal. For example, despite plans to provide census response assistance through mobile questionnaire assistance centers, this was not feasible. Shelter-in-place regulations also disrupted plans for Update-Leave operations in many areas. Nonetheless, our tract-level analysis of response patterns strongly suggests that adoption of the Internet-based online response option also broadened the gap between the hardest-to-count tracts, where there are, characteristically, lower levels of Internet connectivity. Higher national levels of census self-response were achieved with this modernization initiative but dramatic differences within geographic levels (states, counties) remained, giving cause for concern about equity.

Our analyses show that variations in level of average self-response between the highest and lowest quintiles of tracts within California (21.1%) and within Los Angeles County (19.4%) were as great as the state-to-state variations in self-response reported by the Census Bureau (20.4%).[33] The Internet response option worked well—but the resulting disparities in response level stemmed from it working best in the most affluent areas with high levels of broadband connectivity and not so well in the hardest-to-count tracts.

The national-level PES data show that level of Census 2020 self-response is strongly linked to eventual census coverage due to higher levels of household omission in the harder-to-count tracts. Further research (including analysis of vacancy rates along the HTC continuum) will be helpful in definitively quantifying the relationship between level of census response and coverage. Nonetheless, the historical data showing a consistent relationship between level of response and coverage from 1990-2010 for different racial/ethnic groups are consistent with the 2020 tabulations showing the monotonal relationship between self-response and PES-estimated coverage.

Our analyses, coupled with the national data, highlight the fact that it is not simply race/ethnicity and tenure that are related to coverage but that other structural factors of community context must be considered. In large counties such as Los Angeles where a historically-undercounted racial/ethnic group now makes up almost half of the total county population, there is, inevitably, socioeconomic diversity and neighborhood characteristics (many but not all captured in the HTC index) that play a big role in determining census response and, ultimately, coverage. Differential undercount is not simply a problem of racial equity; it is also a problem of socioeconomic and sociopolitical equity and how those patterns interact with census data collection procedures.

Our analysis of the patterns of change in census response from 2010 to 2020 in two large counties—Los Angeles and San Diego counties—shows that although census response continued to be lower in the hardest-to-count tracts of both counties, there were differences between in the dynamics of self-response that appear to be correlated with specific community characteristics, including racial/ethnic composition.

The observation of correlation between tract characteristics and response varying somewhat from county to county suggest the need for more localized planning to overcome barriers to response while, at the same time, raising concerns about geographical inequities in response since the national level analysis demonstrate that census coverage continues to be linked to local level of response, despite efforts to overcome this limitation via NRFU.

Implications for Census 2030

Marked disparities of self-response between hard-to-count and easy-to-count tracts in California need to be assessed in other states and local communities to extend these findings. We hypothesize that a core set of community characteristics, more or less those captured in the Census Bureau's Planning Database, will consistently be associated with level of response but that the weighting of different factors may vary from one local area to another. This suggests, then, the utility of developing a set of HTC index variants as a basis for tailoring census planning—both operations and campaigns to promote census participation—since the dynamics will vary from community to community.

Increased Census Bureau focus on HTC tracts presents opportunities for innovative new modes of collaboration with local stakeholders who best understand specific barriers to self-response. Efforts to more fully understand the "structural" and "system" factors that affect census response will also greatly improve the quality of American Community Survey data since the analysis of 2020 census response and ACS response shows that similar structural factors affect local level of response to both.

The quality of American Community Survey (ACS) data and decennial census data are both essential to fairness in allocation of federal, state, and philanthropic funding based on federal data. A prime example is ESEA Title IA compensatory education funding which is allocated based on numbers of children (based on population estimates derived from decennial data and proportion of children in poor households in a school district (derived from Small Area Income and Poverty Estimates that are generated by modeling that draws on ACS data).[34]

The accuracy of population data on numbers of children and ACS data on household characteristics in each school district in the U.S. are both crucial to equitable and cost-effective investment in education. Consequently, the best possible understanding of structural causes of undercount holds promise for adjusting allocation of funding to optimally meet local school district needs.

Finally, Census Bureau collaboration in nation-wide efforts to improve Internet access and overcome barriers to online response may mitigate the widening "digital gap" in decennial and ACS response. Current federal efforts to invest in broadband infrastructure are a welcome development but they will not suffice without matching and coordinated decade-long investments in digital literacy and multi-stranded cross-agency collaboration to build household use of online resources for services such as telehealth, self-directed learning, and a multitude of other uses that are now commonplace among the affluent, Internet connected segments of the population.

ENDNOTES

[1] O'Hare, William, Jae June J. Lee, "Who Responded in the 2020 Census: Variation in Tract-Level Self-Response Rates in the 2020 U.S. Census" https://www.georgetownpoverty.org/issues/who-responded-in-the-2020-census/

[2] The Hard-to-Count (HTC) Score is based on a collection of specific variables that are associated with response rates and enumeration difficulty (such as high poverty, low education, high rentership, English language difficulty, crowded housing). See the discussion in the next section.

[3] Bruce, Antonio and J. Gregory Robinson. "Tract Level Planning Database with Census 2000 Data." Population Division, U.S. Bureau of the Census, 2012.
https://www2.census.gov/programs-surveys/research/guidance/planning-databases/2000/pdb-tract-2000.pdf
The original HTC score was subsequently replaced in the PDB with the low response score (LRS).

[4] Bruce, Antonio, J. Gregory Robinson, Jason E. Devine. "A Planning Database to Identify Areas That Are Hard-to-Enumerate and Hard-To-Survey in the United States." Population Division, U.S. Bureau of the Census, 2012. Available at http://www.asasrms.org/Proceedings/H2R2012
See, also, Bates, Nancy and Mary H. Mulry. 2011. "Using a Geographic Segmentation to Understand, Predict, and Plan for Census and Survey Mail Nonresponse," Journal of Official Statistics, Vol. 27, No. 4, 2011, pp. 601–618.

[5] Chandra and Nancy Bates (2017), "The Low Response Score (LRS): A Metric to Locate, Predict and Manage Hard-to-Survey Populations", Public Opinion Quarterly, Vol. 81, No. 1, Spring 2017, pp.144–156
https://www.census.gov/content/dam/Census/topics/research/erdman_bates_2017.pdf
The original HTC score was replaced in the Planning Database with the low response score (LRS).

[6] De La Puente M. (DATE??) Using Ethnography to Explain Why People are Missed or Erroneously Included by the Census: The Evidence from Small Area Ethnographic Studies, Center for Survey Methods Research, U.S. Census Bureau https://www.census.gov/srd/papers/pdf/mdp9501.pdf.

[7] Fein DJ, West K. The Sources of Census Undercount: Findings from the 1986 Los Angeles Test Census, Survey Methodology, December, 1988. https://www150.statcan.gc.ca/n1/pub/12-001-x/1988002/article/14590-eng.pdf

See also West K, Fein DJ. Census Undercount: An Historical and Contemporary Sociological Issue, Sociological Inquiry, Vol. 60 #2, 1990.
https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-682X.1990.tb00134.x

[8] California Census 2020, “Identifying California’s Hard To Count In Census 2020”, https://census.ca.gov/california-htc/

[9]Studies have indicated that racial groups are only proxies of more specific causal factors that contribute to enumeration difficulty (J. Gregory Robinson, Antonio Bruce, and Carrie Johanson, “The Planning Database: Decennial Census Data for Historical, Real-time, and Prospective Analysis,” 2007 Joint Statistical Meetings, Salt Lake City ; Word, 1997) www.census.gov/population/www/documentation/twps0019.html. The variables shown in Appendix 1 are factors more directly associated with the barriers to enumeration and better suited for trying to capture the diversity that can occur at the tract level. Census data show that at the local level there can be sharp differences across tracts within race/Hispanic groups on the characteristics that influence enumeration propensities (Bates and Mulry, 2011). https://www.scb.se/contentassets/f6bcee6f397c4fd68db6452fc9643e68/using-a-geographic-segmentation-to-understand-predict-and-plan-for-census-and-survey-mail-nonresponse.pdf

[10] See the Census Bureau’s TIGERweb Decennial website for an excellent interactive tool that can be used with 2000, 2010, and 2020 census tract geography/designation. https://tigerweb.geo.census.gov/tigerweb2020/

[11] To form the HTC index for California, each of the 14 variables was sorted from high to low across all tracts (e.g. sort tracts from the highest percent in poverty to the lowest.). Then each variable was scaled from 1 to 11. The highest values were recoded as 11, down to the lowest values, which were recoded as 0 (e.g. values of 11 are giving to tracts with the highest poverty rates and values of 0 are given to tracts below the California poverty rate tract median.) The sum of the 14 values represents a tract’s CA-HTC Index.

[12] https://www.census.gov/library/stories/state-by-state/california-population-change-between-census-decade.html

[13]Cesar Alesi Perez, Marisol Cuellar, and Hans Johnson, “Immigrants in California”, Public Policy Institute of California, January, 2023 https://www.ppic.org/publication/immigrants-in-california/

[14] The self-response rate represents the percent of household returning the census questionnaire by internet response, a mailed paper questionnaire, or by telephone. See more discussion in the next section.

[15] The 2020 census employed three primary contact strategies. The first and most widely used was “Internet First”, which was used in areas with relatively high internet connectivity, where households were invited to complete the questionnaire online. The second contact strategy was “Internet Choice”, which was used in areas with low internet connectivity, where households were provided with a paper questionnaire to return and information about how to respond online. An English-Spanish invitation to respond and questionnaire was sent to areas with 20% or more

households needing Spanish language assistance. https://scag.ca.gov/sites/main/files/file-attachments/demo30_keynotechristy.pdf?1604697375

[16] https://www.pewresearch.org/fact-tank/2018/09/14/education-levels-of-u-s-immigrants-are-on-the-rise/

[17] Joanna Lee and Jennifer Ito, "Census 2020 in the San Joaquin Valley: An Empirical Assessment of Strategies to Activate Hard-To-Count Populations" University of Southern California Dornsife Equity Research Institute, March, 2023.

[18] https://www.census.gov/library/visualizations/interactive/2020-census-total-response-rates-by-state.html

[19] The census omission rates are based on the results of the 2020 Post Enumeration Survey. Omissions represent people who should have been counted in the census but were not. For more information about the definition of omissions, see pages 4-6 at https://www2.census.gov/programs-surveys/decennial/coverage-measurement/pes/national-census-coverage-estimates-by-demographic-characteristics.pdf

[20] Bruce, Antonio, J. Gregory Robinson, Jason E. Devine. "A Planning Database to Identify Areas That Are Hard-to-Enumerate and Hard-To-Survey in the United States." Population Division, U.S. Bureau of the Census, 2012, Table 2. Available at http://www.asasrms.org/Proceedings/H2R2012

[21] Joanna Lee, Jennifer Ito, and Manuel Pastor, "Census 2020 in the San Joaquin Valley: An Empirical Assessment of Strategies to Activate Hard-To-Count Populations", Equity Research Institute, University of Southern California, April, 2023.

[22] This study focuses on the response rate patterns of the Hispanic population compared to the total population of California. Later analysis could include other Non-Hispanic groups, such as Non-Hispanic Blacks and Asians. However, the Non-Hispanic Blacks form a majority in only 64 tracts in California, which leads to small "sample sizes" of tracts when disaggregated into quintiles and the results for the quintiles with very few tracts can be misleading

[23] Kissam and Robinson are also analyzing response rate patterns in Fresno County, a large agricultural county in Central California but will soon report on those patterns in a subsequent paper. Similar county-level case studies are feasible in other counties and have the promise of showing the extent and type of how small-area variations in response are affected by local community context.

[24] Municipalities in Los Angeles County can be found at http://www.laalmanac.com/government/gl10.php . The number of incorporated cities in San Diego County is from https://www.calcities.org/get-involved/regional-divisions/san-diego-county-division

[25] Number of K-12 school districts in Los Angeles County is from https://www.asumag.com/research/top-10s/article/21152096/largest-school-districts-in-los-

angeles-county-201920 Number of K-12 school districts in San Diego county is from https://sdtoday.6amcity.com/20-school-districts-in-san-diego-california

[26] The number of tracts for Hispanics in Quintile 4 and 5 are relatively small (with only 7% and 3% of the Hispanic majority tracts). The results for these quintiles—such as the higher 2020 response rates increase in Hispanic majority tracts compared to Non-Hispanic majority tracts) should be interpreted with caution.

[27] https://www.census.gov/quickfacts/fact/table/sandiegocountycalifornia/PST045221

[28] Kissam and Robinson's forthcoming analysis of patterns of 2020 census response and 2015-2019 ACS response in Fresno County already shows, for example, that in this distinct sub-state region, a county with a lower level of urbanization than either Los Angeles or San Diego County, that there are significant differences in level of response between the highly-urbanized tracts, the slightly urbanized rural tracts (with significant proportion of households in a fairly rural area designated as an "urban cluster" with a population of 2,500 to 50,000), and those in extremely rural portions of the county.

[29] Salvo J, Arun PL, Misclassifying New York's Hidden Units as Vacant in 2010: Lessons Gleaned for the 2020 Census, Population Research Policy Review, 2013.

[30] Kissam, E., Wadsworth, G, and Intili, J. "Troubled Reflections: Summary of Themes and Implications for Census 2020", San Joaquin Valley Census Research Project, San Joaquin Valley Health Fund, 2019.

[31] Ed Kissam and J. Gregory Robinson, "Dynamics of Census and ACS Response in Fresno County: Implications for Equitable Allocation of ESEA Title IA Funding for Low-Income Hispanic Communities", (forthcoming, May, 2023).

[32] See **Table 6** showing a 1.48% undercount of renter-occupied households and 0.43% overcount of owner-occupied households https://www2.census.gov/programs-surveys/decennial/coverage-measurement/pes/national-census-coverage-estimates-by-demographic-characteristics.pdf

[33] https://www.census.gov/library/visualizations/interactive/2020-census-self-response-rates-map.html

[34] https://www.census.gov/programs-surveys/saipe/technical-documentation/methodology/school-districts/overview-school-district.html