THE CENSUS PROJECT

Supporting quality measures of America's economy and places

STANDARD DEVIATIONS: How Widening Gaps in Decennial Census and American Community Survey Response in Easy and Hard to Count Communities Lead to Inequitable Allocation of Social Program Funding

By Edward Kissam

Standard Deviations blog posts represent the views of the author/organization, but not necessarily those of the Census Project.

Here, I briefly review the evidence suggesting that a combination of long-term "structural" limitations in Census Bureau methodology and the distinctive pandemic-related circumstances that likely led to systematic differential undercount in hard-to-count neighborhoods and communities in 2020 may have seriously impaired data quality for the coming decade.

My observations stem from analysis of ACS response and Census 2020 self-response in a virtual "case study' of a California area (Fresno County) with high concentrations of agricultural workers and other low-income households of Mexican immigrants.[1] In light of this evidence, it is almost certainly the case that currently available decennial census and ACS data will not be "fit for use" for equitably allocating public sector funding in this area and others like it.

I argue, however, that if more detailed operational metrics were made available, it would be possible to assess the extent to which the characteristics associated with low self-response in hard-to-count Hispanic census tracts with concentrations of non-citizens also undermined effectiveness of nonresponse follow-up (NRFU) and led to differential undercount in this stratum of census tracts.

[1] Fresno County is large—with a 2020 population of about 1 million, with a majority Hispanic population (53.8%) and more than 100,000 agricultural workers, almost all of whom are Mexican-origin immigrants. J. Gregory Robinson has been a key collaborator in this effort via meticulous work analyzing published data and expert observations on patterns observed in the merged dataset built from Census Bureau Planning Database (PDB) data, the database developed and distributed by CUNY, and Census Bureau published data on ACS response and 2020 decennial self-response.

Valiant Efforts To Assess and Respond to Data Quality Problems in Census 2020 and the ACS, But Serious Issues Remain Unresolved

It's a pleasure to see the new insights from DA (demographic analysis) and the PES (post-enumeration survey). As Census Bureau Director, Rob Santos, aptly stated at the Bureau's March 10, 2022 press conference, these findings, particularly those related to differential undercount of American Indian/Alaska Native, Hispanic, and Black populations, could be "kept in mind" in assessing whether the officially-tabulated Census 2020 data would be adequate for specific uses.

It has also been heartening to see the valiant and meticulous efforts by Census Bureau analysts to identify biases in the 2020 ACS data by examining year-to-year changes in tabulations of variables where rapid change is extremely unlikely, analyzing how operational challenges gave rise to the data quality problems, and testing approaches to enhancing ACS data quality, by generating the experimental estimates.[2]

Unfortunately, despite these diligent efforts and the detailed delineation of Census 2020 data quality issues at the national level, using DA, PES, and longitudinal analysis of ACS tabulations as a basis for generating experimental re-weighted estimate give rise to inequities in funding allocation. When one looks beneath the surface, beyond the current national and forthcoming state-level analyses, to consider potential "use cases" of decennial census tabulations in conjunction with ACS tabulations, data quality will fall short as a basis for allocating federal, state, and philanthropic funding for social programs—one of several critical "use cases" for census-derived data.

Current efforts to mitigate data quality problems in the 2020-2030 decade by use of a "blended base" in the Population Estimates Program are promising in presenting a practical solution but inequities at the local level will still be more problematic than national and state level analyses indicate. Assessing "fitness of use" of ACS and Census 2020 data only at the national level and only in relation to broad race/ethnic variables obscures much more complex and serious unevenness of differential undercount of vulnerable sub-populations in a wide range of local community contexts. Synthetic adjustment, while imperfect, is viable and will mitigate the serious inequities that would result from use of unadjusted data.

The Indirect Consequence of Differential Response to the ACS and Decennial Census

The accuracy of census population count is a headline-generating concern but detailed demographic and socioeconomic data from the ACS is, arguably, even more important in

[2] Rothbaum, J., et al., "Addressing Nonresponse Bias in the American Community Survey During the Pandemic Using Administrative Data ", 2021 AMERICAN COMMUNITY SURVEY RESEARCH AND EVALUATION REPORT MEMORANDUM SERIES # ACS21-RER-05 and SEHSD Working Paper #2021-24

equitably allocating federal, state, and philanthropic funding for social programs[3] in part because differential non-response and resulting differential non-response is more extreme for the ACS than in the decennial census.[4] As policy analysts and experts consistently observe, equitable funding allocation is an exercise in "cutting up the pie", impacted more by variations in data accuracy than by the size of the pie.

The widening gap in self-response rate between easy-to-count and hard-to-count communities and neighborhoods in 2020, coupled with longstanding limitations inherent in standard NRFU make it clear that socioeconomically and socio-politically disadvantaged local areas in the United States will more often find themselves unfairly deprived of social program resources than at any point in the past four decades. They have no choice but to confront the practical consequences of differential undercount.[5]

Many, but not all, of the quality shortcomings in 2020 census and ACS data certainly did arise from major operational challenges specific to the COVID-19 pandemic. But others stemmed from chronic limitations of Census Bureau methodology such as inadequate Master Address File development[6], unsuitability of standard data collection procedures for "complex households", limitations of availability and accuracy of proxy interviews and administrative records.[7]

---

[3] Social program funding allocation formulae understandably give weight both to jurisdictions' socioeconomic characteristics (only available in the ACS) and population count while a number of other federal funding programs rely more heavily on population count or in combination with other sources of economic data.

[4] Analysis of funding allocation in more than 300 federal programs is clear in showing that census-derived data is a key component in funding allocation for social programs but that ACS data has additional important impacts. See Andrew Reamer, "Counting for Dollars 2020: The Role of the Decennial Census in the Geographic Distribution of Federal Funds", George Washington Institute of Public Policy, April 29, 2020. Examples of programs where municipalities and other local jurisdictions (e.g. school districts and workforce development boards) have a direct state include: CDBG, Title 1 education funding, and WIOA funding.

[5] Practically speaking, it was not viable to examine the uneven quality of Census 2020 data at the tract, zipcode, or county level in connection with the apportionment file. Similarly, the pressing timetable for reapportionment made it infeasible to fully examine the quality of the PL 94-171 file. However, there are now opportunities to give careful consideration of these issues in connection with use of the population/housing tabulations for allocating federal, state, and local program funding.

[6] As shown in the Census Bureau's own evaluation of LUCA in 2010 (Swarz, R, Virgile, P, and B Timko, 2010 Census Local Update of Census Addresses Assessment: Final Report, Geography Division, U.S. Census Bureau. 2012) , the program was implemented very unevenly. Financial constraints, together with unwarranted optimism about high-tech solutions for address listing and reliance on local administrative records, led to the decision to save money by only conducting address canvassing in about 30% of the U.S.

[7] Colleagues and I have documented problems with MAF-building in hard-to-count neighborhoods for several decades by using local community address canvassers trained to identify "low visibility" housing units. See GAO, "Lessons Learned for Locating and Counting Migrant and Seasonal Farmworkers", GAO 03-605, 2003 and Ed Kissam, Cindy Quezada, and Jo Ann Intili, "Community-Based Canvassing to Improve the Census Bureau's Master Address File: California's Experience in LUCA 2018", Statistical Journal of the International Association of Official Statistics, Vol. 35, December, 2018. Our initial pilot of community-based address canvassing in Fresno County indicated that about 8% of the housing units in our sample of HTC tracts were "low visibility" hidden housing units.

Fresno County Farmworker Communities: A Case Study for Assessing he Consequences of Differential Response and Methodological Limitations of NRFU

I focus on the evidence of differential undercount of farmworker communities in Fresno County, California, one of the top three labor-intensive agricultural counties in California—in part because I've done a good deal of research there. But, more importantly, census undercount in the farmworker population—about 90% Hispanic immigrants, more than half of whom lack legal status—is a paradigm case showing how interaction of many different factors can give rise to serious and systematic undercount. This hard-to-enumerate farmworker population is the proverbial "canary in the mine" that provides a basis for assessing the broader societal risks and operational etiology of differential undercount in marginalized communities.[8]

Even more problematic than systematic undercount in the hard-to-enumerate communities is that the result is a distorted demographic and socioeconomic profile that consistently results in loss of funding for the social programs most critical for improving community life in local municipalities with concentrations of low-income immigrants and farmworkers similar to those highlighted here across the U.S.[9] In the overarching policy context, the Catch-22 is that the very conditions associated with differential undercount are those that these programs where funding is allocated using census-derived data seek to ameliorate.

The 2020 Decennial Census and ACS in Fresno County: Patterns of Differential Response

Data analyzed by J. Gregory Robinson, drawn from the ACS-based Census 2020 Planning Database, published reports of Census 2010 mail return and mail response, and Census 2020 tract-level reports of self-response, show that introduction of online response exacerbated differential self-response between the hardest-to-count and easiest-to-count census tracts in Fresno County. [10] Robinson has also pointed to similar disparities in urban counties as well—most notably Los Angeles County.

---

[8] The most recent data on farmworker legal status and race/ethnicity from the National Agricultural Worker Survey tabulation of California farmworkers is for FY2015-2016. It is, therefore, comparable to the 2013-2017 ACS-based 5-year estimate of persons employed in agriculture. The total number of persons employed in agriculture in each ZCTA (the population tabulated in ACS) includes supervisors, managers, and clerical support and is computed in part using data from the Population Estimates program.

[9] Martin, P., Kissam, E., and Gabbard, S. ("The impact of migrant travel patterns on the undercount of Hispanic farm workers". In *Proceedings of the Bureau of the Census Research Conference on Undercounted Ethnic Populations.* Richmond, VA: Bureau of the Census, 1993. We assessed undercount via analysis of skewed demographic and socioeconomic profile of the census-identified farmworker population (using "long form" Census 1990 and National Agricultural Worker Survey data).

[10] Unpublished analysis, J. Gregory Robinson, February, 2022 (data/analysis available on request). Robinson has analyzed at the tract level changes in response between the 2010 and 2020 decennial censuses, the relation of these changes to ACS response rates, relationship between HTC score and self-response in relation to selected

Robinson's analysis shows that in the hardest-to-count quintile of 196 census tracts in Fresno County (with an average California HTC score of 96), achieved only a 59.7% self-response rate despite major state and philanthropic investments in outreach/assistance, while the easiest-to-count quintile of tracts achieved a 79.3% self-response rate in Census 2020.

There's much more work to be done and Census Bureau release of zipcode-level or operational metrics for small clusters of tracts would be helpful in assessing how non-response contributed to final undercount. Nonetheless, what is clear is that, given observed levels of non-response to the ACS in the 2013-2017 dataset and the published tract-level data on self-response to the 2020 decennial census, there are likely to be very serious data quality problems in the hard-to-count farmworker communities in Fresno County.

An example of a Fresno County sub-region of particular concern is a sub-county area known as "the Westside". One community in this sub-county area is Mendota, sometimes known as "Cantaloupe Capital of the World". ACS-based estimates show that 58% of the labor force in Census Tract 83.01 that includes much of Mendota are agricultural workers. With a 57.4% 2020 Census response rate and an ACS 2013-2017 response rate of 13.8%, there is very likely to have been a serious community undercount in the decennial.

A few miles to the south of Mendota, there is census tract 82.00 with several smaller farmworker communities—San Joaquin, Tranquility, and Cantua Creek—where almost half of the labor force (an estimated 45%) are agricultural workers. The 2020 decennial self-response rate in the tract was 45.2% while the 2013-2017 ACS response rate was 20.2%. Internet connectivity in both of these Westside communities is also low—51.5% in Mendota, 58.4% in the San Joaquin, Tranquility, Cantua Creek area.

As might be expected, other known correlates of low census response are evident throughout this sub-county region—with about two-thirds of adults lacking a high school education and about half speaking English "not well". It is not a surprise, given the hard-to-count scores in tracts like these (98 for Mendota, 87 for the San Joaquin area) that ACS response is extraordinarily low and that decennial response is low. But the magnitude of the differentials is a matter of real concern.

Another concern in different sub-regions of Fresno County and other similar areas is that, despite the very welcome utilization of bilingual census forms in areas with concentrations of limited-English households, census tracts with less than 20% Spanish-speaking linguistically isolated households did not receive bilingual census response forms and the Census Bureau's planned use of mobile questionnaire assistance centers (MQACs) became infeasible due to the pandemic.

---

tract characteristics (e.g. >50% Hispanic, Internet First vs. Internet Choice response mode. Robinson has also examined patterns of differential self-response in Los Angeles County and in the other counties of the southern San Joaquin Valley (Kern, Tulare, Kings, Madera, Merced, and Stanislaus), and San Diego County.

External Benchmarks for the Reliability of ACS Data on Agricultural Workers

The efforts by Rothbaum and his Census Bureau colleagues to assess ACS sample bias by comparing the ACS tabulations with high-quality data from external sources is an attractive and promising approach to understanding the ways in which the COVID-19 pandemic and/or other factors may have distorted ACS tabulations of key population characteristics.

However, their analyses are not well-suited for examining ACS bias in the hard-to-count tracts with concentrations of undocumented immigrants—because two important data sources they rely on for weighting responses (IRS tax returns and SSA program benefit data) are not uniformly available or reliable in the low-income immigrant-dense neighborhoods and communities.
Researchers studying migrant and seasonal farmworkers, have for, at least the past half century, explored alternatives to census data as a source of information on U.S. farmworkers. It is now possible, especially in California, to examine high-quality alternative data sources to the ACS. The 2013-2017 ACS-derived zipcode-level data on occupation identifies 415,000 agricultural employees in California but an estimate based on state unemployment insurance records suggests there were actually almost 990,000 California farmworkers in 2016.[11]

My own estimate of the size of the California farmworker population using a different analytic methodology drawing on Census of Agriculture and National Agricultural Worker Survey data, suggest there were about 650,000 California farmworkers in 2017.[12] There are complexities, and uncertainties related to either of these alternative estimates—but there is, at least, very good reason to believe that the tally of ACS-identified farmworkers indicate an ACS underestimate of about 30% in the 15 communities of Fresno County with high concentrations of farmworkers.[13]

---

[11] The estimate of ACS-identified farmworkers in California is based on ACS zipcode-level data downloaded by Richard Mines, for his collaboration with Ed Kissam and colleagues in identifying the levels of COVID-19 vaccination in California communities with concentrations of farmworkers (unpublished white paper prepared for California Department of Public Health vaccination campaign efforts, Heather Riden, Ed Kissam, and Rick Mines, "The Need to Micro-Target Vaccination Outreach in California Farmworker Areas" Western Center for Agricultural Health, University of California, Davis, August 8, 2021). The estimate of California agricultural workers based on UI data is from Phillip Martin, Brandon Hooker, and Marc Stockton, "Ratio of farmworkers to farm jobs in California increased to 2.3 in 2016", California Agriculture, Volume 73, #2, 2019.

[12] The analytic approach I used for estimating the number of California farmworkers based on contracted and hired farm labor expenses and National Agricultural Worker Survey data on farmworker employment patterns is described in depth in an earlier national estimate developed for the National Legal Assistance and Defenders Association to support Legal Services Corporation allocation of scarce funding (Kissam and Williams 2013). My estimate here is based on updating our Census of Agriculture/NAWS base with 2017 farm labor expense data and FY 2015-2016 NAWS data on farmworker wages.

[13] We defined farmworker-communities based on ACS estimates of number of persons employed in agriculture as a proportion of the total employed labor force. In the 15 Fresno County ZCTA's we identify as farmworker

Was Differential Self-Response in Fresno County "Cured" by NRFU in 2020?

As Nancy Mathiowetz and other experts (William O'Hare, Matthew Barreto) testified in litigation regarding the potential addition of the citizenship question to Census 2020, levels of non-response have always been correlated with estimates of differential undercount for racial/ethnic populations and population sub-groups (e.g. children)[14] observed at the national level (although correlation coefficients have varied from census to census).[15] Analyzing and quantifying the relationship between ACS and decennial census response, and differential undercount at the local level is, indeed, challenging but, at the same time, crucial; although coefficients may vary, the patterns can be expected to be similar. Although the Census Bureau has been reluctant to release sub-state operational metrics due to privacy concerns, these concerns should not be relevant for release of operational metrics for different types of tracts (e.g. for aggregates of census tracts based on HTC scores, for low-income tracts with concentrations of non-citizens, for tracts categorized by census data collection modality).

An observed pattern of <u>differential self-response</u> within states or within counties is a red flag indicating the possibility of <u>differential undercount</u> and biased demographic and socioeconomic tabulations if NRFU procedures or ACS weighting fail to work effectively in a local area.[16] In principle, NRFU efforts might have compensated for non-response in 2020 in Fresno County—but it is uncertain whether NRFU compensated reliably in the hard-to-count tracts where proxy interviews were difficult to secure and often unreliable, where administrative records were not available or unreliable for a substantial proportion of the immigrant population, and where hot-deck imputation was likely to be biased because the

communities, the % of total employment in agriculture (as identified in ACS) varies from 16% to 76%. There are small numbers of agricultural employees in these communities who are managers, etc., but not many.

[14] O'Hare, William P.; Mayol-Garcia, Haris; Wildsmith, Elizabeth; and Alicia Torres, "The Invisible Ones: How Latino Children Are Left Out of Our Nation's Census Count", Report from the Child Trends Hispanic Institute and NALEO Education Fund, April, 2016.

[15] "Plaintiffs' Corrected Proposed Findings of Fact", Kravitz, et al. v. United States Department of Commerce, et al. (Case 8:18-cv-01041-GJH Document 151-1 Filed 02/18/19), pp. 125-147. J.G. Robinson reviewed Census Bureau reports on mail response in 1990, 2000, and 2010 in relation to post-enumeration survey estimates of undercount and confirmed Mathiowetz's general observation (J.G.Robinson, unpublished spreadsheet, March 1, 2019, available on request). Based on review of Mathiowetz testimony and Robinson's analysis of the historical relationship between self-response level and undercount, and our SJVCRP survey findings on Latino immigrants' willingness to respond I projected a 35% response differential between White Non-Hispanic and Hispanic immigrant households and a 9%-13% undercount of Latino immigrants in areas such as Fresno County.

[16] There has been extensive analysis of the relationship between demographic and socioeconomic characteristics and self-response (e.g. O'Hare, Robinson, and many others) but much less on the robustness of ACS methodology for dealing with non-response or the reliability of the decennial census NRFU procedures that had not been fully tested (in part due to budget constraints for planned dress rehearsal studies).

Latino immigrant households who were most willing to respond to the census were older and smaller than those who were less willing to respond.[17]

Exactly what the observed 19.6% average differential self-response gap between the hardest-to-count and easiest-to-count census tracts in Fresno County in 2020 (compared to an 11.4% gap in 2010) indicates with respect to eventual differential undercount is not easy to quantify because to do so it is necessary to consider the operational factors that affected quality of NRFU in different types of census tracts.

The Census Bureau should provide operational metrics on these strata (e.g. for each quintile of tracts ranked on the HTC index) to assess how NRFU operations might have impacted reliability of enumeration. It would be advisable to consider whether specific factors, as distinct from the overall HTC score of a tract, had problematic and differing operational consequences among the hard-to-count stratum of tracts—e.g. in tracts with relatively low or high proportions of non-citizens? In tracts receiving Internet First or Internet Choice or Update-Leave treatment?

However, the Census Bureau has not to date agreed to release operational metrics regarding the level of proxy interviews, use of administrative records, and whole person imputations for smaller geographies than the state level.[18] Until such metrics are available, data analysts will be unable to even determine whether Census 2020 and ACS data are fit for use at the tract, neighborhood, or community level. To reliably assess the implications of variations in self-response rate it would be necessary to provide information at the sub-county level—because there is substantial variation in characteristics of tracts that affect both self-response and NRFU, especially in large counties with concentrations of Hispanics such as Los Angeles County.

Given the distinctive operational challenges faced in conducting NRFU during a period when COVID-19 incidence was generally high NRFU household interviews were completed in only slightly more than one-third of all non-responding households in California[19]. There is no doubt

---

[17] In the SJVCRP we observed that the sub-population of Hispanics most likely to respond to the 2020 Census and to contacts by enumerators in the course of NRFU were middle-aged naturalized Mexican immigrants (who have higher levels of educational attainment and English-language ability) than either heads of household in "mixed status" households (including undocumented and/or LPR or citizen individuals).

[18] I modeled the relationship between low response and eventual differential undercount using data from the San Joaquin Valley Census Research Project 2018 research—a survey of low-income Hispanic households in the southern San Joaquin Valley. See Ed Kissam, "How low response among Latino immigrants will lead to differential undercount if the United States' 2020 census includes a sensitive question on citizenship". Statistical Journal of the IAOS 35 (2019) 221–243 221 DOI 10.3233/SJI-19050. Based on assumptions about MAF quality, proxy interview response, patterns of response within the Latino immigrant population, and household structure, drawn from detailed survey data, the model projected a 51% self-response rate among Latino immigrant households in 2020.

[19] https://www.census.gov/library/visualizations/interactive/2020-census-operational-quality-metrics.html Census Bureau state-level reports of operational metrics in 2020 for California shows that 31.42% of households did not self-response and that 12.21% of the cases of non-responding households were resolved via direct household interview.

that the operational challenges were still more severe in low-income communities of racial/ethnic minorities living in crowded housing.

An additional consideration is that PES-based and DA-based national analyses of potential undercount in relation to race/ethnicity cannot capture the full extent to which socioeconomic and sociopolitical factors affect self-response and, ultimately, patterns of differential undercount at the local level. Fresno County overall is about 54% Hispanic but well-known correlates of census response vary tremendously: poverty at a level of about 46% in the hardest-to-count tracts and 7.4% in the easiest-to-count tract, 25.3% non-citizens in the hardest-to-count tracts vs. 4.0% in the easiest to count tracts. These same correlates can be presumed to negatively impact willingness to respond to NRFU—in-person interviews, but quite possibly, still more seriously, response to CATI or CAPI or exhortations to respond online.

How Does the ACS Do in Securing Response in Hard-to-Count Tracts of Fresno County?

Differential response is more problematic in the ACS than in the decennial census. Planning Database tabulations of 2013-2017 ACS response show that the lowest quintile of hardest-to-count tracts averaged 29.7% while the ACS response rate in the easiest to count tracts was 63.9%--a differential of 34.2%. It is not surprising that differential response gap in the ACS is higher than the differential self-response gap in the decennial census given the more limited resources available to the ACS for promoting response, and for addressing non-response.

Specific Evidence of Limitations to Census 2020 NRFU Methodology in Hard-to-Count Tracts with Concentrations of Low-Income Latino Immigrants?

In 2018, colleagues and I designed a survey (San Joaquin Valley Census Research Project) to assess the likely impact of the citizenship question on low-income Latino households. In addition to asking survey respondents about their willingness to respond to the "traditional" census (without the proposed citizenship question) and, subsequently, about a census with the question, field researchers asked about other factors that might affect census participation: availability of tablets or home computers for online response, living in a housing unit where they received US mail delivery, as well as willingness to participate in proxy interviews and ability to provide reliable responses if willing.

Address List Barriers to Self-Response

We learned that willingness to respond to a census without the citizenship question was high but that barriers to response were high too. About out of eight low-income Latino respondents (13%) only received mail at a PO Box; a similar proportion (12%), most of them living in "complex" doubled-up households, shared a mailbox for a single USPS/MAF address (where presumably only one census form or ACS survey would be sent). And 3% had no mail delivery at all. Would they be included in the census response from the "main" family at the address. Not likely (despite instructions to "include everyone living here").

About three-quarters of the respondents interviewed in our 2018 research did have Internet access, although it varied from place to place. However, most Internet access was via cellphone and less than half had a tablet or home computer to use in order to respond.[20]

Reluctance To Participate in Proxy Interviews

Census Bureau tabulation of operational metrics for Census 2020 in California shows that statewide about half of all cases of non-responding households were resolved via proxy interviews. However, we learned that only one out of five of the low-income Latino SJVCRP survey respondents (19%) would be willing to participate in a proxy interview to help an enumerator learn how many people lived in a nearby dwelling.

Probing about attitudes showed a strong consensus among respondents that it was not their business to provide that information, that it was dangerous, and that neighbors would be angry if they learned that someone had provided this information to strangers (even from the Census Bureau).[21] Moreover, one-third of the survey respondents willing to participate in a proxy interview said they really didn't know enough about their neighbors to respond with much certainty. Previous field research experience had shown us that landlords, especially those renting out unauthorized "hidden" housing units would not provide such information and that some, in fact, pressured their unauthorized tenants not to talk to a census enumerator if one stopped by.

Census Bureau release of operational metrics regarding reliance on proxy interviews as a substitute for direct response in the low-income hard-to-count Hispanics tracts (even at the level of two or three adjacent and similar census tract clusters) will be required to adequately assess data quality in these sorts of tracts.

Availability of Administrative Records in Tracts with Concentrations of Immigrants Lacking Legal Status

How useful might we expect the administrative records used in the decennial census and ACS to be for generating reliable data on non-responding households in tracts with concentrations of Hispanic non-citizens in Fresno County and other San Joaquin Valley agricultural counties? Half of our SJVCRP survey respondents (49%) were undocumented and residual estimates by experts—Enrico Marcelli and Manuel Pastor, and by Robert Warren at the Center for Migration

[20] Kissam, E., "How low response among Latino immigrants will lead to differential undercount if the United States' 2020 census includes a sensitive question on citizenship", Statistical Journal of the international Association of Official Statistics, June, 2019

[21] Gail Wadsworth, Ed Kissam, Cindy Quezada, Jo Ann Intili, "Troubled Reflections: Latino Immigrants' Thinking About Census 2020", San Joaquin Valley Census Research Project, February, 2019. This report analyzed details of qualitative data on beliefs, attitudes regarding Census 2020 (including details about responsiveness to proxy interview requests).

Studies of New York, show that about 42% of foreign-born residents 18+ years old in the region are undocumented.[22] Obviously, PIKs are not available for these individuals, nor is Medicare or SSA data.[23]

The Census Bureau has also relied on IRS data as a source of "high quality" administrative records data. However, there are many uncertainties about the extent to which undocumented households file income tax returns (e.g., using ITINs). But it is clear that not all undocumented immigrants file. One analysis—oriented toward demonstrating that some undocumented earners do file income tax returns estimates that about half of undocumented heads of household filed.[24] Conceivably, employer-issued W-2's and 1099's might provide useful data but many studies over the past several decades have shown there is widespread use of "borrowed" SSN's[25]

Another serious problem is that OMB/Census Bureau ambiguity in residence rules means that IRS tax return data for one household does not accurately capture data for other households living at the same address (supposed to be included on the census form but not supposed to be included in household income report).

Moreover, address matching in neighborhoods of low-income workers, some of whom move frequently due to problems paying rent is challenging. We know, also, that, of course, the landlords who rent rooms or sections of a house to other families typically do not report rent receipts and urge the other families in these "doubled up" households to not respond to contacts from any government representative.

The bottom line is that standard NRFU procedures are unreliable as a tool for overcoming the consequences of extremely high and uneven non-response in the 2020 decennial in these neighborhoods and communities that are already known to be the hardest to enumerate.[26]

---

[22] Ed Kissam computation for San Joaquin Valley Census Research Project based on analyses by Enrico Marcelli, Manuel Pastor, and methodology refined by Robert Warren. See, for example, Robert Warren and John Robert Warren "Unauthorized Immigration to the United States: Annual Estimates and Components of Change, by State, 1990 to 2010," in the *International Migration Review*, Vol. 47, No. 2, Summer 2013.

[23] Bhaskar, R., *et al*. "Assimilation and Coverage of the Foreign-Born Population in Administrative Records". CARRA Working Paper Series #2015-02. U.S. Census Bureau: Center for Administrative Records Research and Applications, April 21 2015; and in Statistical Journal of the International Association of Official Statistics, June, 2018.

[24] Institute on Taxation and Economic Policy, "Undocumented Immigrants' State and Local Tax Contributions", July 13, 2015. The ITEP model includes the estimate of income tax compliance and references several relevant studies.

[25] Sarah Horton, "Identity loan: The moral economy of migrant document exchange in California's Central Valley", American Ethnologist, Vol. 42, No. 1, February, 2015

[26] Research on the impact of the COVID-19 pandemic on farmworkers shows that household life was seriously disrupted by loss of employment and illness (California Institute of Rural Studies, 2020). The farmworker communities in Fresno County had, during 2020 (before vaccine was available) a cumulative incidence of COVID-19

The approach used by Rothbaum and colleagues for re-weighting ACS data seems very promising for "use cases" that rest on analyses of national data and, quite probably, even for some states and counties—but not all and not for jurisdictions with concentrations of unauthorized immigrants.

The limited availability and accuracy of SSA and IRS data in low-income communities with substantial non-formal economies, hidden housing units, and high concentrations of undocumented immigrants, imply that adjustment at higher geographic levels (national? state? county?) will slightly mitigate inequity stemming from differential undercount but that serious state-level and community-level inequities in funding allocation—due to undercount in the decennial and sample bias in the ACS and limitations to re-weighting and imputation will persist.

The Trajectory of Census Improvement—Mixed Evidence

The issues noted here about differential self-response giving rise to problematic NRFU (especially in tracts where enumeration workload is particularly high) is a chronic methodological challenge.[27] However, there were, as noted in the Census Bureau's discussion of challenges confronted in the 2020 ACS, also distinctive ones in 2020 related not only to the impact of COVID-19 constraints on NRFU but, also, the new, inadequately-tested introduction of online response.[28]

As Robinson and others have pointed out, the gap in response rates between easy-to-count and hard-to-count tracts widened dramatically in 2020—in large measure, because online self-response appears to have worked very well in easy-to-count tracts in Fresno County while not so well in most hard-to-count tracts.[29]

---

that was 1.5 time higher than in the county as a whole (Kissam 2020b). It is likely that this eroded propensity to respond to the census but still more seriously had a negative impact on NRFU.

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

[28] I have not been able to determine whether the Census Bureau tested its online response form for connection via mobile phone. Our observations prior to Census 2020 were that respondents who were not relatively digitally-literate would at least have a good deal of difficulty in entering their ID number. As can be seen in the CUNY Census 2020 database, high proportions of households in low-income areas that do have Internet access and may have been designated as Internet First have only smartphone online access. This is consistent with our SJVCRP survey responses throughout the San Joaquin Valley.

[29] For example, average self-response rate in the 6 tracts where >25% of households were update-leave was 49.1% as compared to 70.1% in the 172 tracts where <1% of all addresses were update-leave. Internet connectivity in the quintile of tracts with the highest HTC scores that, as projected, had the lowest levels of 2020 self-response averaged 68.1% as compared to 90.1% connectivity in the quintile of tracts with the highest HTC scores. J.G Robinson, "Combined Self-Response Rates, 2010 Response Rates, and 2013-2017 Planning Database Demographic, Housing for Tracts in Fresno County, California: Sorted on the Basis of Hard-to-Count Scores", March 20, 2022.

In Fresno County, the self-response rates in the easiest-to-count tracts in 2020 were 5.5% higher than in 2010 while they were (despite best efforts by state outreach contractors and the Census Bureau itself) 2.6% lower than in 2010. It is generally recognized that Internet availability varied greatly throughout the nation, in the San Joaquin Valley, and within Fresno County. Additionally, Latino immigrants' households' ownership of devices for broadband access varies greatly (with cellphone access to online resources very common but not optimal for census response).

Real-World Implications of Low Levels of Response to the ACS and Census 2020

The observations shared here are not meant to be a final or definitive assessment of the extent to which differential non-response and problematic NRFU have resulted in differential undercount in these communities. They are the tip of the iceberg for beginning to assess the extent to which 2020 Census data or ACS data will be "fit for use" in matters affecting the well-being of these communities and hundreds like them where a combination of factors led to low response at the same time that Internet response facilitated response in easy-to-count tracts.

In that context, national-level assessment of either ACS data or Census 2020 data is inadequate to assure equitable allocation of social program funding. The current PES-based estimate of net undercount of 4.99% of Hispanics at the national level, which represents a widening of the differential undercount estimates going back four decades to the 1980 Post-Enumeration Program, are very likely to have not adequately captured the full magnitude of inequity experienced by these communities. Director Santos' acknowledgement that municipalities impacted by differential undercount could "keep in mind" the PES and DA findings, and, implicitly, the ACS research findings, should not be simply a footnote in a report. It should, instead, be a basis for action—advocacy and litigation as needed by undercounted municipalities to assure appropriate adjustments to official tabulations of decennial census and ACS data used in standard funding allocation.

Recommendations put forward in the recent white paper by The Census Project, "America's Essential Data At Risk: A Vision To Enhance and Preserve the American Community Survey" (March, 2022)[30] provide a good foundation for proactive efforts to improve ACS data quality. However, improved "on the ground" ethnographic and mixed-methods survey research will so be needed to assure that Census Bureau data products used in a multitude of funding decisions assure economically-disadvantaged communities more equitable access to the funding they so critically need. Similarly, the Census Scientific Advisory Committee's first recommendation at its March 18, 2022 meeting highlights a practically valuable initiative that could support improved equity in funding allocation by adjustment in the population base derived from analyses of differential undercount in the Demographic and Housing Characteristics (DHC) file as a key component in the population estimates program.[31]

Ironically, it was the Census Bureau itself that pioneered the kind of groundbreaking research in the 1980's and 1990's that is now needed more than ever to address persistent and systematic methodological shortcomings in current Census Bureau data collection as the United States becomes increasingly diverse and the socioeconomic divide between "haves" and "have nots" widens.[32]

---

[30] https://thecensusproject.org/acs-americas-data-at-risk/

[31] Census Scientific Advisory Committee Memo to Robert Santos, March 18, 2022.

[32] Methodologies used by Kirsten West and David Fein in their "triple enumeration" research conducted as part of the 1987 Test of Census-Related Operations in the Los Angeles basin (Fein, David J. and Kirsten West. "The Sources of Census Undercount: Findings from the 1986 Los Angeles Test Census", Survey Methodology, December,1988.)

Sophisticated triangulation of data from diverse administrative records can make significant improvements to the usability of census-derived data. However, what is still more urgently needed is a renewed commitment to use available survey and ethnographic research methodologies to better understand and, hopefully, overcome the widening gap between the assumptions underlying traditional data collection methodologies and the 21st century realities of day-to-day life in an increasingly diverse nation where trust in government continues to decay.[33]

At the same time, recognizing the shortcomings inherent in analysis of differential undercount at the national level, it would be valuable to consider innovative approaches to enhance equitable allocation of funding at the local level. The current research products disseminated by the Census Bureau could and should be supplemented with analyses that enhance utility of tabulations of data in smaller geographies (state, county, sub-county levels).[34]

- *Ed Kissam is a researcher who has focused on farmworkers and immigrant settlement in the U.S. for three decades. Since the beginning of the pandemic, he has worked on strategic response to COVID-19 in farmworker communities as part of several collaborative networks of grassroots organizations, healthcare providers, and university researchers. He is co-trustee of the WKF Fund which provides support to a range of social justice initiatives.*

and the 31 ethnographic "alternative enumeration" studies designed by Leslie Brownrigg and Manuel De La Puente were exemplary. See De La Puente, Manuel, "Using Ethnography to Explain Why People Are Missed or Erroneously Included by the Census: Evidence from Small Area Ethnographic Studies", Center for Survey Methods Research, U.S. Census Bureau, 1993.

[33] See, for example, Putnam, Robert, "*E Pluribus Unum*: Diversity and Community in the Twenty-first Century: The Johan Skyte Prize Lecture", Nordic Political Science Association, 2007 for a powerful analysis of decreasing social trust. The Census Bureau's own research in its CBAMS initiatives has delineated other dimensions of decrease in trust. See also Czajka, John L. and Amy Beyler, "Background Paper--Declining Response Rates in Federal Surveys: Trends and Implications", Mathematica Policy Research Report, June 15, 2016.

[34] For example, analysis of PES data for hard-to-count tracts vs. easy-to-count tracts or well-defined strata of tracts (e.g. Hispanic-majority tracts vs. non-Hispanic White majority tracts, along with analysis of operational metrics for these distinct strata of census tracts could provide valuable insights to support synthetic adjustment. Similar analyses should, of course, be explored to improve equity in other areas and communities where differential undercount at the local level is likely to be significantly higher than at the national level due to concentrations of chronically-undercounted populations (e.g. Blacks and American Indian/Alaska Natives).