A Cascade Model: How Latino Immigrants' Lowered Response Will Lead to Differential Undercount in Census 2020 — Executive Summary

San Joaquin Valley Census Research Project
San Joaquin Valley Health Fund

By Edward Kissam

San Joaquin Valley HEALTH FUND

THE CENTER at Sierra Health Foundation

CALIFORNIA INSTITUTE FOR RURAL STUDIES

January 2019

Executive Summary

A Cascade Model: How Latino Immigrants' Lowered Response Will Lead to Differential Undercount in Census 2020

This is the second in a series of six reports by the San Joaquin Valley Census Research Project based on Fall 2018 survey research assessing the likely impact that adding a citizenship question to Census 2020 will have on Latino first- and second-generation immigrant undercount in the region.

The Census Bureau has been consistently optimistic about the viability of streamlined census procedures introduced as part of modernizing and re-engineering Census 2020. However, the new procedures, while cost-effective and probably satisfactory for easier-to-count populations and communities, are likely to have serious limitations when utilized in neighborhoods and communities such as those of the San Joaquin Valley.

Most problematic, the reliability of the planned data collection strategies has not been assessed or tested in the distinctive societal context where addition of the sensitive question about citizenship is expected to lead directly to pronounced response bias within an already hard-to-count population. We expect the re-engineered procedures will exacerbate differential undercount, at least in immigrant communities, and quite possibility in others with high proportions of low-income households.
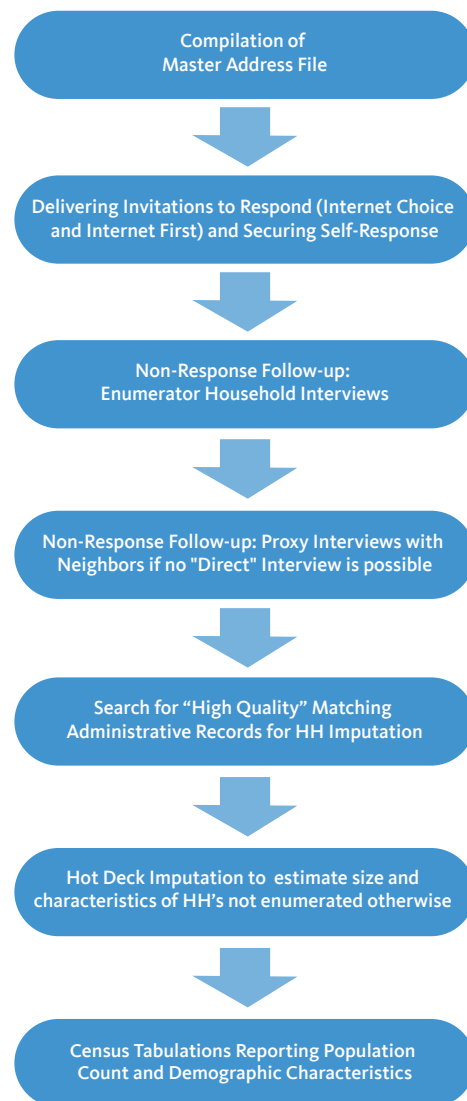
The cascade model presented here draws both on the research done in the San Joaquin Valley and presented in the first of these reports and on previous analysis and research. It describes how lowered response rates are likely to affect each stage in the census process of data collection/processing.

This San Joaquin Valley Census Research Project-based model explains how lowered response is transformed into undercount. It is referred to as a "cascade" model because the level of success and failure at each stage in decennial census operational procedures determines the parameters for census operations at the next stage. Level of self-response, for example, determines extent of reliance on enumerator efforts to secure information from the households that fail to self-respond. Enumerator success in this endeavor then determines the extent of reliance on proxy interviews for information on households. Cumulative success at this stage, then, determines the

extent of efforts to secure data from administrative records. And finally, cumulative success determines the need for reliance on count and whole-person imputation. The accuracy of the census count (and demographic profile of the population) depends on the level of reliance on each data-collection or analytic operation, since some (e.g. proxy interviews) are known to be more error-prone than others.

The cascade of census stages in decennial census collection/imputation are visualized in the cascade model of undercount of the San Joaquin Valley Latino first- and second-generation immigrants in Figure 1.

Figure 1—The Cascade Model of Diminishing Data Quality in Successive Census Operations

Compilation of Master Address File

↓

Delivering Invitations to Respond (Internet Choice and Internet First) and Securing Self-Response

↓

Non-Response Follow-up: Enumerator Household Interviews

↓

Non-Response Follow-up: Proxy Interviews with Neighbors if no "Direct" Interview is possible

↓

Search for "High Quality" Matching Administrative Records for HH Imputation

↓

Hot Deck Imputation to estimate size and characteristics of HH's not enumerated otherwise

↓

Census Tabulations Reporting Population Count and Demographic Characteristics

The current paper incorporates empirical data collected in the San Joaquin Valley Census Research Project survey to estimate the Census Bureau's success/failure in the following stages of the process: MAF-building, self-response, NRFU direct interview response, and proxy interview response. It also relies on the project's survey data to estimate the systematic undercount resulting from use of hot-deck imputation due to differences in size between the households likely to respond to the census and those likely not to respond.

Despite the valiant but compromised efforts by the Census Bureau to generate accurate census tabulations in the face of greatly elevated and uneven non-response, we believe the result will be attrition in data quality that ultimately results in flawed tabulations of both the size and demographic characteristics of the region's population.

The stakes are high for the San Joaquin Valley because the hard-to-count population of Latino immigrants makes up more than one-third of the entire population in the region. The model estimates the level of Latino immigrant undercount in the region as being 11.7% if the citizenship question is added. Given the size of the Latino immigrant population likely to be undercounted, it is reasonable to expect a 4.1% undercount in the total population of the San Joaquin Valley.

One of the issues it will be particularly important to consider is not just the overall flawed tabulations, but the differential undercount of sub-groups within the Latino population. Different levels of census response among undocumented immigrants, legal residents, naturalized citizens and the U.S.-born second-generation will skew the census-derived demographic profile of Latinos, as well as give rise to geographic disparities in census count. The model also identifies some potential ways for the Census Bureau to collaborate with local stakeholders in combined efforts to ameliorate likely undercount. This paper makes it clear that "Get Out The Count" campaigns focused primarily on impacting respondent motivation will not yield adequate results unless they also incorporate strategies to improve operational processes of census data collection.

It appears that the Census Bureau's view about the efficacy of its procedures to "cure" widespread non-response stemming from inclusion of the citizenship question in Census 2020 is misplaced. The cascade model in its present (initial) stage is essentially an exercise in hypothesis generation—tracing how patterns of non-response ripple onward through NRFU into flawed tabulations. We cannot yet definitively determine the model's predictive accuracy, in part because details on some aspects of Census 2020 operations (particularly those relating to reliance on administrative records and algorithms for hot-deck imputation) are unclear or unavailable.

We also recognize that the San Joaquin Valley Census Research Project initial findings about the prevalence and structure of complex households need to be further researched, due to the variety of housing accommodation and living arrangements, and the need to better understand how adding a citizenship question would exacerbate pre-existing patterns of partial household undercount in these sorts of crowded housing. The contribution of the project's initial research in this specific area is to highlight issues that have not yet been adequately addressed by the Census Bureau.

The current analysis and estimate of Latino immigrant undercount can and should be refined as Census 2020 operations are finalized. Nonetheless, we think it is critical at this juncture of census planning to think clearly and practically about operational adjustments that might contribute to an accurate and fair census.

Our hope is that the analysis presented here provides a useful framework to re-assess how the re-engineered decennial census operations will affect differential undercount in different regions and among ethnic groups with specific demographic profiles. The model's projection of the likely magnitude of differential undercount in Latino immigrant communities, even if subsequently adjusted, suggests the need for a commitment to carry out the research needed to yield fine-grained measurement of Census 2020 differential undercount and to use ethnographic research and demographic analysis in addition to dual-system estimation. It is unfortunate that the Census Bureau's ethnographic research efforts, which so powerfully illuminated crucial understanding of multiple causes of differential undercount, have languished over the past decade. Such research might well have shown, as the

San Joaquin Valley Census Research Project has sought to do, that the dynamics of census undercount cannot be adequately understood in isolation, that real-world context and operational implementation need to be carefully considered concurrently.

Being a work in progress, the San Joaquin Valley Census Research Project will update the analyses in the model to incorporate forthcoming survey-based findings about patterns of census response among non-Latino immigrants in the San Joaquin Valley when they become available in February 2019.

Meanwhile, we encourage readers who are concerned about the possibility of differential undercount in communities with high concentrations of low-income minority and immigrant households to consider using the cascade model analytic framework in combination with local survey and ethnographic research to examine the distinctive configuration of operational risks they face if Census 2020 includes the citizenship question.

www.shfcenter.org/sjvhealthfund