



Estimating the Potential Impact of LUCA-based Census Address List Improvement on California Differential Undercount¹

by Ed Kissam

Overview

Census undercount is correlated with a broad range of factors which include individual respondent socioeconomic characteristics, household characteristics, and neighborhood and community characteristics. Consequently, different ‘undercount’ analysts focus on one set of factors or another. Historically, the most common focus has been on the well-established correlation between race/ethnicity and undercount.

Beginning in the 1970’s, Census Bureau researchers began to look more deeply into the ways in which social system functioning and census operational systems interacted to give rise to differential undercount of minorities. This research literature - which includes post-enumeration surveys in past decennial censuses, ethnographic studies of the causes of census undercount, and a unique study of undercount conducted as part of the 1986 Los Angeles Test census² - provides valuable insights about the importance of Census Bureau’s Master Address File (**MAF**) quality as a foundation for a successful census. However the implications of this research for understanding how important census address list quality was for census accuracy, and how omissions of addresses from the Census MAF compromised census quality, have not received much attention.

These studies and their implications are crucial because they provide insights about the proportion of aggregate census undercount which can be attributed to total household omission (which is highly correlated with an address not being in the MAF), and the extent to which undercount stems from partial household omission (typically thought of as resulting from respondent error). This, in turn, makes it possible to estimate the specific contribution that MAF improvement can make to census accuracy.

Strategic Planning to Improve the MAF and Estimates of MAF Improvement Impact

In order to plan strategically to improve MAF quality and reasonably assess the likely impact of MAF improvement on ultimate census accuracy, it is useful to focus on neighborhood and community characteristics in the course of the Census Bureau’s own address canvassing and in



LUCA (Local Update of Census Addresses) - based efforts to improve the MAF. This provides the most practical guidance for targeting MAF improvement efforts. Ideally, such targeted local address canvassing efforts, jumpstarted in the implementation of LUCA, might morph into an ongoing “continuous” MAF improvement efforts. Given the introduction of the option of online census response as a mode for census participation, such efforts might be extended up to, say, one month before Census Day.³

In this paper, two approaches are described for estimating the extent to which LUCA-based local address canvassing might improve the Census Bureau’s Master Address File (MAF) and, thereby, decrease differential undercount and improve overall enumeration and census accuracy. The analysis presented here is for California but the approach described can be used in assessing the benefits of well-targeted local in-field address canvassing in any state.

The first model for estimating the potential impact of MAF improvement is based on readily-available data on *local poverty rates* (which is strongly, but imperfectly, correlated with prevalence of low-visibility housing units, “unusual” housing accommodations, crowded “complex” households). The second model is estimated directly based on assumptions regarding the prevalence of low-visibility housing based on the Los Angeles research and Census Bureau analysis of occupied housing units omitted in Census 2010. In both models, it is assumed that census tract level data on the demographic/socioeconomic profile of the local population and characteristics of local housing can be coupled with attention to data on *local concentrations of immigrants* and, particularly concentrations of undocumented immigrants to further enhance efforts to identify and target neighborhoods with higher concentrations of low-visibility housing. Data on the distribution of undocumented and recent immigrants, sub-populations particularly likely to live in low-visibility housing are currently available from Center for Migration Studies or can be straightforwardly derived from analysis by the University of Southern California’s Center for the Study of Immigrant Integration to determine the distribution of immigrants by legal status in census tracts with significant numbers of foreign-born residents.

There are very good reasons to believe that examining both sets of data (from the Census Bureau’s Planning Database and analysis of the distribution of undocumented immigrants in California) provides the best basis for estimating prevalence of low-visibility and concealed housing quarters - and, therefore, will yield the best indicator of impaired MAF quality and provide a good basis for locally targeted improvements to the MAF. It has not yet been possible to integrate the ACS-based data on the distribution of undocumented immigrants, and their key socioeconomic characteristics, into the analysis. However this supplemental analysis would be helpful in further refining the model presented here for targeting address canvassing. This is because the most recent immigrants are often the most socially and economically marginal, and are most likely to live in sub-standard housing, and in situations where the housing is actively concealed.⁴



Variations in MAF Quality

An important implication of multi-variate models of census undercount which include neighborhood characteristics is that MAF quality will vary greatly from neighborhood to neighborhood. While this presents challenges in assuring the reliability of a state-level estimate of MAF quality and potential improvement, it also provides good strategic guidance by indicating the importance of well-targeted and well-managed local address canvassing as part of LUCA—since “local knowledge” can provide additional guidance to supplement the analysis of extant data from the ACS and local government sources.

The optimal strategy for cost-effective MAF improvement will, as always, be to deploy inevitably limited resources to focus on adding low-visibility addresses in neighborhoods where overall housing patterns suggest the MAF is worst while, at the same time, remembering that there may well be deep pockets of low-visibility housing in otherwise affluent neighborhoods. Local LUCA managers’ ability to discern “where to look”, coupled with local address canvassers ability in knowing “how to look”, i.e. how to establish trust of potentially reluctant residents, communicate well with them, knowing clues that there is a concealed housing unit at a location, can be powerful tools for effective MAF improvement.

LUCA-based address canvassing, by relying on local knowledge, can powerfully supplement the number of census blocks to be targeted for address canvassing/MAF improvement by eliminating “false positives” (areas, for example, where poverty is prevalent and concentrations of immigrant families but regular, relatively high-visibility housing units) and “false negatives” (areas such as north San Diego County where encampments of recently-arrived immigrant farmworkers live in cardboard sheds in ravines in very high-income areas).

The current, preliminary model detailed here makes reference to data patterns observable in the Census Bureau’s planning database which provides a basis to identify “low response” census tracts, block groups, and blocks based on analysis of 2010 decennial census data and American Community Survey (ACS) data from 2008 and 2014.

However, it is important to remember that “low response” refers to rates of mail return of census forms, not specifically to MAF quality (since some high-income households may be as likely to not return a mailed census form as some very low-income households and, in fact, the Census Bureau does not and cannot uniformly measure MAF quality). There is some correlation between “low response score” and low MAF quality but it is imperfect. However, because the database is so comprehensive additional analyses focused specifically on assessing MAF quality (not mail response) are feasible. The WKF Fund has recently begun exploratory research, using quantitative and qualitative research methods, to develop an enhanced approach to identifying and targeting census tracts with low-quality MAF.



Extending and Refining This Analysis

The current modeling exercise provides only a bird's-eye view of the most obvious and direct financial benefits to California and Californians from MAF improvement. It is, nonetheless, very important to recognize that decennial census quality (as well as operational procedures pioneered in the decennial) profoundly affect the quality of American Community Survey (ACS) data which is actually the primary data source used for most business and social program planning. This is because differential undercount of the most economically and social marginal sub-populations skews the demographic and socioeconomic profile of all communities. Improving the quality of ACS-based profiles has tangible, but complex and difficult-to-estimate positive implications and important implications for California's Citizens Redistricting Commission which will rely on ACS data to assure that redistricting after 2020 will be equitable.

The analysis presented here is based on the Census Bureau's data at the tract level but, ultimately, it is likely that the best basis for targeting LUCA-based local address canvassing will be at the block group or block level of census geography. Local LUCA-based review of target areas can be expected to further refine the targeting to the blocks within each tract with the most low-visibility housing units. Better spatial resolution in targeting efforts toward MAF improvement would, of course, be expected to result in more cost-effective MAF improvement processes.

Whichever level of spatial resolution is available in standard data used for targeting address canvassing (e.g. block, block group, tract, PUMA) an important benefit of LUCA-based local address canvassing is that local knowledge provides valuable insights into the specific sorts of housing infrastructure which likely to have low-visibility housing accommodations (e.g. invisible, occupied basements or garages, clusters of "unusual" housing at a property with a single postal address).⁵

The ideal approach for optimally targeted local address canvassing would be to test a few variant approaches to implementing LUCA-based local address canvassing (e.g. alternative targeting models, types of partnerships between local government and grassroots organizations intimately familiar with local low-income housing and hard-to-count populations). However, it is not yet clear if that would be feasible given the Census Bureau's operational timeline and uncertainties about how such pilots could be funded.

Importance of Decreasing the Number of Addresses Missing in the Census Address File (MAF)

In reporting their findings from the Causes of Undercount Survey Fein and West (1988) state that 50% of all "non-match" cases were ones where the address was not on the census address list, i.e. missing from the MAF. This results in total household omissions.



It is worthwhile to note, however, that the extent of total household omission (due to a household not being included in the MAF) varies from community to community, neighborhood to neighborhood. Research on census undercount in California farmworker communities reported aggregate omissions, as had Fein and West, and categorized them either as “partial household omissions” (some individuals living at an address were not included in census responses) or “total household omissions”.⁶ Total household omissions ranged from 6.4% in Arvin (Kern County) to 27.6% in Parlier (Fresno County).⁷ Overall 24% of the low-income immigrants and farmworkers in the study neighborhoods were left out. Ultimately, total household omission in the five rural communities in the study accounted for 58% of this aggregate undercount. That is, 13.9% of the census omissions in these rural low-income neighborhoods resulted from entire households being missed.

In a similar coverage measurement study in connection with Census 2010, focusing on 33 “hard to count” census tracts in eleven rural California counties in the San Joaquin Valley, Central Coast, and South Coast, it was estimated that total household omission accounted for about 7% of a aggregate undercount of about 9.9%-- i.e. two-thirds of the entire undercount.⁸ It is useful to note that overall census operations in the Census Bureau’s Los Angeles region where the study took place had improved in the decade from 2000 to 2010 and also that our 2010 study areas were more diverse and had less dense concentrations of immigrants than in the research on undercount in rural communities, i.e. the research area had 75% Latino households, of which 44% were farmworker households.

Nonetheless, we found correlates of undercount similar to those reported by Fein and West. For example, we found that 27% of the housing units categorized by our field researchers as “unusual” had not been enumerated while only 3% of the households living in standard, usually single-family, homes had been missed. We also found that 17.7% of the households where a recent immigrant was the census respondent had not been enumerated, although Census Bureau mailings of Spanish-language census forms had dramatically improved response.

Proposed Strategy for Targeted LUCA-based Local Address Canvassing to Improve the MAF

The estimate of the potential impact of MAF listing assumes that it is feasible to target for LUCA-based local address listing, approximately 20% of all the census tracts in California, that is, those where the poverty level is 20% or higher.⁹ This would entail address canvassing in approximately 1,500 census tracts presumed to be “low-quality MAF”.

Narrower targeting would decrease costs but also decrease the number of non-listed housing units which could be added; broader targeting might be desirable but the associated costs might be prohibitive.¹⁰



Furthermore, from a research management perspective, it would be better to focus more on quality of in-field address canvassing than on quantity. This underscores the importance of sound training and supervision for in-field canvassers to assure they can reliably detect likely low-visibility housing units and effectively interact with residents to determine actual housing arrangements even when housing units or specific housing arrangements at a place are generally actively concealed. Therefore, the estimation of the potential impact of MAF improvements offered here is based on efforts to improve the MAF only in tracts with poverty rates of 20% or higher.

This targeting strategy which assumes that LUCA-based MAF improvement can reach each of the census tracts with >20% poverty rates takes into account the fact that more than 20% of the population in those targeted census tracts are poor (since the average poverty rate in these tracts is about 29%); while the low-visibility households in the 70% of non-targeted tracts (where the poverty rate averages about 8%) will still be missed.¹¹ The higher-than-expected prevalence of poverty in the targeted tracts is likely to offset the fact that there would be no MAF improvement and, thus, no impact in the non-targeted tracts.

Two Models for Estimating the Impact of LUCA-based MAF Improvement on Census Accuracy

There are, in principle, several alternative ways to estimate the potential impact of MAF improvement on eventual differential undercount in Census 2020. Two models are presented here, both based on the research on “structural” causes of undercount:

- ***Model 1 Based on Population Profile and Risk of Total Household Omission in Poor Neighborhoods:*** immigrant-head households and US-born headed households are missed at different rates so analysis of population profile in low-income neighborhoods with concentrations of minorities and immigrants provides an indicator of possible low-quality MAF and, thus, the likely impact of MAF improvement. This model indicates a potential improvement in the number of housing units captured in the address list of 2.54% in the targeted “hot zones” with low MAF quality and an overall statewide improvement in differential undercount of 0.63%.
- ***Model 2 Based on Incidence of Low-Visibility Housing in Poor Neighborhoods*** Different sorts of low-visibility housing units in low-income neighborhoods with higher-than-average concentrations of minorities and immigrants are omitted from the MAF at different rates so analysis of Los Angeles Test Census data provides possible low-quality MAF and, thus, the likely impact of MAF improvement. This model indicates a potential improvement of 3.24% in the targeted “hot zones” with low MAF quality and an overall statewide improvement in differential undercount of 0.66%.



These models provide independent methodologies for estimating potential impact of MAF improvement via LUCA-based local address canvassing. However, both models assume that the same operational strategy is used, that is, one which identifies “high risk” census blocks or tracts where MAF quality is likely to be lower and conducts local in-field canvassing in those areas to add low-visibility housing units.

It deserves note that the Census Bureau’s analysis of Census 2010 data shows that nationally 1.9% of all occupied housing units were omitted and that omissions of renter-occupied housing units, for example, was still higher: 2.5%.¹² The models used in the analysis, therefore, require only that LUCA-based local address canvassing successfully identify about one-third of the low-visibility housing units missed in 2010.

It will certainly be possible to further refine both models (or develop a still-better one) but the modeling exercise here makes it clear that wise investment in LUCA-based local address canvassing will yield a dramatic return.

Part of state-level planning for a program of state assistance to counties and municipalities to implement LUCA-based local address canvassing should be to refine the general model presented here to target LUCA assistance as well as possible to low-quality MAF areas. A complementary element at the level of local government planning should be to identify, within the boundaries of standard census geography (e.g. census tracts) where to most usefully focus limited in-field canvassing efforts. Sound management, including guidance on effective implementation of in-field address canvassing in neighborhoods with concentrations of low-income minority and immigrant residents, will significantly affect the eventual impact.

Both of these preliminary models for estimating the potential impact of MAF improvement on census accuracy in 2020 discussed here look at the impact stemming from improvement in both census enumeration of California immigrants and their children and native-born families. They incorporate the assumptions stemming from the practical limitations of a MAF improvement initiative, i.e. the need to target wisely given finite financial resources available for LUCA.

Both models assume California’s MAF improvement efforts should be targeted primarily to areas with higher levels of poverty and higher proportions of relatively recent immigrants (available from the Center for Migration Studies’ database—based on 2014 ACS data) because it is likely that the poorest families are those who are most likely to live in the sorts of low-visibility and/or “unusual” housing which is most likely to be omitted from the MAF.

Both models assume there will be no dramatic changes in federal census-driven funding and that the historic relationship between the “official” poverty level and prevalence of sub-standard, low-visibility housing remains unchanged (although there is, in fact, much evidence, e.g. in the San Francisco Bay area), that housing costs continue to account for greater and



greater proportions of family income, resulting in more crowding and “unusual” living arrangements.

Neither model presented here takes into account the possible impact that LUCA-based address canvassing might have if there were ways to identify and facilitating reporting by “households” (distinct social/family units) within housing units which are typically categorized as “partial household omissions” due to the Census Bureau’s decision to interpret the ordinary language ambiguity of “household” to reference people living under one roof rather than as referring to “social/family unit” (the concept which drives a good deal of partial household omissions).¹³

The modelling exercise leaves no doubt that there are huge potential returns on investments in supporting LUCA as the most practical approach to MAF improvement.

Data Elements in Model 1

Fein discusses the findings from the “triple enumeration” analysis—of data from the test census itself, the PES component of the test census, and the “Causes of Undercount Survey” (CUS) which provides particularly good evidence of the extent of high immigrant undercount—with omission rates of 15% for households where the respondent spoke little or no English, 11% for cases where the respondent had no high school education, and 13.8% for non-citizens.

Model 1 requires specification of the proportions of adult immigrant heads of household and their foreign-born children, US-born children of immigrants, and adult US-born heads of household.

- ***Proportion of Immigrants in California***

Based on ACS data which, of course, underestimates the number of immigrants in California due to differential undercount, the Public Policy Institute of California (PPIC) estimated that, in 2011, immigrants made up 27% of California’s population. California’s immigrant population has probably not increased since 2011 but because the ACS data are likely to incorporate an undercount, I estimate that 29% of Californians are immigrants.

- ***Proportion of US-born children in California in a household with an Immigrant Parent***

The Migration Policy Institute, using 2014 ACS data, estimates that there are about 4,232,000 children of immigrant parents in California and that 90.3% are U.S.-born, meaning that about 3,867,000 are US-born children of immigrant parents—about 10% of California’s total 2014 population of 38.8 million.¹⁴ Therefore, improvements in census enumeration of immigrant-headed households will also dramatically increase the census enumeration of their non-immigrant U.S.-born children.



- ***Proportion of US-born adults in California***

Since 29% of California's population is foreign-born, the remaining 71% are U.S.-born. However, since 10% of the US-born population are children living in a household with an immigrant parent, US-born adults make up only about 61% of California's total population.

- ***Poverty as a correlate of census omission***

Prevailing levels of poverty in a neighborhood is arguably the most straightforward indicator to use as a proxy for higher-than-average prevalence of low-visibility, crowded, and or/ "unusual" housing units and, thus, low-quality MAF, areas where there is a high risk of total household omission. While the current preliminary model was developed based on the Census Bureau planning database variable of % of a census tract in poverty, a more refined model might use an indicator based on prevalence of "deep poverty" (with a benchmark set at some lower level, e.g. 70% of poverty level, or higher level in areas with unusually high local housing costs). Fein's analysis of national post-enumeration survey (PES) data from the 1980 census shows that the overall census omission rate was 9.1% for households in poverty, i.e. those with 1979 incomes of \$7,500 or less.¹⁵ Poverty is also co-variant with lower levels of education, respondent being a racial/ethnic minority, limited-English, and other indicators of socioeconomic disadvantage which correlate with undercount so this suggests that targeting based on poverty, perhaps supplemented with additional ACS-derived variables, is viable.¹⁶

The Public Policy Institute reports that in 2014, that 16.4% of Californians lived in poverty, but that using a more comprehensive indicator developed by PPIC and the Stanford Center on Poverty and Inequality, the California Poverty Measure (CPM), the poverty rate would be 21.8%. Additional analyses from the Center show that 29% of California immigrants live in poverty (vs. 19% of non-immigrants).¹⁷

Based on these considerations, Model 1 incorporates a conservative estimate that 25% of immigrants and 15% of non-immigrant census respondents live in low-income neighborhoods where the risk of housing accommodations being omitted from the MAF is higher than average and where there will be on-the-ground MAF improvement efforts.

Model 2- Proportion of Low-visibility Households Likely to Have Been Omitted from the MAF

The second model utilizes only the observed prevalence of low-visibility housing and observed levels of census omission for populations living in each of two types of low visibility housing: "attached units" and "strange or unusual types of housing or location" in the Los Angeles 1986 Test Census area (with more than 4,000 housing units in the Causes of Undercount Survey sample).



In the 1986 Los Angeles Test Census analysis Fein reports that 6.7% of the housing in the study was classified as a “strange unit type of location” and 17.6% of this housing was omitted, i.e. an overall total household omission rate of 1.19%. Another 8.9% of the housing consisted of low-visibility “single family attached units” and 23.5% were omitted, i.e. an overall household omission rate of 2.09%

This results in an aggregate household omission rate of 3.28%; and, by adding low-visibility housing units in areas where the MAF is of low quality, provides an indicator of the level of possible MAF improvement in those areas.

Financial Impacts On Census- Driven Federal Funding Stemming From Improved Enumeration

The estimate of financial impacts from MAF improvement is based on the Brookings Institution’s excellent analysis of overall census-driven federal funding nationally. This paper, however, adjusts the dollar amounts to account for projected rates of inflation and population growth in the next 5 years. The assumption used here is that California’s 2020 population will be about 40 million and federal funding in California will be about \$83 billion per year.

The next sections examine the different estimates of impacts generated by the two different analyses of strategies for MAF improvement.

Model 1 Calculations for estimating impact of LUCA-based MAF improvement

This model for estimating the potential impact of MAF improvement on census enumeration in 2020 projects financial impacts in the post-census decade is based on the following assumptions:

- 20% of California’s immigrant live in low-income neighborhoods where there is high risk of an incomplete MAF and total household omission, while 15% of California’s native-born population live in these low-quality MAF neighborhoods
- LUCA-based MAF improvement by reliance on local government-community partnerships to conduct local address canvassing can decrease total household omissions for immigrants living in the high-risk neighborhoods by 4% and for non-immigrants by 2%.

Extent of MAF improvement in California Census 2020 enumeration is then estimated as a function of the number of Californians living in a neighborhood with a high risk of incomplete MAF and the likely prevalence of households omitted from the MAF in these areas, which then indicates the amount of possible improvement which could be expected as a result of adding low-visibility units



The calculated result is that LUCA-based MAF improvement might decrease aggregate California undercount by about 0.63%--with a 0.29% of the improvement stemming from additional enumeration of foreign-born adults and children, 0.18% from additional enumeration of non-immigrant adults, and a 0.16% stemming from additional enumeration of US-born children living in households with an immigrant parent. This implies additional enumeration of about 200,000 persons stemming solely from MAF improvement and a corresponding impact of about \$525 million per year in census-driven federal funding for California in the 2021-2030 decade.

An advantage of Model 1 is that it provides insight about the extent to which MAF improvement will benefit from using a targeting algorithm which assures intensive local address canvassing in immigrant neighborhoods reflecting the particularly high rate of immigrant housing which is “low visibility”.

Model 2 Calculations for estimating impact of LUCA-based MAF Improvement

Using the alternative model based on Fein’s report of the prevalence of low-visibility housing units (separately estimated for “attached units” and “strange or unusual” units in the Los Angeles Test Census) and assuming that these can successfully be added to the MAF in each of the 20% of high-risk California census tracts to be targeted it appears that targeted LUCA-based local address canvassing might be expected to decrease the aggregate California undercount by 0.66% yielding an increase of about \$545 million per year in additional census-driven federal funding.

Range of Likely Financial Impacts of LUCA-based MAF improvement as a result of local address canvassing

In summary, the benefits of MAF improvements as a result of a targeted program of LUCA-based local canvassing reaching about 20% of California’s tracts where MAF problems are concentrated could produce benefits of at least \$400 million per year in increased census-driven federal revenues.

Both models described here can and should be further refined. This preliminary modeling effort shows clearly that locally-targeted LUCA-based local address canvassing can make a dramatic positive contribution in working toward the goal of a California Complete Count. Such an effort is particularly important in the context of the Census Bureau’s reduced budget and revised operational plans for its own address canvassing (AdCan).

If we were to assume that targeted LUCA-based address canvassing were only able to successfully add half of the low-visibility addresses omitted from the MAF the benefits



stemming from an increase in census-driven federal funding would still amount to more than \$250 million per year.

Clearly, an important component in implementing a strategy of LUCA-based local address canvassing with local government partnering with grassroots organizations to add low-visibility housing units to the census address list is effectiveness of canvassing. A GAO report reviewing California Rural Legal Assistance's program where low-visibility addresses identified by community workers were submitted directly to the Los Angeles Census Bureau Regional Office in Census 2000 determined that 73% of those submitted were valid, suggesting that local address canvassing is, indeed, viable.¹⁸

END NOTES

¹ LUCA is a Census acronym denoting their operations related to Local Update of Census Addresses. Historically, LUCA has been based primarily on “in-office” review of address lists but there has been some experience with proactive street-level “in field” address canvassing to add low-visibility housing units.

² This research is detailed in two major research documents. The first is David J. Fein and Kirsten West, “The Sources of Census Undercount: Findings from the 1986 Los Angeles Test Census”, *Survey Methodology*, December, 1988. The second is, David Fein, “The social sources of census omission: Racial and ethnic differences in omission rates in recent censuses”, Ph.D., dissertation, Princeton University, 1989. Findings from the ethnographic/alternative enumeration studies are summarized in Manuel de La Puente, “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. De La Puente’s summary includes a bibliography listing all 29 studies. Studies which are particularly relevant to MAF improvement include those by: Pamela Bunte and Rebecca Joseph (on Cambodians in Long Beach), Boanerges Dominguez and Sarah Mahler (on Mexicans in the South Bronx), Dale Martin Montoya (on farmworkers in Woodburn, OR), Nestor Rodriguez and Jacqueline Hagan (on Guatemalans in Houston), and Alex Stepick and Carole Stepick (on Haitians in Miami). Philippe Bourgois’s research on inner-city attitudes about census response are particularly relevant to understanding the challenges in neighborhoods with actively concealed irregular housing arrangements.

³ The desirability of “continuous” MAF improvement is a recommendation forcefully presented in Joseph Salvo 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



⁴ Additional Census Bureau and local data on local housing conditions, while not included in this analysis, are also relevant for targeting and for developing refined estimates of the local financial impact of adding low-visibility housing units in areas where the MAF is of relatively poor quality.

⁵ In my research on census undercount of farmworkers in rural California communities, for example, Oxnard had much fewer total household omissions than in some other communities because there are many houses where there have been a succession of additions to a single-family dwelling which results in partial household undercount. In contrast, Parlier typically had “back houses” (Jennifer Sherman, Don Villarejo, and Anna Garcia, “*Finding Invisible Farmworkers: The Parlier Survey*”, California Institute of Rural Studies, 1997). The ratio of total and partial household omissions is a function of housing/living patterns.

⁶ Ed Kissam and Ilene Jacobs, “Practical Research Strategies for Mexican Indigenous Communities Seeking to Assert Their Own Identity”, in Jonathan Fox and Gaspar Rivera-Salgado (Eds.) *Indigenous Mexican Migrants in the United States*, Center for U.S.-Mexican Studies, 2006. The research strategy in this study was adapted from the methodology used in the Census Bureau’s Center for Survey Methods Research ethnographic alternative enumeration studies developed by Leslie Brownrigg and utilized in the 29 ethnographic studies in the 1990 decennial census.

⁷ A resident ethnographer (Aline Doignon) and consulting anthropologist (Anna Garcia) from my research team subsequently mapped Arvin block-by-block in 2003 in preparation for a USDA-funded study of immigrant settlement in rural areas of the U.S.--the New Pluralism Study. Our observation in the course of mapping Arvin in 2003 using the same methods we had used to map Parlier in 1989 explains the relatively low rate of total household omissions observed in Arvin in our study of Census 2000 undercount. The town has remarkably orderly housing and represents a better-than-average scenario for rural MAF development. Ed Kissam, “Migration networks and processes of community transformation: Arvin, California and Woodburn, Oregon”, *Journal of Latino and Latin American Studies*, 2007.

⁸ Ed Kissam, “Census Enumeration of Immigrant Communities in California: Dramatic Improvements but Challenges Remain”, Conference on Hard-to-Survey Populations, American Statistical Association, December, 2012. The methodology used here was an adaptation of the Census Bureau’s post-enumeration survey approach.

⁹ Based on a California extract from the Census Bureau’s Planning Database.



¹⁰ The proportion of census blocks or tracts targeted in a community for local address canvassing in LUCA is obviously a function of available funding to support in-field canvassers. Higher levels of in-field address canvassing will further improve the MAF.

¹¹ Average poverty rates computed from the Census Bureau's Planning Database without adjusting for variations in census tract population (which is relevant but which would have been a labor-intensive task). The model can also be refined by factoring in data on local housing costs so as to adjust expectations regarding the relationship between income and quality of housing available to low-income families.

¹² Census Bureau, Decennial Statistical Studies Division, "2010 Census Coverage Measurement Memorandum Series #2010-G-06" (2012). This analysis, based on data from the Census Bureau's 2010 post-enumeration survey provides valuable information on the proportion of vacant and occupied housing units omitted in different types of neighborhoods.

¹³ Ordinary language dictionary definitions of "household" clearly show the ambiguous and potentially inconsistent semantics of the term which does, in fact, generally refer both to dwelling/housing unit and to family/social unit. This assumes that there is one family/social unit per housing unit. This is the case in the vast majority of U.S. households but not in the most economically stressed and socially marginal populations which are most severely undercounted. Given the nature of immigration, immigrants are particularly likely to live in crowded housing where multiple social/family units—several nuclear families, a nuclear family and unrelated individuals, a primary family and unrelated additional nuclear families share space but consider themselves to be distinct "households" (social/economic units).

¹⁴ PPIC, the Migration Policy Institute, and the Pew Research Center have published comparable estimates of the numbers of U.S.-born children with immigrant parents. These slight differences do not greatly affect the analysis. My estimate based on PPIC's estimate of children of immigrants as percent of California children and Pew's estimate of proportion of children of immigrants born in the U.S. and MPI's estimate both suggest that about 9.7% of California's population are US-born children of immigrants.

¹⁵ It might reasonably be assumed that decennial census quality would improve in the four decades from 1980-2020. This is not necessarily the case since lower budget authorization for the Census Bureau has compromised its' ability to implement key portions of the decennial census process—limiting the scope, for example, of test census operations, address canvassing, and approaches to non-response follow-up. See Terri Ann Lowenthal, Census Project blog, December 15, 2015 for a good discussion. The Census Bureau's Census 2020 Operational Plan also provides detailed information on re-engineered operations which includes a mix of promising and worrisome developments.



¹⁶ Chapter 3 Fein Ph.D. dissertation. Fein's logistic regression model for census omission in the 1980 PES dataset is a very detailed one. The most relevant observation here is that poverty was, as expected, co-variant with race/ethnicity (immigration status was not available) as was education and was statistically significant as a factor in the best model.

¹⁷ The California Poverty Measure, one of several "supplemental" poverty measures which have been developed, is very attractive for targeting LUCA in states with high proportions of undocumented immigrants such as California because immigrants who lack legal status have to stretch their earnings further to cover housing costs. The analysis here uses the "official" poverty measure because it is the variable which is available in the Census Bureau's Planning Database. Of course, the relationship between local poverty level and likelihood of living in sub-standard, low-visibility housing also varies in relation to local housing costs. Such additional adjustments would be useful in prioritizing census tracts or blocks for LUCA-based local address canvassing. For additional details on the California Poverty Measure, see Christopher Wimer et al, "Poverty and Deep Poverty in California", The Stanford Center on Poverty and Inequality, 2014. The measure is updated periodically see <http://inequality.stanford.edu/cpi-research/area/poverty>

¹⁸ GAO-03-60 Decennial Census: Lessons Learned for Locating and Counting Migrant and Seasonal Farmworkers (2003)