



Census Block-Level Tabulation of the Citizen Voting-age Population (CVAP) In Violation of Federal Law on Census Confidentiality

*by Ed Kissam and Jo Ann Intili,
WKf Fund, October 2018*

The Threat to Civil Liberties

The rationale given by Secretary Wilbur Ross, Department of Commerce, for adding a citizenship question to Census 2020 was that this change to the decennial census was necessary in order to accommodate the Department of Justice's request for tabulations of CVAP (the citizen voting-age population) at the census-block level in order to better enforce the Voting Rights Act. He argued this would result in better information on CVAP in local political jurisdictions.

Secretary Ross's proposal to allow tabulations of census data has not attracted much notice to date because the general public and policymakers think of a census block as a reasonably large area with lots of people so that individuals' responses are effectively confidential. That is not the case. Census blocks are defined by land area, not by population so, although the average census block has about 34 housing units (HU's) the number in any particular census block varies greatly from place to place. In California, for example, almost one-third (30.6%) of the census blocks in the state are "small blocks" with only 1-10 housing units in them.

Title 13 of U.S. Code forbids use of information provided by a census respondent for anything other than the purpose for which the information was collected.¹ Title 13 also prohibits disclosure of any household's individual census response.² Secretary Ross's decision to acquiesce to the Department of Justice's request violates both of these central provisions of Title 13 because the proposal includes: a) plans to utilize census-derived information for Department of Justice enforcement purposes, and b) census block level tabulations of CVAP in small census blocks that will result in disclosure of individual census responses.³

Tabulation of CVAP at the census block level would be a violation of federal law if it included the small census blocks. And if there were only tabulation of a purposefully selected sub-set of larger census blocks with more than 10 housing units, the already-questionable analysis of CVAP would inevitably become statistically worthless since the tabulation would omit almost one-third of the housing units in the state, affecting the integrity of CVAP data for all jurisdictions. This short paper discusses how the Department of Commerce's proposed illegal tabulation would affect the state of California and the people living in the state.

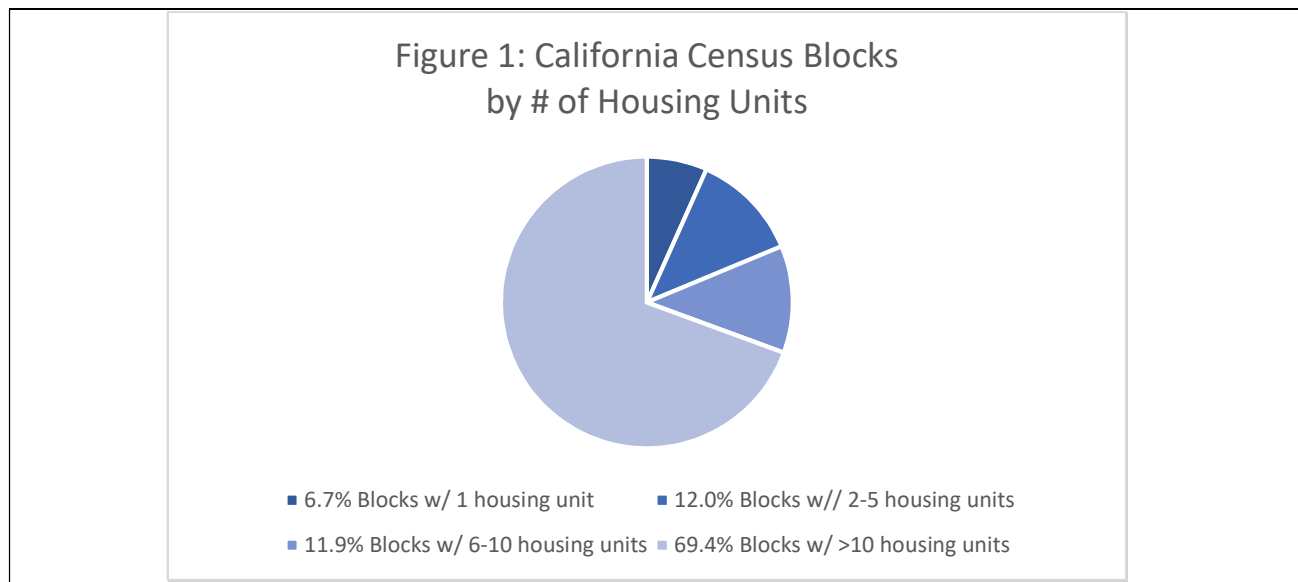


How Systematic Tabulation of CVAP at the Census Block Level Threatens Census Confidentiality

The proposed plan for tabulating CVAP at the census block level is a straightforward violation of Title 13 for census blocks with only one housing unit. Tabulation of census blocks with 2-10 housing units does not definitively violate Title 13 but does violate current guidelines for federal data tabulation promulgated by OMB and the Census Bureau. For example, in a tabulation reporting the citizen voting-age population in a census block with only 2 housing units, it might not be definitively clear which responses were from Household A or Household B but both would, understandably, consider that their privacy was threatened. And it would be.

Threats to Confidentiality at the State Level: The Case of California

This is not a small, marginal problem. It is a huge one. **Figure 1** below shows the size distribution of census blocks in California.



As can be seen in this figure, block-level tabulation leads to violation of Title 13 confidentiality safeguards in 6.7% of the census blocks of the state, poses extreme threats to privacy in the 12% of census blocks with 2-5 households, and overall compromises the privacy of the one in three California households in small census blocks with 10 or fewer households.

County by County Variations in Threat to Confidentiality

The extent to which block-level tabulation threatens confidentiality varies in relation to housing density. Review of the implications of block-level tabulation shows that the threat to privacy varies significantly from county to county.

Table 1 shows variations in the privacy impact of census block-level tabulations in five of California's 58 counties with distinct patterns of housing—Alameda, Los Angeles (primarily urban



but with some less-densely settled areas), Fresno, Kern (primarily rural agricultural but with urban metro centers), and Sierra (remote rural). It provides a grim picture of the extent to which block-level tabulation of CVAP would threaten the privacy of Californians in different areas of the state. It shows the prevalence of census blocks with only one household or very few households (1-10 HU/block) in the five counties highlighted here.

Table 1: Distribution of Census Blocks in a Sample of California Counties by Number of Housing Units

	ALAMEDA (Urban)	LA (Urban)	KERN (Agric.)	FRESNO (Agric.)	SIERRA (Remote Rural)	CALIFORNIA
# and % of Census Blocks w/ 1 HU's w/ 1-10 HU's per block						
Median HU's/block	26	29	14	16	3	19
% Foreign-born	32%	34.5%	20.3%	21%	4.3%	27%
Tot. Census Blocks w/ HU's	14,246	74,805	13,356	13,386	345	411,610
# of Census Blocks w/ only 1 HU	659	2,968	1,574	821	88	27,389
% of Census Blocks w/ only 1 HU	4.6%	3.2%	11.8%	6.1%	25.5%	6.7%
# of Census Blocks w/ 1-10 HU's	2,654	12,678	5,620	4,516	277	125,777
% of Census Blocks w/ 1-10 HU's	18.60%	17.1%	42.1%	33.7%	83.2%	30.6%

**Block data from U.S. Census Bureau January 18, 2018 census block by block address count file used by local government entities participating in LUCA. This analysis includes only census blocks with housing (some census blocks have no housing in them) and does not include GQ's. Data on foreign-born population are from ACS 2012-2016. Tabulations for all 58 California counties are available on request from edkissam@me.com.*

It is clear there is a significant number of California households (27,389, with about 80,000 people living in them) that are the only housing unit in the census block. Consequently, Census Bureau tabulation of CVAP at the census block level would inevitably result in clear-cut violation of Title 13 since the block-level tabulations of blocks with a single household would reveal the census responses of individual householders living in that single-household block. In remote rural Sierra County, for example, confidentiality would be violated for about one out of four households. In urban areas such as Los Angeles or Alameda counties, lower proportions, but larger numbers, of households would have their privacy irrevocably violated, for example, with about 3,000 housing units in Los Angeles County being the only one on their block.⁴



Statewide there are an additional 570,000 households, with more than 1.7 million Californians living in them, in census blocks with 2-10 housing units, where census confidentiality would be seriously compromised as a consequence of block-level tabulation. Although these tabulations would not definitively reveal an individual household's census response they could, if made available to non-statistical agencies such as the Department of Justice and the Department of Homeland Security, be used for non-statistical purposes. In the counties with high concentrations of foreign-born individuals, block-level tabulation of CVAP would be particularly threatening as they might be used for targeted immigration enforcement.

Census Block-level Tabulations Do Not Conform to Federal Guidelines on Safeguards for Tabulation of “Sensitive” Information

Secretary Ross's mandate violates current federal data suppression guidelines designed to guarantee confidentiality of “sensitive” information⁵—including citizenship status--in federal reports. Census Bureau guidelines currently suppress public reporting of ACS-derived citizenship data for all geographies below the census block group level—a geography with, on the average, about 500 households, and 1,500 people.

The justification for caution in allowing small-area tabulations of sensitive information collected by the Census Bureau is well-founded. Census Bureau block-level tabulations were the basis for targeted detentions of Japanese-Americans in World War II. After Census Director William Lane Anderson, who had opposed relaxation of confidentiality guidelines to allow these tabulations to be shared with the FBI and military intelligence agencies, was replaced with a compliant Census Director, J.C. Capt, the tabulations were authorized under the Second War Powers Act and used to detain Japanese-Americans in California, Arizona, Wyoming, Colorado, Utah, Idaho and Arkansas.⁶

Prospects for Resolving the Conflict Between the Administration's Plans for Small-area Tabulation by “Balancing” the Trade-off Between Confidentiality Protection and Accuracy

Current Census Bureau planning envisions achieving a “tunable” balance between accuracy and confidentiality.⁷ The consequences would still be serious and Census Bureau in-house decisions might be affected by external political pressure.

Although the newly-emerging field of “disclosure avoidance” is making technical progress in this arena, the issues are not all technical.⁸ In the context of current national policy vis-à-vis immigrants and high-level political dissent as to whether apportionment might, in some states, reflect only the distribution of citizens, not the entire population in each jurisdiction, tabulation/reporting of CVAP (which includes the data element “citizenship”), the Secretary of Commerce's alleged reason for block-level tabulation is extremely worrisome.



If the Census Bureau’s Disclosure Review guidelines were irresponsibly modified to permit tabulation of CVAP for census blocks with as few as 11 households (i.e. blocks with an average population of about 29 persons), as can be seen in **Table 1** above then about one-fifth of the households in Alameda County, one-third of those in Fresno County, and virtually all of those in Sierra County (83%) would need to be protected using data suppression or data disclosure avoidance techniques.

This level of data suppression or utilization of other “disclosure avoidance” techniques to protect confidentiality would generate such a high level of missing and/or masked data for the smaller blocks with 1-10 households to make analysis of CVAP meaningless.⁹ Neither traditional privacy protection such as data-swapping nor modern disclosure avoidance strategies can be designed to provide meaningful privacy protection for these very small geographies.¹⁰ Under these circumstances, accurate analysis for mapping CVAP would actually have been degraded by the administration’s politically-motivated plan to tabulate at the block level.¹¹

Even if the Census Bureau’s policy decision were to ignore the very serious concerns about data disclosure which threatened but didn’t definitively compromise confidentiality and tabulations were produced for all blocks except those with one household (the level where individual responses would indisputably be illegally revealed) there would still be 6.7% of census blocks in California where data on CVAP (or other household characteristics) would not be available, making any analyses of CVAP that included such blocks highly questionable.¹²

However, even if block-level tabulations or tabulations of clusters of blocks were attempted, the reliability of these tabulations would be highly questionable as a result of increased levels of census non-response and the subsequent inaccuracies inherent in hot deck imputation, the Census Bureau’s last resort.¹³

Secretary Ross envisions “modelling” to overcome the dramatic increases in level of missing data which would result from adding the citizenship question. However, such modeling is infeasible for small-area geographies and would be particularly questionable for analysis of CVAP in census tracts with concentrations of non-citizens since census response would be biased on the key variable the data collection was striving to examine (citizenship status).

If the citizenship question were to be included in Census 2020, responsible guidelines for tabulation of citizenship data would almost certainly require that no tabulations below the block-group level be produced.



Will Recourse to Administrative Records Provide a “Magic Bullet” to Solve the Problem of Degraded Quality of CVAP data stemming from skewed non-response?

Secretary Ross’s proposal pretends to overcome the problem of non-response by linking census records for non-responding households to administrative records to “fill in” information on citizenship status. However, Census Bureau researchers have shown that administrative records cannot be successfully linked to many immigrant households—especially those that are “linguistically isolated” and those headed by recent immigrants.¹⁴

Moreover, other research shows serious discrepancies between tabulations of citizenship status derived from administrative records and American Community Survey research—so that, even in the cases where some sort of linkage is feasible, it is difficult to determine which of the two types of data sources is accurate (in part because key administrative databases are not up to date).

Publication of block-level analyses of “census data” derived from administrative records such as SSA’s Numident database or tax returns submitted to the IRS would still be illegal (as well as unreliable—especially in areas with high levels of household mobility) since the administrative data would have been imported into census records and, therefore, fall under the purview of Title 13 also.

It also deserves note that, even in cases where there is an administrative record for a head of household, key data elements needed for analyzing the CVAP (citizenship status and age) may not be available for other individuals in the household.¹⁵ This is an unsurpassable barrier to reliance on administrative records in areas with many “complex” households where housing is shared with unrelated family/social units.

Further Concerns Regarding the Legality of Collecting Data on the CVAP and Census Block-level Tabulation

An August 7, 2018 report by the Committee on National Statistics of the National Academy of Sciences (CNSTAT) highlights a particularly troubling facet of the Department of Commerce’s plans for use of citizenship data it seeks to gather in Census 2020. The report notes that the citizenship data might be used to create a national registry of citizens. If such a plan were implemented, census respondents’ confidentiality would be further compromised and there would be good reason to view this particular use of census data as being for a purpose other than the intended one. The CNSTAT report observes that use of census data to create such a dataset would be illegal under Title 13.¹⁶



Summary of Concerns about Confidentiality/Accuracy of Block-Level Tabulation of CVAP

Our conclusion is that the decision to include a question about citizenship on census 2020 and tabulate CVAP at the census block level in order to better inform implementation of the Voting Rights Act is fatally flawed in terms of data analysis. At the same time, it violates federal law.

Even if it were to be determined in the course of current litigation that the Secretary of Commerce had the authority to mandate inclusion of the citizenship question and that the Department of Justice actually had initiated the request (despite evidence to the contrary), his decision is arbitrary and capricious because implementation of such a decision inevitably entails either:

- 1) publicly-funded development and dissemination of bogus analyses of CVAP, or
- 2) violation of federal law.

Block-level Tabulation-Bogus Analytic Results

Even if census block level tabulation were permitted, it would not yield more accurate analysis of the citizen voting-age population (CVAP) than current tabulations of ACS data at the block-group level because of the large error terms generated by missing data, the infeasibility of securing consistently valid data on non-responding households from administrative records, from proxy interviews or via “hot deck” imputation.¹⁷ It would actually degrade the quality of analyses of CVAP patterns.

The resulting analysis of CVAP in other, larger, blocks in many political jurisdictions in California, and other states with significant immigrant populations would be fatally flawed. Non-response levels in census blocks where more than 30% of households are headed by non-citizens would be so high that the margins of error would be close to the size of the block, rendering the data analysis unusable.¹⁸ Efforts to accurately map patterns of CVAP at the block level, already made virtually impossible due to bias stemming from differential non-response stemming from inclusion of the citizenship question, would be further degraded, riddled with “holes” of missing, suppressed data (presumably, the single-household blocks) and “distortions” from utilization of either traditional data swapping or “modern” data disclosure avoidance techniques.

Block-level tabulation of small census blocks including those with a single household—illegal and vacuous

Tabulating CVAP and reporting to the Department of Justice, for all census blocks, including those of one household, clearly violates the promised confidentiality guaranteed by Section 13, as referenced in footnotes 1 and 2. Practically speaking it grossly threatens privacy and increases the risk that census data will be used illegally for non-statistical purposes in the census blocks



with more than one but, nonetheless, relatively few housing units—e.g. those with 2-10 housing units.

The maps used for detention of Japanese-Americans during World War II were derived from block-level census data and thematic coding for the maps was based on each dot representing Japanese-origin individuals.¹⁹ In an era where White House rhetoric and administrative strategies have explicitly sought to detain any undocumented immigrant, including those in mixed-status households, threat of non-statistical use of small-area tabulations of CVAP is significant, especially if decisions regarding the trade-off between confidentiality and accuracy is made by “policy decision makers” within the Census Bureau, Department of Commerce, Department of Justice, or White House.

The Bottom Line

Efforts to add the citizenship question to the decennial census are not only politically-motivated but analytically fraudulent due to their impact in degrading data quality by increasing the extent of differential non-response (based on household characteristics including the citizenship variable). Tabulation of CVAP data at the census block level, if the citizenship question were added to Census 2020, would be still more scientifically and ethically unjustified and would further degrade the quality of resulting analyses compromise the integrity of the Census Bureau.

Secretary Ross should be enjoined from an indisputably bogus statistical endeavor which violates Americans’ civil liberties at the same time that it undermines the Census Bureau’s constitutionally mandated mission of accurately enumerating the population of the US in 2020 and commits the Bureau to collusion with the Department of Justice in utilizing either confidential data or fabricated data analyses in the course of litigation in connection with the Voting Rights Act or for other non-statistical purposes.



End Notes

¹ U.S. Code Title 13 § 9 (a) states that “Neither the Secretary, nor any other officer or employee of the Department of Commerce or bureau or agency thereof, or local government census liaison, may...(1) use the information furnished under the provisions of this title for any purpose other than the statistical purposes for which it is supplied...” Although there are exceptions to this prohibition (13 U.S. Code § 402) they point specifically to sharing of census data on businesses with “Designated Statistical Agencies” which are explicitly identified: The Bureau of Labor Statistics and the Bureau of Economic Analysis. Moreover, 13 U.S. Code § 8 (b) explicitly references tabulations in its discussion of “restriction on use [of census data]” and states, “the Secretary may furnish copies of tabulations and other statistical materials which do not disclose the information reported by, or on behalf of, any particular respondent...”

² Title 13 provisions go on to identify further restrictions on census data use, stating that information provided by respondents cannot be used to “(2) make any publication whereby the data furnished by any particular establishment or individual under this title can be identified;”

³ See Paul Overberg and Janet Adamy, “Trump Administration Plans to Check Your Answers on Census Citizenship Question”, [Wall Street Journal](#), April 3, 2018. They report, “Mr. Ross said asking the question and cross-checking answers [through use of administrative records] ‘will permit the Census Bureau to determine the inaccurate response rate for citizens and noncitizens alike using the entire population.’”.

⁴ Los Angeles County includes rural as well as urban areas. The only California county which is entirely urban is San Francisco County. It has, on the average, 80 housing units per census block.

⁵ See Tourangeau, Roger and Ting Yan, “Sensitive Questions in Surveys”, [Psychological Bulletin](#) 2007, Vol. 133 (5) 859-883 for a detailed discussion of considerations which make a question “sensitive”. In a survey which includes immigrants, a question about citizenship is clearly “sensitive” based on all three elements in Tourangeau and Yan’s analysis: intrusiveness, threat of disclosure, and social desirability. Widely-accepted research guidelines do not prohibit asking sensitive questions in surveys but review research design to assess harm to human subjects. This review typically requires additional scrutiny to security of sensitive data collected, consequences of inadvertent release of confidential data, and limitations on tabulations that rely on the sensitive data.

⁶ Lori Aratani, “Secret use of census info helped send Japanese Americans to internment camps in WWII”, [Washington Post](#), April 6, 2018. Aratani’s reporting is based on historical research published in 2000 and 2007 by historian Margo J. Anderson of the University of Wisconsin and statistician William Seltzer of Fordham University, “After Pearl Harbor: The Proper Role of Population Data Systems in Time of War”, presentation to Population Association of American Annual Meeting, March, 2000. Their distinction between uses of macrodata, mesodata, and microdata are valuable in considering the implications of type of tabulation in relation to potential uses.



⁷ Simson L. Garfinkel, “Modernizing Disclosure Avoidance: Report on the 2020 Disclosure Avoidance Subsystems as Implemented for the 2018 End-to-End Test (Continued)”, presentation to 2017 Census Scientific Advisory Committee Fall Meeting, September 15, 2017.

⁸ A good review of the technical issues confronting the Census Bureau can be found in Simon Garfinkel, John M. Abowd, and Satah Powazek, “Issues Encountered Deploying Differential Privacy”, pre-print, Workshop on Privacy and Electronic Security (WPES), September 6, 2018. The paper includes a good discussion of technical issues involved in determining sensitivity of tabulated data and, consequently, where to set the balance between accuracy and disclosure avoidance. However, sensitivity is not explicitly defined in the paper although implicitly it refers to potential consequences of unauthorized disclosure. Secretary Ross’s proposed tabulation would very clearly be in violation of the sort of balance the authors argue for.

⁹ See Federal Committee on Statistical Methodology “Report on Statistical Disclosure Limitation Methodology”, Office of Information and Regulatory Affairs, Office of Management and Budget, December, 2005 for a comprehensive discussion of a range of statistical disclosure issues and strategies, any of which would render the block-level tabulations unreliable.

¹⁰ The Census Bureau has recently argued (Dr. John Abowd, “Research Matters Blog: Protecting the Confidentiality of America’s Statistics: Adopting Modern Disclosure Avoidance Methods at the Census Bureau”, August 17, 2018) that sophisticated disclosure avoidance techniques provide strong confidentiality protection by introducing noise into tabulations. However, the presentation by Simson Garfinkel, Chief of the Census Bureau’s Center for Disclosure Research, transparently states that there are trade-offs between privacy and accuracy and that decisions about “tuning” these will be made by policymakers. Moreover, there is no evidence that the analysis includes attention considered the “small block” problem and that the results of PL-94-171 block-level tabulations of CVAP produced for redistricting might be used for law enforcement targeting (a non-statistical use). In the case of census blocks with one household, presumably, the tabulation of the single household in the block might include a modification of some variables in the household record but the key construct variable (CVAP—the cross-tabulation of two underlying variables: age and citizenship status) could not be modified without complete loss of accuracy (since it’s a binary variable). We believe the Bureau’s mathematical analysis also disregards the fact that the accuracy of more than 30% of the census household records in many census blocks, particularly those with concentrations of non-citizens (legal and undocumented), will be imputed due to high levels of non-response and, therefore, the underlying data will be of questionable reliability. Anderson and Seltzer’s research found maps for targeting Japanese-American representation where each dot in a thematic map represented 10 persons.

¹¹ In his testimony to the House Committee on Oversight and Government Reform, May 9, 2018, Professor Justin Levitt astutely observed that the plan to “improve” analyses of CVAP sought to degrade accuracy to create the illusion of precision.



¹² See OMB “Standards and Guidelines for Statistical Surveys” (2006) for detailed guidelines of issues to be resolved in any federal survey design. Many of these guidelines (which include attention to survey design and data analysis, as well as data collection procedures) would be violated if the Department of Commerce/Census Bureau operations and block-level tabulations were implemented as planned.

¹³ Analysis of CVAP requires cross-tabulation of two separate variables—citizenship status and age. Hot deck imputation is not a reliable technique for such an analysis because there are serious uncertainties when there is a systemic bias affecting households’ likelihood of response (as is the case when the citizenship question is included). Hot deck imputation is unreliable even in determining the size of a non-responding household and fatally flawed in imputing citizenship status and age to each person living in a non-responding household. Errors in underlying variables used in creating a “construct variable” such as CVAP are additive.

¹⁴ Bhaskar, Renuka; Fernandez, Leticia E.; and Sonya R. Porter, “Assimilation and coverage of the foreign-born population in administrative records”, *Statistical Journal of the IAOS*, 34 (2018), pp. 191-201.

¹⁵ The OMB/Census Bureau “residence rules” incorrectly equate “all persons living in a housing unit” with standard sociological definitions of “household” since many households/housing units in low-income neighborhoods with crowded housing are “doubled-up” family/social units. In these households, administrative records on P1 (the census respondent, presumed to be the head of household) say nothing about other family/social units living under the same roof.

¹⁶ Task Force on the 2020 Census, Committee on National Statistics, National Academy of Science, “Letter Report on the 2020 Census”, August 7, 2018. <http://nap.edu/25215> In addition to providing a detailed explanation of the risk posed by the administration’s intent to create the “population registry” with citizenship data, CNSTAT’s review of the Census Bureau’s plans also includes a detailed and expert discussion of the statistical problems of block-level tabulations.

¹⁷ In California tabulation of ACS data shows that about 10% of all census tracts are ones where more than one-third of households are headed by non-citizens. A preliminary model of the cascade of non-response due to inclusion of the citizenship question and additional concerns leading to non-response among immigrant households (including both non-citizens and naturalized citizens) suggests that actual census response in these tracts will be so low that the characteristics of about one out of four households will need to be imputed—although erroneous enumerations are very high for households “enumerated” based on a proxy interview with a neighbor or via hot deck imputation (due to a strong systemic bias in non-response which makes “donor” households unreliable as proxies for non-responding ones).

¹⁸ Computation of margin of error rests on sample size and distribution of responses. Margin of error for the constructed/computed CVAP variable is a function of the margins of error for the underlying variables—age and citizenship status. For the analysis presented here, for census blocks with 1-10 households in a community context if there is an overall non-response rate of 30% but be as high as 100% in some blocks, the margin of error approximates the value of the computed CVAP in a high proportion of cases in any political jurisdiction. It also should be noted that statistical analysis is particularly



inappropriate when there is non-sampling bias in the survey dataset being analyzed which stems from respondents' refusal to participate in the survey (unit non-response) and/or refusal to answer the question (citizenship status) which is the source of key data in the analysis (item non-response). Finally, it is relevant to note that Census Bureau research on item non-response to the citizenship status question (skipping the question) in the ACS showed overall Hispanic non-response rates of 13%-15% (John Abowd, January 19, 2018 memo to Wilbur Ross). Since the U.S. Hispanic population is about 65% native-born, and assuming that ACS non-response to the citizenship question among Hispanics is concentrated among non-citizens, the item non-response rate for the citizenship question among Hispanic immigrants may actually be as high as 45%--meaning that levels of missing data on citizenship status will be dramatically higher than overall levels of missing data. The bottom-line issue is that these multiple sources of census error are less statistically problematic at higher levels of census geography/data analysis. This means that the block-level tabulations provide only an illusion of accuracy, an illusion which is, in fact, more likely to be misleading than reliable.

¹⁹ See the discussion of the WWII use of mesodata in Seltzer and Anderson (2000). They write "The usefulness of mesodata displayed cartographically can also be seen from figure 6 of the final report on the evacuation submitted by General DeWitt to the Army Chief of Staff [US Army. Western Defense Command and Fourth Army. 1943: 83] which reproduces a set of five population density maps of the "Japanese Population - Western Defense Command Area: 1940." The figure cites the "US Bureau of the Census" as the source and carries the legend, "each dot represents 10 people."