



**Center for Neighborhood  
Knowledge**



## **Evaluating 2020 Census Data for Los Angeles County, California Comparison with Parcel-Based Housing Counts**

Paul Ong and Jonathan Ong<sup>1</sup>

UCLA Luskin School of Public Affairs and UCLA Center for Neighborhood Knowledge  
November 8, 2021

This factsheet is the second in a series to evaluate the quality of the 2020 census enumeration. Such an evaluation is needed because no effort to count all Americans can be perfect given the enormity and difficulties of such a monumental task. Inevitably, some individuals are missed and others are counted more than once. Given this reality, a major question is whether there are systematic biases in results. A census can be roughly accurate in the net count if the number of excluded is offset by the number over counted; nonetheless, there can be systematic bias if the errors are not random. These non-random errors can produce what is known as a differential undercount where some populations and neighborhoods types are relatively undercounted. Previous research by the U.S. Census Bureau finds a differential undercount people of color and other disadvantaged groups.<sup>2</sup> Because of racial and class segregation, the differential undercount is also manifested as systematic biases among neighborhoods. In other words, the count is likely to be artificially lower in disadvantaged places than other places. This form of geographic data inaccuracy further marginalizes communities already suffering from pre-existing burdens. Spatial differential undercount has political ramifications in the process of drawing electoral districts and economic implications in the allocation of public funds and services.

As discussed previously, there is significant concern about the completeness and accuracy of last year's counts, potentially creating substantial differential undercount. There is

---

<sup>1</sup> Paul Ong is a Research Professor at UCLA School of Public Affairs. Jonathan Ong is a researcher at Ong and Associates, a public-interest consulting firm. Affiliations are for identification purpose only, and authors are solely responsible for the content.

<sup>2</sup> U.S. Census Bureau, "Census Bureau Releases Estimates of Undercount and Overcount in the 2010 Census," May 22, 2012, [https://www.census.gov/newsroom/releases/archives/2010\\_census/cb12-95.html](https://www.census.gov/newsroom/releases/archives/2010_census/cb12-95.html)

circumstantial evidence of a sizable problem for 2020 enumeration, an issue rooted in the unforeseen events. The COVID-19 pandemic severely disrupted the 2020 census, creating unforeseen challenges and hurdles not experienced in previous decades. The severity of last year's disruptions forced the Census Bureau to revise the collection process and extend their timeline. The Trump administration's controversial and politically motivated push to include a citizenship question on the questionnaire further complicated the enumeration. Although the effort was unsuccessful, it nonetheless created fear among immigrants, both legal and undocumented. Because the negative pandemic and political impacts were unevenly distributed across neighborhoods, there is a real possibility that the spatial differential undercount would be worse than in previous decades.

We have developed a three-part method to assess the quality of the 2020 census. The first approach compares the PL94 population counts (the redistricting version of the enumeration) at the tract level with population estimates from the ACS (American Community Survey) estimates. We found that the 2020 counts are less correlated with the 2015-19 estimates than the 2010 counts are with the 2005-19 estimates, and the discrepancies are systematic along racial and economic lines. This is consistent with the hypotheses that the pandemic and ex-President Trump's rhetoric adversely affected the accuracy of the 2020 census, and that the differences are consistent with a differential undercount.<sup>3</sup> The second part of the assessment is to compare PL94 data and administrative data. The final part is collecting qualitative information on the field operations.

This technical brief summarizes the second assessment, which compares tract-level counts of housing units from 2009 and 2019 parcel records from the Los Angeles County Assessor, 2010 and 2020 housing-unit counts with PL94, and housing-unit estimates from the 2015-19 American Community Survey.<sup>4</sup> We put the data into a common set of geographic units, 2010 tract boundaries.<sup>5</sup> We utilize the Census Bureau's geographic crosswalk to allocate 2020 PL94 block data to 2010 tracts, applying area weights if a block is split into two or more tracts. The parcel records include all units on residential property, but not those for other use (e.g., commercial, industrial). Some of the excluded units could be "permanent residents," such as single-room occupancy hotels. The parcel information does not include some types of housing included in PL94 and ACS (e.g., residential houseboats, recreational vehicles, vans), nor illegal units.

The analysis compares the correlations among the data sources: 2009 parcel and 2010 PL94, 2019 parcel and 2020 PL94, and 2019 parcel and 2015-19 ACS. Based on the assumption that the pandemic and Trump adversely affected the enumeration, we hypothesize that the 2020

---

<sup>3</sup> <https://knowledge.luskin.ucla.edu/2021/08/18/cnk-research-finds-census-inequity-in-l-a-county/>

<sup>4</sup> The 2009 and 2019 parcels are better aligned with the master address files used by the 2010 and 2020 census, respectively.

<sup>5</sup> We do not include the 2005-09 ACS because of the difficulties and uncertainty of spatially allocating the information into 2010 tract boundaries. It is technically possible, but we believe it is unreliable for our analysis.

counts would be less correlated with parcel data than the 2010 count. There is no *a priori* expectation with regard to whether the 2020 counts would be less or more correlated with parcel data than the ACS estimates. While the ACS was not affected by the pandemic nor Trump, it has sampling error. The comparison is nonetheless potentially useful.

The following table provides descriptive statistics for the datasets. These are for tracts with at least one unit in each of the parcel datasets. Within each time period, the overall difference between the average parcel-based counts and the PL94 counts are roughly the same, although the gap is slightly larger in the later time period. Not surprisingly, the average number of housing units in the 2019 parcel dataset is lower than in 2020 PL94 and 2015-19 ACS.<sup>6</sup> (However, if we include housing units on non-residential parcels, the difference is in the opposite direction.) There are also slight differences in the spread (from the 1<sup>st</sup> to the 99<sup>th</sup> percentiles), with the largest range for the 2020 PL94 data.

Table 1: Descriptive Statistics by Data Source

	1st Pctl	Median	Mean	99th Pctl	Count
2009 Housing Units on Residential Parcels	141	1,315	1,410	3,229	3,265,116
2010 Housing Units in PL94	199	1,382	1,487	3,433	3,443,691
2019 Housing Units on Residential Parcels	152	1,350	1,449	3,365	3,356,460
2020 Housing Units in PL94	189	1,430	1,550	3,674	3,589,977
2015-19 Housing Units in ACS	170	1,420	1,529	3,636	3,539,562

Overall, the two sets of analyses find significant differences, with the 2020 PL94 being less associated with the parcel information than the 2010 PL. Interestingly, 2015-19 ACS performs as well or better than the 2020 PL94 counts. These results are consistent with the hypothesis that the recent census enumeration was seriously disrupted.<sup>7</sup>

The table below presents the results from comparing 2010 and 2020 PL94 counts with their respective parcel-based counts. The first row reports the results for all tracts as defined earlier, with tracts weighted by the number of 2020 parcel-based housing units to account for the size differences among tracts. The results show a statistically significantly higher correlation for the earlier period. This qualitative result holds when no weights are used.

We conducted additional analyses using different samples or specifications to test the robustness of the first set of results. Geographically misaligned tracts are defined as those where a 2020 PL94 block data is split into two or more 2010 tracts. By restricting the sample to only those that are perfectly spatially aligned, we eliminate the potential problem of

<sup>6</sup> We do not include the 2005-09 ACS because that available information is reported in 2000 tract boundaries, thus incompatible with the other data.

<sup>7</sup> It should be noted that discrepancies in the housing count are not equivalent to discrepancies in the population count. There are several factors that can create differences between the housing and population counts: vacancies, enumeration response rates in occupied units, item response rate, accuracy of information by proxy, and accuracy of administrative records.

misallocation contaminating the statistics. The second modification eliminated outliers, those with a percent difference between 2020 PL94 and 2019 parcel below the 1<sup>st</sup> percentile or above the 99<sup>th</sup> percentile. By using 2020 PL94, this favors that source over others. The log of the counts deemphasize tracts with large counts, thus minimizing their influence (leverage) on the correlation coefficient. Finally, we restricted the sample to those in the mid-range, which is defined as the tracts where the number of parcel-based housing units is between 667 to 2,000 housing units. This would be roughly equivalent to the normal range of the population size of tracts (approximately from 2,000 to 6,000 persons). Regardless of the changes, the results show a statistically significant lower correlation for the 2019-20 period than the 2009-10 period. These findings are consistent with the hypothesis that the recent census enumeration was noticeably disrupted by the pandemic and politics.

Table 2:

Correlations of Housing Counts in Parcel Files and PL94 Files					
	PL94 2010	PL94 2020	Difference	N Tracts	P-Value
Housing Units on Residential Parcels					
Weighted, all	0.963	0.938	0.025	2315	<.001
Unweighted, all	0.944	0.927	0.017	2315	<.001
Weighted, w/o misaligned	0.967	0.938	0.029	1794	<.001
Unweighted, w/o misaligned	0.965	0.950	0.015	1794	<.001
Weighted, w/o outlier	0.982	0.970	0.012	2282	<.001
Unweighted, w/o outlier	0.981	0.975	0.007	2282	<.001
Weighted, log	0.958	0.942	0.016	2312	<.001
Unweighted, log	0.947	0.929	0.018	2312	<.001
Weighted, mid-range	0.936	0.924	0.012	1761	<.01
Unweighted, mid-range	0.935	0.925	0.010	1761	<.05

We replicate the above analysis by comparing 2015-19 ACS and 2020 PL94 counts with the 2009 parcel-based counts. As mentioned previously, there are offsetting effects that can influence the outcomes. The 2020 disruptions did not affect the ACS, but it is based on a sample of the housing units, thus subject to sampling error. The average margin of error is approximately 3%, so the ACS estimates are inherently disadvantaged relative to the decennial census due to the difference in data collection method (a survey versus an enumeration). Moreover, the estimates are based on inflation factors (weights), which are subject to other potential error.<sup>8</sup> Despite this limitation, we find that the ACS generally performs as well as (statistically not significant) or better than PL94 (statistically significant higher correlation), as reported in Table 3. These findings reinforce the hypothesis that the recent census enumeration is less accurate when compared to other benchmarks.

<sup>8</sup> U.S. Census Bureau, "Population and Housing Unit Estimates," <https://www.census.gov/programs-surveys/popest.html>.

Table 3:

Correlations of Housing Counts in 2019 Parcels, ACS 2015-19, 2020 PL94					
	ACS 2015-19	PL94 2020	Difference	N Tracts	P-Value
Housing Units on Residential Parcels					
Weighted, all	0.962	0.938	0.024	2315	<.001
Unweighted, all	0.947	0.927	0.020	2315	<.001
Weighted, w/o misaligned	0.969	0.938	0.031	1794	<.001
Unweighted, w/o misaligned	0.966	0.950	0.017	1794	<.001
Weighted, w/o outlier	0.975	0.970	0.005	2282	<.01
Unweighted, w/o outlier	0.976	0.975	0.001	2282	N.S.
Weighted, log	0.963	0.942	0.021	2312	<.001
Unweighted, log	0.954	0.929	0.025	2312	<.001
Weighted, mid-range	0.920	0.924	-0.004	1761	N.S.
Unweighted, mid-range	0.921	0.925	-0.004	1761	N.S.

The findings presented here, along with evidence from other sources cited earlier, strongly indicate that the 2020 census should be viewed with caution. Clearly, additional research will be needed to determine the existence of a systematic differential undercount, and equally as important, to determine the magnitude and pattern of the disparities. This is critical because census numbers have profound political, economic and social consequences. It is vital that we have an accurate portrait of America that is inclusive, representative and fair, and this is likely to require adjustments to the official counts.

---



---

#### References:

Akee, Randall; Ong, Paul M; Rodriguez-Lonebear, Desi. "US Census Response Rates on American Indian Reservations in the 2020 Census and in the 2010 Census." Technical Report. UCLA Center for Neighborhood Knowledge and American Indian Studies Center.

Ong, Paul M and Ong, Jonathan. August 18, 2020. "Persistent Shortfalls and Racial/Class Disparities". Technical Report, UCLA Asian American Studies Center, UCLA Center for Neighborhood Knowledge, and Ong & Associates, 2020.

Ong, Paul M; Ong, Elena; Ong, Jonathan. May 7 and May 12, 2020 "Los Angeles County 2020 Census Response Rate Falling Behind 11 Percentage Points and a Third of a Million Lower than 2010," Technical Report. UCLA Center for Neighborhood Knowledge.

Ong, Paul M and Ong, Jonathan. June 11, 2020. "Persistent Shortfall and Racial/Class Disparities, 2020 Census Self-Response Rate." UCLA Luskin School of Public Affairs and UCLA Center for Neighborhood Knowledge. UCLA Latino Policy & Politics Initiative and Center for Neighborhood Knowledge.

Paul Ong and Jonathan Ong. August 18, 2021. "Potential Differential Undercount in 2020 Census Redistricting Data: Los Angeles County, California," UCLA Center for Neighborhood Knowledge.

---

---

We are thankful to the following individuals for their comments and suggestions: Dowell Myers, Andrew Reamer, Steve Pierson, Christopher Dick, Coleman Harris, and William Frey. This brief is made possible through the authors' generous pro bono time and efforts. As a land grant institution, the UCLA Center for Neighborhood Knowledge acknowledges the Gabrielino and Tongva peoples as the traditional land caretakers of Tovaangar (Los Angeles basin, Southern Channel Islands) and that their displacement has enabled the flourishing of UCLA.